

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO
FACULTAD DE INGENIERIA



ALGUNOS ALGORITMOS DE IDENTIFICACION DE
SISTEMAS DINAMICOS DISCRETOS.

TESIS

que para obtener el título de
Ingeniero Mecánico Electricista
presenta

RAFAEL MAURICIO CANETTI WASSER

Director: ALEJANDRO GUARDA AURAS

México D.F. 1985

Handwritten marks in the bottom right corner, including the letters 'M', 'S', and 'E'.



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

COMENTARIOS PREVIOS

Las técnicas de identificación han tenido un gran impacto en el área de control y ha interesado a numerosos grupos de investigadores de estecampo.

La posibilidad de combinar estas técnicas con herramientas conocidas de control, ha abierto nuevos horizontes de Ingeniería, a través de la implantación de estrategias robustas y eficientes en el control de diversos tipos de sistemas.

Este trabajo fué motivado por el interés que existía en varios grupos de la UNAM, de utilizar técnicas de identificación de sistemas dinámicos discretos.

El objetivo inicial era implementar algunos algoritmos y verificarlos en experimentos simulados para que estuvieran a disposición de utilizarse sobre datos experimentales.

Creemos que estos objetivos fueron ampliamente cubiertos. Los programas están siendo utilizados en el Instituto de Ingeniería y en la División de Estudios de Postgrado de la Facultad de Ingeniería de la UNAM. En cuanto al informe, los capítulos 5, 6 y 7 de este trabajo documentan los fundamentos y propiedades de los algoritmos en cuestión, y en el apéndice D se encuentran los listados de los programas.

Debido a la enorme dispersión de la literatura sobre este tópico aún en desarrollo, la comprensión de las técnicas de identificación obligó a la lectura de gran cantidad de trabajos. Lectura acrecentada además, por el enorme desconocimiento con que inicié este estudio.

Esto fue lo que motivó el contenido de este trabajo; servir por una parte como Informe de lo realizado, para fines de referencia personal, y por otro lado servir a otros compañeros para saber donde recurrir en la literatura, al tratar de aprender algo sobre los problemas asociados a las técnicas de identificación. Por eso se agrupan los capítulos 2, 3 y 4 y los apéndices A y B.

Este trabajo, en resumen, no tiene la aspiración de ser un texto (como podría malinterpretarse por su volumen), sino que en primer lugar es un trabajo de Tesis para aspirar al grado de Ingeniero y en segundo lugar, recoge algunos resultados dispersos en la literatura sirviendo a los fines de referencia que se mencionan en el párrafo anterior.

Quiero agradecer la valiosísima ayuda prestada por amigos e Instituciones, que permitieron el desarrollo de esta Tesis:

- Al Instituto de Ingeniería por posibilitar materialmente su edición
- A la Fundación Javier Barrios Sierra por la ayuda en la mecanografía de este trabajo
- Al Instituto de Investigación en Matemáticas Aplicadas y Sistemas por los valiosos recursos que puso a mi disposición, especialmente en Servicios de Cómputo y Biblioteca
- Al Dr José Luis Farah por su constante orientación y consejos
- Al Dr Antonio Alonso Concheiro por su gran apoyo y aliento
- A los doctores M.D. España, F. Kuhlman y G. Medrano por la paciente lectura del manuscrito
- A los que desarrollaron el área de identificación, especialmente a los Doctores B. Furht, L. Ljung y P. Wellstead, quienes en brevísimos encuentros me aclararon enormes dudas, me alentaron y orientaron con el envío de la literatura
- A quienes colaboraron en la edición de este volumen, en particular a Rosa Chávez, Yolanda Rivera, Esther Santos y Salustio Faugier
- Muy especialmente a mis padres y abuelos, sin cuyo impulso y apoyo permanente, ni este trabajo, ni muchas otras cosas en mi vida hubieran sido posibles.

INDICE

	Pág.
1. INTRODUCCION	1
1.1 Propósito de la identificación	1
1.2 ¿Qué es la identificación paramétrica?	5
1.3 Observaciones sobre el enfoque del problema	7
1.4 Propósito y desarrollo del presente trabajo	9
2. ALGUNOS ELEMENTOS DE ESTIMACION PARAMETRICA	13
2.1 Planteo del problema	13
2.2 Definiciones básicas	15
2.3 Varianza del estimador	22
2.4 Estimación por máxima verosimilitud	46
2.5 Estimación por mínimos cuadrados	57
2.5.1 Mínimos cuadrados y el modelo lineal	57
2.5.2 Interpretación geométrica	61
2.5.3 Enfoque estadístico	63
a) errores independientes	63
b) hipótesis de normalidad	67
c) errores dependientes. Estimación de Markov	82
2.5.4 El problema de mínimos cuadrados en general	86

	Pág.
3. SIMULACION DE RUIDO BLANCO. IDENTIFICACION NO PARAMETRICA	89
3.1 Procesos estocásticos	90
3.1.1 Definiciones	90
3.1.2 Estimación de las funciones de correlación	94
3.1.3 Densidad espectral	97
3.2 Ruido blanco	98
3.2.1 Definición	98
3.2.2 Ruido continuo de banda limitada	101
3.2.3 Señal binaria aleatoria a intervalos discretos	105
3.2.4 Señales binarias pseudo-aleatorias	107
3.2.5 Señales de excitación persistente	111
3.3 Simulación de ruido blanco discreto	113
3.3.1 Método de simulación	113
3.3.2 Pruebas sobre las señales simuladas	118
3.3.3 Señales simuladas	134
3.4 Identificación no paramétrica de sistemas continuos	136
3.4.1 Convolución y correlación	136
3.4.2 Determinación de la respuesta a impulso	139
3.4.3 Respuesta en frecuencia y análisis espectral	146

	Pág.		Pág.
4. MODELOS PARA LA IDENTIFICACION	149	5.2 Mínimos cuadrados y métodos de correlación	201
4.1 Sistemas lineales discretos	149	5.3 Errores correlacionados	203
4.1.1 Introducción	149	5.4 El esquema recursivo (Plackett)	208
4.1.2 Respuesta a pulso y función de transferencia	154	5.5 El esquema recursivo (SROOT)	218
4.1.3 Interpretación operacional de Z^{-1}	156	5.6 Cómputo de la matriz de covarianzas	228
4.1.4 Causalidad. Notación	159	5.7 Mínimos cuadrados y filtro de Kalman	235
4.2 Modelado de las perturbaciones	161	5.8 Extensiones a sistemas variables en el tiempo	238
4.2.1 Respuesta de sistemas lineales discretos a procesos estocásticos	161	5.8.1 Olvido exponencial	239
4.2.2 Una aplicación: identificación no paramétrica de sistemas discretos por correlación	165	5.8.2 Marcha aleatoria multidimensional y filtro de Kalman	246
4.2.3 El teorema de factorización espectral	166	5.9 Mínimos cuadrados repetidos	250
4.2.4 Procesos AR, MA y ARMA	169	5.10 Determinación del retraso	253
4.3 Forma canónica para el sistema y perturbaciones	170	5.11 Aplicación a la estimación de la respuesta a pulso	254
4.3.1 El modelo ARMAX	170	5.11.1 Entradas no aleatorias	254
4.3.2 Relación señal a ruido	176	5.11.2 Entradas aleatorias	255
4.3.3 Casos particulares	180	5.12 Pruebas con datos simulados	257
Conclusiones	184	5.12.1 Las simulaciones	257
5. IDENTIFICACION POR MINIMOS CUADRADOS	187	5.12.2 Precisión e intervalo de estimación	266
5.1 El método y propiedades de convergencia	187	5.12.3 Propiedades asintóticas y efecto del ruido	277
5.1.1 Formulación del problema	187	5.12.4 Estimación de la varianza de los estimadores	299
5.1.2 Convergencia	196.	5.12.5 Estimación de la varianza del ruido	305
		5.12.6 Inicialización exacta vs. arbitraria	311

	Pág.		Pág.
5.12.7 Extensiones a sistemas variables en el tiempo y sistemas con término constante	317	6.4.3 Identificación del sistema E con el algoritmo 2	430
5.12.8 Ruido correlacionado	332	6.4.4 Algunos resultados comparativos	455
5.12.9 Perturbaciones no gaussianas	335	6.4.5 Identificación del sistema F con el algoritmo 2	460
5.12.10 Determinación de la respuesta a pulso	342		
5.12.11 Mínimos cuadrados repetidos. Determinación del orden. Determinación del retraso	355	7. MINIMOS CUADRADOS GENERALIZADOS RECURSIVOS (MCGR)	465
6. MINIMOS CUADRADOS GENERALIZADOS Y ESTIMACION DE MARKOV	387	7.1 El método. Algoritmo 3, Algoritmo 4	465
6.1 Introducción	387	7.2 Implantación de los algoritmos 3 y 4	476
6.2 Estimación de Markov	388	7.3 Identificación del sistema E	478
6.3 Mínimos cuadrados generalizados	391	7.4 Convergencia de MCGR y otros algoritmos recursivos	491
6.3.1 El método	391	7.4.1 Introducción	491
6.3.2 Algoritmo 1	394	7.4.2 Algoritmos recursivos de Máxima Verosimilitud (1) y (2)	493
6.3.3 Implantación del algoritmo 1	397	7.4.3 Forma general del algoritmo recursivo. Ecuación diferencial ordinaria asociada	502
6.3.4 Algoritmo 2	402	7.4.4 Convergencia de MCGR	510
6.3.5 Convergencia del algoritmo 2	405	7.4.5 Convergencia de los algoritmos HCR, MCGR, MVR(1) y MVR(2)	512
6.3.6 Implantación del algoritmo 2	413		
6.4 Experimentos con datos simulados	416	APENDICES	
6.4.1 Las simulaciones. Sistemas E y F	416	A - ALGUNOS PROBLEMAS NUMERICOS	515
6.4.2 Identificación del sistema E con el algoritmo 1	418		

	Pág.
B - TEOREMAS DE CONVERGENCIA	525
C - SEÑALES SIMULADAS DE RUIDO BLANCO	537
D - LISTADO DE LOS PROGRAMAS	559
REFERENCIAS	589

- Antes de iniciar el estudio de estrategias para controlar una planta es necesario tener modelos que describan, en forma conveniente para el control, su comportamiento y la influencia del medio exterior. Sin ese modelo la síntesis de un algoritmo de control resulta imposible.

Muchas veces nos enfrentamos a la situación en que arribar a un modelo adecuado de la planta implica un esfuerzo muy grande, y probablemente muy costoso.

Para obtener ese modelo se recurre a las relaciones físico-químicas básicas que resultan del análisis de los procesos que ocurren en la planta. Aun cuando este análisis pueda dar una estructura general del modelo matemático, difícilmente establece con cierto grado de precisión el valor de sus parámetros.

Una situación así puede ser salvada implantando estrategias de control complejas, lo suficientemente "robustas" como para ser en cierto modo insensibles a cuales sean los valores exactos de los parámetros del modelo.

Un enfoque más sencillo es el de planificar experimentos que obtengan la información faltante o incorporar a las estrategias de control procedimientos que calculen los parámetros de la planta, con base en mediciones sobre el sistema.

- Existen situaciones donde se desconocen las leyes que rigen la di-

CAPÍTULO 1

1. Introducción

1.1 Propósito de la identificación

Antes de definir formalmente qué entendemos por identificar sistemas dinámicos, conviene introducir algunos problemas que justifican la identificación.

Del mismo modo que se tropieza con la dificultad de definir el problema de control, resulta imposible definir el problema que da lugar a la identificación.

Podemos, sin embargo caracterizar algunas situaciones típicas a las que la identificación busca dar una respuesta:

námica de procesos de interés, por lo que no es posible obtener un modelo por procedimiento analíticos.

- Muchos sistemas poseen dinámicas variables en el tiempo. Si las variaciones de sus parámetros son lentas, es posible pensar en utilizar procedimientos de estimación de los valores de los parámetros que sigan sus variaciones, y con base en éstas, calcular estrategias variables de control.
- Otra situación similar se produce en el control de sistemas no lineales, aproximables por una linealización en torno a su punto de operación. Pueden obtenerse estrategias variables de control de la misma manera que antes, si con base en mediciones sobre la operación actual del sistema, se estima un modelo cuyos parámetros se actualicen junto con las variaciones del punto de operación.
- En control estocástico es de suma importancia poseer modelos adecuados de sistemas con perturbaciones aleatorias, que separen dichas perturbaciones de la dinámica de interés.

¿Es posible obtener esos modelos a partir de observaciones sobre el sistema?

¿En qué condiciones y con qué precisión?

Hay varias cosas en común en estas situaciones:

- 1) El modelo: se trata de obtener un modelo matemático de la planta.
- 2) Los datos: se recurre a la experimentación y las observaciones sobre las variables del sistema. En particular las entradas y las salidas.
- 3) El propósito: el objetivo de identificar el modelo es el de diseñar e implantar estrategias de control o al menos predecir el comportamiento del sistema.

Estos tres elementos están estrechamente vinculados:

Si el objetivo final es construir estrategias de control, los modelos a identificar deben ser apropiados para propósitos de control. Las técnicas de correlación, que sirven como se verá, para estimar funciones de respuesta a impulso, son poco apropiadas para control por asignación de polos, o control de varianza mínima.

No es casual la velocidad y la dirección de desarrollo del tema de identificación en las últimas décadas. Junto a las técnicas de control clásico existen numerosas técnicas de determinación de modelos adecuados, el análisis en frecuencia. En el control moderno la tendencia es a utilizar más frecuentemente modelos paramétricos de descripción en variables de estado. El deseo de determinar dichos modelos con base en datos experimentales ha sido un gran impulsor de las técnicas modernas de identificación.

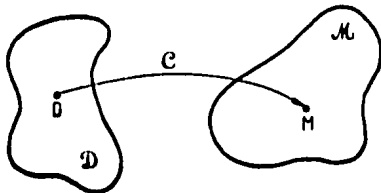
Se pueden construir modelos matemáticos de formas muy diferentes. Dos

formas extremas podrían ser a partir exclusivamente de consideraciones teóricas, o solamente de datos experimentales. La identificación consiste en la determinación de modelos matemáticos a partir de datos experimentales. Como se verá, estos métodos permiten incorporar la información "a priori" que se tenga a partir de consideraciones teóricas. La experiencia en esta área ha confirmado que se obtienen resultados más precisos cuando se incorpora este conocimiento apriorístico.

1.2 Qué es la identificación paramétrica

Desde un punto de vista más general podemos caracterizar la identificación por tres elementos en correspondencia con los que se mencionaron en la sección anterior.

- Los datos $D \in \mathcal{D}$ (conjunto de señales).
- Los modelos $M \in \mathcal{M}$ (conjunto de modelos)
- El criterio C .



El problema de identificación puede resumirse como:

Dados los datos $D \in \mathcal{D}$, encontrar el modelo $M \in \mathcal{M}$, que explique mejor los datos D en el sentido del criterio C .

De esta forma el problema de identificación se transforma en un problema de optimización.

El criterio. El criterio C está vinculado con el objetivo de la identificación y debe medir qué tan bueno es un modelo para explicar los datos observados. En general el criterio se pone en términos de la minimización de una función de costo.

Los datos. Son mediciones sobre el sistema. En lo que se trata en este trabajo son colecciones secuenciales de entradas y salidas del sistema de interés.

Los modelos. Restringiremos el espacio de búsqueda de modelos a un conjunto \mathcal{M} , que en nuestro caso será el de sistemas lineales, invariantes en el tiempo, de parámetros concentrados, en tiempo discreto y de orden finito.

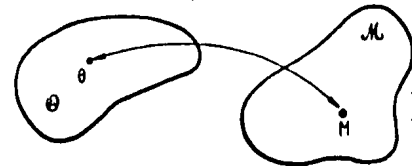


fig 1.2

Este conjunto \mathcal{M} está parametrizado en θ , vector de parámetros. Es decir que hay una correspondencia biunívoca entre los vectores $\theta \in \Theta$, y los sistemas $M \in \mathcal{M}$.

Identificar el modelo M será entonces equivalente a encontrar un vector θ de parámetros, que determine dicho modelo.

Por razones que resultan obvias, esta clase de identificación se denomina identificación paramétrica.

1.3 Observaciones sobre el enfoque del problema.

En todo lo que sigue no haremos distinciones entre los sistemas y los modelos matemáticos que representan su comportamiento.

Hay una serie de preguntas y de imposiciones que debemos hacer a los métodos de identificación, plantearemos algunas de ellas.

Los primeros resultados que se obtuvieron en identificación partían de la base que los datos D eran generados por un sistema $M_0 \in \mathcal{M}$. En virtud de ello la identificación buscaba reconstruir M_0 .

Esta suposición es muy restrictiva pues supone que el sistema real S que produce los datos D es exactamente representable en la clase \mathcal{M} . Para efectos de control supondremos una clase \mathcal{M} muy limitada, mientras que los sistemas reales son por lo general no lineales, de orden no finito, variables con el tiempo, etc. A lo sumo podremos aspirar a aproximar el compor-

tamiento de S con un $M \in \mathcal{M}$. Pero si realmente $M_0 \in \mathcal{M}$, los algoritmos deben ser consistentes, es decir que su resultado debe ser el propio M_0 (o un modelo muy próximo).

Sin olvidar que los sistemas que trabajaremos tienen perturbaciones aleatorias debemos responder a dos interrogantes en la identificación:

i) ¿Cuál es el modelo estimado?

ii) ¿Con qué precisión se obtuvo?

Para responder a esto es muy útil poner el problema en un contexto estadístico y poder usar algunos de los poderosos resultados de la estadística. Es lo que haremos. Debemos destacar, sin embargo, que la mayor parte de los criterios que usaremos en un contexto estadístico tienen sentido aún en un contexto determinístico, por ser.

Los datos que se usarán son experimentales, cabe preguntar entonces si hay formas óptimas de planificar los experimentos:

. ¿qué señales de entrada deben usarse en la experimentación?

. ¿hay señales óptimas, en el sentido de dar resultados más precisos?

Así como la identificación es recurrir a datos experimentales para llegar a modelos matemáticos, también en la búsqueda de respuestas a algunos problemas teóricos de identificación se recurre a la simulación.

Como se verá, la mayor parte de los resultados son de carácter asintótico, es decir que son válidos para un número de observaciones $N \rightarrow \infty$. La primera pregunta que surge es:

- ¿Qué tan válidos son esos resultados para un número finito N de observaciones?

Para responder se han hecho extensísimas simulaciones y experimentos con distintos métodos.

Otras preguntas que son problemas de identificación son:

- ¿Es posible identificar cualquier sistema?
- ¿Es posible identificar indistintamente en lazo abierto o cerrado?
- ¿Cómo afectan las perturbaciones al proceso de identificación?

1.4 Propósito y desarrollo del presente trabajo

Nos limitaremos, como se ha dicho, a la identificación de sistemas lineales en tiempo discreto, de orden finito, invariantes en el tiempo. Las descripciones que veremos son de entrada-salida.

Se estudiarán algunos de los métodos más usados en los últimos años: mínimos cuadrados, mínimos cuadrados repetidos, mínimos cuadrados generalizados, variables instrumentales, máxima verosimilitud y sus versiones (o aproxima-

ciones en algunos casos) recursivas.

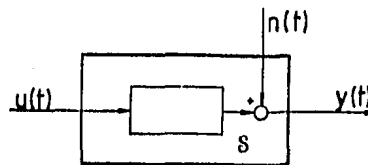


fig 1.3

En la primera parte trataremos los resultados de Estadística más importantes y los criterios que se utilizarán luego. El capítulo 2 está dedicado por eso a la estimación paramétrica.

En el capítulo 3 se unifica la notación de algunos términos de procesos estocásticos. Se generan señales que servirán luego como entradas y perturbaciones de los sistemas simulados.

Se discuten en ese mismo capítulo, métodos clásicos de identificación no paramétrica de sistemas continuos.

La clase de datos D es la de sucesiones $\{u(t), y(t)\}$ de entradas y salidas al sistema (ver figura 1.3). Ambas sucesiones son de elementos acotados para $t = 1, 2, \dots, N$. Junto con los métodos, se determina qué tipo de señal debe ser $u(t)$. (Se introduce el concepto de "excitación persistente").

En el capítulo 4 se estudia la clase de modelos M es decir de los sistemas discretos, lineales. ..., con perturbaciones aditivas de espectro racional.

Los capítulos 5, 6 y 7 están dedicados a los métodos de identificación, o más propiamente a los criterios C y a la discusión de las propiedades de los modelos identificados según cada criterio.

En estos capítulos se realizan simulaciones con sistemas discretos y se identifican éstos para verificar las propiedades teóricas de los algoritmos de identificación estudiados.

Se introducen a través de ejemplos algunas técnicas de determinación del orden y aspectos de la estructura del modelo.

En los apéndices se incluyen los programas que se implantaron para llevar a la práctica los métodos descritos y se discuten algunas consideraciones de carácter numérico y resultados importantes de la estadística.

En resumen, el trabajo puede considerarse dividido en dos grandes partes:

1) la que corresponde a los capítulos 5, 6, 7 que trata de los algoritmos de identificación, y 2) la de los capítulos 2, 3 y 4 que integra una serie de resultados y definiciones usadas a lo largo de la exposición de los algoritmos de identificación.

Si el vector θ no es conocido, pero la variable aleatoria X es observable, se pueda tratar de estimar θ , es decir de inferir su valor a partir de la información contenida en observaciones $X_1, X_2, X_3, \dots, X_n$ de X .

Esto puede hacerse de muchas maneras y con distintos métodos. Se distinguen dos clases especiales de estimación:

- a) Estimación puntual - A partir de las observaciones se estima un valor del parámetro desconocido θ .
- b) Intervalo de estimación - A partir de las observaciones se tiene como resultado un intervalo de valores posibles para θ .

Nos concentraremos en la estimación puntual. Veremos dos métodos de estimación: por mínimos cuadrados y máxima verosimilitud. Otros métodos como por ejemplo: métodos de momentos, mínima distancia, estimadores bayesianos, mínima entropía, etc, pueden verse en M.G. Kendall & A. Stuart [33] (1961); A. Mood, F. Graybill, D. Boes [37] (1974).

Se discuten algunos criterios que permiten juzgar qué tan bueno es un método y comparar distintos métodos entre sí en diversas situaciones.

2. ALGUNOS ELEMENTOS DE ESTIMACION PARAMETRICA

2.1. Planteo del problema

Sea X una variable aleatoria (v.a.) de densidad de distribución (d.d.) conocida en la forma $f_X(x|\theta)$, dependiente del parámetro θ . Se considera que θ es un parámetro vectorial de dimensión finita k :

$$\theta^T = (\theta_1, \dots, \theta_k)$$

donde $\theta_i, i = 1, \dots, k$ son parámetros escalares.

El vector θ determina completamente la función f_X , es decir f_X es un elemento de la familia de funciones $f_X(x|\theta)$, parametrizada en $\theta \in \Theta$ (conjunto de valores posibles de θ).

2.2 Definiciones básicas

Notación. Las variables aleatorias se escriben con mayúsculas, y sus valores con minúsculas.

Convergencia

a) Sea $\{X_i; i = 1, 2, \dots\}$ una sucesión de variables aleatorias vectoriales con funciones de distribución $\{F_i; i = 1, 2, \dots\}$, se dice que X_i converge en distribución (o en ley) a la variable aleatoria X de distribución $F(x)$ si la sucesión $F_i(x)$ converge a $F(x)$ en todo x donde $F(x)$ es continua.

Se escribe: $X_i \xrightarrow{\text{dist.}} X$ (2.1)

b) Se dice que la sucesión $\{X_i; i = 1, 2, \dots\}$ de v.a. X_i converge en probabilidad a una v.a. X , si dados $c, \delta > 0$, existe un natural N tal que para todo $i > N$ se cumple:

$$P\{|X_i - X| > c\} < \delta \quad (2.2)$$

Se escribe

$$X_i \xrightarrow{\text{prob.}} X \quad (2.3)$$

c) Se dice que una sucesión $\{X_i; i = 1, 2, \dots\}$ de v.a. X_i , de segundos momentos finitos, converge en media cuadrática a una v.a. X si:

$$\lim_{i \rightarrow \infty} E\{(X_i - X)^T (X_i - X)\} = 0 \quad (2.4)$$

Se escribe:

$$X_i \xrightarrow{\text{m.c.}} X \quad (2.5)$$

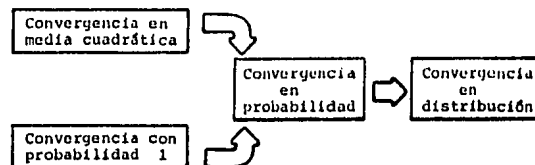
d) Se dice que la sucesión $\{X_i; i = 1, 2, \dots\}$ de v.a. X_i converge con probabilidad 1 (o "casi seguramente", o "excepto en un conjunto de medida nula"), a una v.a. X si dados $c, \delta > 0$, existe un natural N tal que:

$$P\{|X_i - X| < c \forall i > N\} > 1 - \delta \quad (2.6)$$

Se escribe:

$$X_i \xrightarrow{\text{prob. 1}} X \quad (2.7)$$

Las anteriores definiciones de convergencia no son independientes entre sí, guardan las siguientes relaciones:



Obsérvese que no se ha restringido X a ser escalar.

Muestra aleatoria

Diremos que $X_1 \dots X_n$ es una muestra aleatoria (m.a. o muestra aleatoria simple) de largo n de la v.a. X de densidad de distribución $f_X(x|\theta)$ si:

- i) X_i , $i = 1 \dots n$ es a su vez una v.a. de d.d.d. $f_X(x_i|\theta)$.
- ii) La densidad de distribución conjunta de $X_1 \dots X_n$ es:

$$f_{X_1 \dots X_n}(x_1 \dots x_n | \theta) = f_X(x_1 | \theta) \dots f_X(x_n | \theta) \quad (\text{independencia estocástica}).$$

Estadística. Es cualquier vector cuyos elementos son funciones conocidas de la muestra. Siendo sus argumentos variables aleatorias, la estadística es a su vez una variable aleatoria.

P. ej: una estadística T de dim. k es un vector de la forma:

$$T^T = \{T_1(x_1 \dots x_n), T_2(x_1 \dots x_n), \dots, T_k(x_1 \dots x_n)\}$$

De acuerdo a esto T_i , $i = 1 \dots k$ son también estadísticas (escalares).

Una estadística establece una correspondencia $T: \mathbb{X} \rightarrow \mathbb{R}^k$ donde \mathbb{X} es el conjunto de valores posibles de $X_1 \dots X_n$.

$$\mathbb{X} = \{x_1 \dots x_n : f_{X_1 \dots X_n}(x_1 \dots x_n | \theta) > 0\}.$$

Cuando notemos a una estadística como T_n , estaremos diciendo que se trata de una estadística de una muestra de largo n .

Estimador. Diremos que una estadística T es un estimador de una función $\tau(\theta)$ (es vectorial en general), si el valor de T estima el valor de $\tau(\theta)$.

La estadística T , como función se denomina estimador de $\tau(\theta)$, el valor que asume T para una muestra particular $x_1 \dots x_n$ se denomina estimado de $\tau(\theta)$.

Para indicar que T es estimador de $\tau(\theta)$ escribiremos $\hat{\tau}(\theta)$ en lugar de T . Como caso particular, la estadística que estima el parámetro θ se escribe $\hat{\theta}$.

Pasemos a ver cuales propiedades serfa deseable que tuvieran los estimadores.

Un estimador T , como toda estadística es una variable aleatoria, con alguna densidad de distribución $f_T(t|\theta)$. En la figura 2.1 se han dibujado algunas densidades f_T arbitrarias. Para que la estimación sea muy próxima al valor real de $\tau(\theta)$ se podría pedir que los estimadores tuvieran sus densidades de distribución más concentradas alrededor de $\tau(\theta)$. Por ejemplo

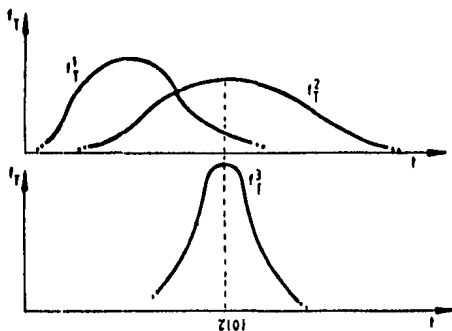


Figura (2.1)

que su media fuera $\tau(\theta)$ y su varianza pequeña. En forma intuitiva se ve que una densidad como f_T^2 es mejor que una como f_T^1 , y una como f_T^3 es mejor que ambas.

Formalicemos algunos conceptos:

Estimador consistente. Se dice que una sucesión T_n (n es el largo de la muestra) de estimadores de cierta función $\tau(\theta)$ del parámetro θ es consistente, si para $n \rightarrow \infty$ T_n converge con probabilidad 1 a $\tau(\theta)$:

$$T_n \xrightarrow{\text{prob. 1}} \tau(\theta)$$

Se dice que la sucesión T_n es debilmente consistente si la convergencia es en probabilidad

$$T_n \xrightarrow{\text{prob.}} \tau(\theta) \quad (2.1)$$

Obsérvese que esta definición de consistencia no implica ninguna condición para muestras finitas, es una propiedad asintótica, para muestras crecientes.

Estimador insesgado. Un estimador T_n de $\tau(\theta)$ es insesgado si su media es $\tau(\theta)$: $E(T_n) = \tau(\theta)$ para todo n y todo $\theta \in \Theta$.

El símbolo E se usa para indicar media (valor esperado).

Notemos la diferencia entre ambas propiedades: la primera implica que el valor esperado tiende al valor real $\tau(\theta)$ y la dispersión se achica arbitrariamente conforme $n \rightarrow \infty$. La segunda no dice nada respecto a la dispersión pero asegura que en todo n la esperanza del estimador es $\tau(\theta)$.

La elección de la media $E(T_n)$ como medida de la bondad de la estimación es arbitraria, se podía haber elegido la moda, mediana, etc.

Esta elección tiene sin embargo grandes ventajas en el cálculo.

Si un estimador no es insesgado, se dice que es sesgado o que tiene sesgo.

Ejemplo 2.1. Sea X una v.a. de media μ y varianza σ^2 . Sea X_1, \dots, X_n una muestra aleatoria de X . Se quiere estimar μ y σ^2 ; $\theta^T = \{\mu, \sigma^2\}$ y se proponen los siguientes estimadores:

$$\hat{\theta}^T = \{\hat{\mu}, \hat{\sigma}^2\}, \text{ con}$$

$$\begin{cases} \hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{cases}$$

$$\text{entonces: } \begin{cases} E(\hat{\mu}) = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} E\left[\sum_{i=1}^n x_i\right] = \frac{1}{n} n\mu = \mu \quad \therefore \text{ es insesgado} \\ E(\hat{\sigma}^2) = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right] = \frac{n-1}{n} \sigma^2 \quad \therefore \text{ es sesgado} \end{cases}$$

Se puede ver (resultado de la ley fuerte de los grandes números, apéndice B) que ambos estimadores son consistentes, y por lo tanto el estimador $\hat{\theta} = \{\hat{\mu}, \hat{\sigma}^2\}^T$ es un estimador consistente y sesgado de $\theta = \{\mu, \sigma^2\}^T$.

Puede mostrarse fácilmente que el estimador $\hat{\theta}^* = \{\hat{\mu}, S^2\}$ es consistente e insesgado, donde $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

2.3 Varianza del estimador

Para un mismo vector $\tau(\theta)$ pueden haber varios estimadores, incluso todos ellos consistentes e insesgados. ¿Qué otro criterio puede permitir la comparación entre estos estimadores? Si un estimador es insesgado, naturalmente surge la varianza como una medida de su bondad. Si el estimador T es real (unidimensional), es claro (ver figura (2.2)) que cuanto menor

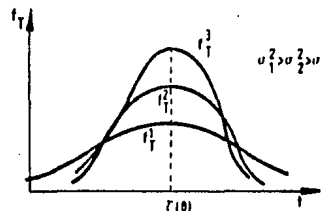


Figura (2.2)

sea su varianza, al estar su densidad de distribución más concentrada alrededor de $\tau(\theta)$, mejor será la estimación. Esta idea se ha usado desde mucho tiempo atrás (Gauss, Laplace), y resultados más modernos (C.R. Rao, Crámer, Aitken, Silverstone, Bhattacharyya, y otros) han mostrado que existen cotas mínimas para la varianza de un estimador.

Si T es un estimador de $\tau(\theta)$ (vector), entonces T es una va-

riable aleatoria multidimensional.

Sean

$$\tau(0) = \begin{bmatrix} \tau_1(0) \\ \vdots \\ \tau_S(0) \end{bmatrix}, \quad T = \begin{bmatrix} T_1(x_1, \dots, x_n) \\ \vdots \\ T_S(x_1, \dots, x_n) \end{bmatrix}$$

Supongamos que T es insesgado, i.e. $E\{T\} = \tau(0)$.

Se define la matriz de covarianzas de T como:

$$\text{cov}\{T\} = \{\sigma_{\ell j}^2\} \quad \text{donde} \quad \sigma_{\ell j}^2 = \text{cov}\{T_\ell, T_j\} \quad (2.9)$$

Esto es

$$\text{cov}\{T\} = E\{(T - \tau)(T - \tau)^T\} \quad (2.10)$$

Nótese que los elementos diagonales son las varianzas de T_ℓ ,

$\ell = 1, \dots, S$

$$\sigma_{\ell\ell}^2 = \text{var}\{T_\ell\} \quad (2.11)$$

Elipsoide de concentración

Se define el elipsoide de concentración de un estimador T insesgado de $\tau(0)$ como el interior y la frontera del elipsoide:

$$[T - \tau(0)]^T \text{cov}\{T\}^{-1} [T - \tau(0)] = S+2 \quad (2.12)$$

El elipsoide de concentración es una medida de la dispersión del estimador T en torno a su valor esperado $\tau(0)$. La distribución de un estimador T' cuyo elipsoide está contenido dentro del elipsoide correspondiente a otro estimador T'' , está más concentrada alrededor de $\tau(0)$.

Se dibuja en la fig. (2.3) el elipsoide de concentración de un estimador arbitrario T , de dim: $S = 2$.

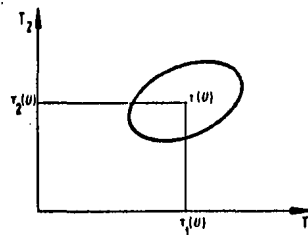


Figura (2.3)

Se puede interpretar el elipsoide de concentración como una generalización del concepto $\mu \pm K\sigma$ en R^1 . ($K = \sqrt{3}$)

Como el determinante de la matriz de covarianzas: $\text{cov}\{T\}$, es proporcional al cuadrado del volumen de su correspondiente elipsoide de concentración, se puede generalizar el concepto

de varianzas:

Varianza generalizada de Wilks. Sea $T = (T_1, \dots, T_S)^T$ un estimador insesgado de $\tau(\theta) = (\tau_1(\theta), \dots, \tau_S(\theta))^T$, se define la varianza generalizada de Wilks como el determinante de la matriz de covarianzas de T :

$$v_w^2 = \det [\text{cov}(T)] = \det [E\{(T - \tau(\theta))(T - \tau(\theta))^T\}] \quad (2.13)$$

Antes de ver cotas mínimas de la varianza de un estimador introduciremos los conceptos de estadísticas suficientes y función de verosimilitud.

Función de verosimilitud de una muestra

Si la densidad de distribución de una v.a. X es $\delta_X(x|\theta)$, que depende de un parámetro θ desconocido, llamaremos función de verosimilitud de una muestra aleatoria X_1, \dots, X_n de X a:

$$L(x_1, \dots, x_n|\theta) = \delta_X(x_1|\theta) \delta_X(x_2|\theta) \dots \delta_X(x_n|\theta) \quad (2.14)$$

Esto es la densidad de distribución conjunta de X_1, \dots, X_n . Se introduce esta notación (por R.A. Fisher) cuando se la considera como función de θ dadas las observaciones x_1, \dots, x_n , distinguiéndola de la densidad conjunta que es función de las observaciones posibles, dado el valor del parámetro θ .

Ejemplo 2.2. Sea X una v.a. de distribución normal $N(\mu, \sigma^2)$. Consideremos una muestra aleatoria de X : X_1, \dots, X_n , la función de verosimilitud resulta:

$$L(x_1, \dots, x_n|\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_2-\mu)^2}{2\sigma^2}} \dots \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$$

luego

$$L(x_1, \dots, x_n|\mu, \sigma^2, T) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \quad (2.15)$$

Teorema 2.1. (Variables aleatorias en función de variables aleatorias)

Sean X_1, \dots, X_n v.a. continuas con densidad conjunta de distribución $\delta_{X_1, \dots, X_n}(x_1, \dots, x_n|\theta)$, y el conjunto $\mathbb{X} = \{(x_1, \dots, x_n) : \delta_{X_1, \dots, X_n}(x_1, \dots, x_n|\theta) > 0\}$. (2.16)

Sea la transformación $Y = G(x)$, $G: \mathbb{X} \rightarrow \mathbb{Y} \subset \mathbb{R}^n$ (donde $G(x_1, \dots, x_n) = (y_1, \dots, y_n)$).

Si se cumple:

- 1) G define una transformación uno a uno entre \mathbb{X} y \mathbb{Y} .
- 2) Existe el jacobiano de la transformación inversa G^{-1} , y

es continuo en \mathcal{Y} :

$$J = \begin{vmatrix} \frac{\partial X_1}{\partial Y_1} & \dots & \frac{\partial X_1}{\partial Y_n} \\ \vdots & & \vdots \\ \frac{\partial X_n}{\partial Y_1} & \dots & \frac{\partial X_n}{\partial Y_n} \end{vmatrix} \quad (2.17)$$

J) El jacobiano J de la transformación g^{-1} no se anula en \mathcal{Y} .

Entonces:

La densidad conjunta de las variables aleatorias Y_1, \dots, Y_n así definidas es:

$$\delta_{Y_1, \dots, Y_n}(y_1, \dots, y_n | \theta) = \begin{cases} |J| \delta_{X_1, \dots, X_n}(g^{-1}(y_1, \dots, y_n) | \theta) & \text{si } (y_1, \dots, y_n) \in \mathcal{Y} \\ 0 & \text{si } (y_1, \dots, y_n) \notin \mathcal{Y} \end{cases} \quad (2.18)$$

Corolario. Sea X una v.a. con d.d.d. $\delta_X(x | \theta)$, $\delta_X > 0$ en el eje real. Sea g una función $g: \mathbb{R} \rightarrow \mathbb{R}$ invertible.

Llamemos $Y = g(X)$, Y es una v.a.; si se cumple $\frac{d g^{-1}(y)}{dy}$ es continua y no nula en el eje real, entonces para Y_1, \dots, Y_n una muestra aleatoria de Y su función de verosimilitud es:

$$L(y_1, \dots, y_n | \theta) = \left| \frac{d g^{-1}(y_1)}{dy} \dots \frac{d g^{-1}(y_n)}{dy} \right| \delta_X(g^{-1}(y_1) | \theta) \dots \delta_X(g^{-1}(y_n) | \theta) \quad (2.19)$$

Estadísticas suficientes. (Referencias [37], [33], [43], [29])

Al utilizar una estadística para estimar un parámetro, decíamos que se trataba de extraer información de la muestra para obtener el valor del parámetro.

Tomando una estadística cualquiera, ¿estará esa información? ¿estará toda la información que contiene la muestra sobre el parámetro?

Esencialmente, la muestra contiene información sobre el parámetro puesto que su densidad de distribución:

$$\delta_{X_1, \dots, X_n}(x_1, \dots, x_n | \theta) \quad (2.20)$$

es una función que depende del parámetro θ .

Las muestras de variables aleatorias escalares son vectores en \mathbb{R}^n ; más exactamente son vectores de un conjunto $\mathbb{I} \subset \mathbb{R}^n$, donde \mathbb{I} está definido por el rango de variación posible de X_1, \dots, X_n . (ver (2.16)).

Una estadística $T_n = (T_{n1}, \dots, T_{nC})^T$ es una función

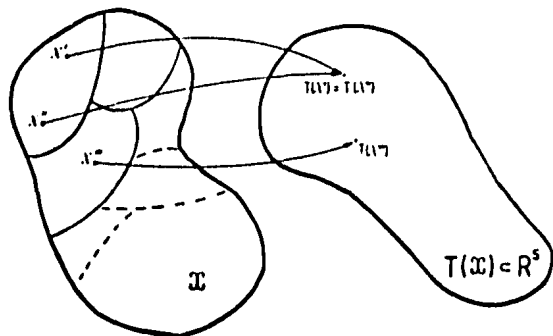
$T_H: \mathbb{X} \rightarrow \mathbb{R}^S$, es decir que toma un valor en \mathbb{R}^S para una muestra cualquiera. Esta estadística establece una partición de \mathbb{X} en clases de equivalencia definidas por la relación de equivalencia

$$x' \sim x'' \Leftrightarrow T_H(x') = T_H(x'')$$

con $x', x'' \in \mathbb{X}$.

De esta forma lo que interesa, más que el valor de la estadística es la estructura de la partición. Es decir, si los elementos x (muestras) de una misma clase contienen todos la misma información sobre el parámetro, el valor de la estadística para esa clase de equivalencia puede considerarse como representante de dicha información.

Nótese que en el sentido de (2.20), la afirmación de que los elementos de una misma clase de equivalencia contengan la mis-



ma información sobre θ quiere decir que dado el valor de la estadística (es decir dada la clase de equivalencia), la expresión (2.2) es independiente de θ dentro de la misma clase.

Formalizando este concepto se dan dos definiciones equivalentes:

Definición 1. Sea X_1, \dots, X_n una muestra aleatoria de una v.a. X de d.d.d. $f_X(x|0)$. Una estadística $T = \{T_1, \dots, T_S\}^T$ se dice que es suficiente para el vector de parámetros θ si y sólo si la distribución condicional de X_1, \dots, X_n dado el valor de la estadística $T = t$, $\{t^T = \{t_1, \dots, t_S\}\}$, no depende de θ para cualquier valor t de T .

Definición 2. Sea X_1, \dots, X_n una m.a. de una v.a. X con d.d.d. $f_X(x|0)$. Se dice que una estadística $T = \{T_1, \dots, T_S\}^T$ es suficiente para el parámetro θ si la d.d.d. condicional de S dado T es independiente de θ , para cualquier estadística S

$$S = S(X_1, \dots, X_n) = \{S_1, \dots, S_S\}^T$$

Nota: Si T es suficiente, se dice que las estadísticas escalares T_1, \dots, T_S son conjuntamente suficientes para θ .

Obsérvese que en la definición 1 se traducen directamente las ideas discutidas antes:

La estadística suficiente contiene toda la información de la muestra sobre el parámetro puesto que una vez conocido el valor de la estadística, la muestra no agrega información sobre el parámetro. En efecto: ¿qué se puede inferir sobre el parámetro θ de una muestra que tiene una distribución que es independiente de θ ?

La definición 1, clara como definición, no es muy útil desde un punto de vista práctico; en general es dificultoso derivar una expresión para la densidad de distribución condicional de la muestra, dada la estadística.

La definición 2 es útil para encontrar cuando una estadística no es suficiente.

Los siguientes teoremas dan herramientas prácticas de trabajo:

Teorema 2.2. Si una estadística T es suficiente, entonces cualquier estadística S tal que S sea resultado de una transformación uno a uno de T , también lo es.

Teorema 2.3. (Criterio de factorización, o criterio de Neyman)

Sea X_1, \dots, X_n una m.a. de una v.a. X de d.d.d. $f_X(x|\theta)$. Una estadística T ($T = \{T_1, \dots, T_s\}^T$), es suficiente para el parámetro vectorial θ si y sólo si la densidad de distribución conjunta de X_1, \dots, X_n puede ser factorizada como:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n | \theta) = g(T | \theta) h(x_1, \dots, x_n) \quad (2.21)$$

donde: $\left\{ \begin{array}{l} \text{i). } h(x_1, \dots, x_n) \geq 0, \text{ no depende del parámetro } \theta \\ \text{ii) } g(T | \theta) \geq 0, \text{ depende de la muestra } X_1, \dots, X_n \text{ sólo a través de la estadística } T \\ \text{iii) el rango de } X_\ell, \ell = 1, \dots, n \text{ es independiente de } \theta \text{ una vez fijado } T. \end{array} \right.$

Observaciones:

- 1) Los teoremas 2.2 y 2.3 son poderosos para construir estadísticas suficientes y para analizar cuando una estadística es suficiente o no.
- 2) Nótese que la expresión (2.21) es exactamente la función de verosimilitud, por lo que el teorema 2.3 podía haberse referido expresamente a ella, y cambiar (2.21) por:

$$L(x_1, \dots, x_n | \theta) = g(T | \theta) h(x_1, \dots, x_n) \quad (2.22)$$

- 3) De acuerdo al teorema de factorización (e incluso a la propia definición 1) cualquier m.a. tiene al menos una estadística suficiente: la propia muestra. En este caso $h(x_1, \dots, x_n) = 1$, $g(T | \theta) = \delta_X(x_1 | \theta) \dots \delta_X(x_n | \theta)$.

Ejemplo 2.3. Siguiendo el mismo caso de los ejemplos 1 y 2, la m.a. X_1, \dots, X_n de una densidad de distribución normal $N(\mu, \sigma^2)$, tenía función de verosimilitud (2.15):

$$L(x_1, \dots, x_n | (\mu, \sigma^2)^T) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \quad (2.23)$$

luego:

$$L(x_1, \dots, x_n | (\mu, \sigma^2)^T) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{n}{2\sigma^2} \left(\frac{T_1}{n} - 2\mu T_2 + \mu^2 \right)} \quad (2.24)$$

donde:

$$T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i^2 \\ \frac{1}{n} \sum_{i=1}^n x_i \end{bmatrix} \quad (2.25)$$

por el teorema 3, la estadística T es suficiente para el parámetro $\theta = (\mu, \sigma^2)^T$.

Distribuciones que poseen estadísticas suficientes

Se vio que la muestra es siempre una estadística suficiente para el parámetro θ de la distribución que la generó. Esto es que X_1, \dots, X_n son un conjunto de estadísticas escalares suficientes para θ . La primer pregunta que surge es si no hay estadísticas $T = (T_1, \dots, T_S)^T$ con $S < n$, suficientes para θ .

La respuesta a esta interrogante es que hay distribuciones que sí poseen esta propiedad. Una respuesta extensa y formal queda fuera del alcance de este trabajo, por lo que se presentará un resultado general, de Huzurbazar (1976) que es extensión de un primer resultado de Koopman y Pitman.

Definición. Una estadística suficiente $T = (T_1, \dots, T_k)^T$ es mínima si para cualquier estadística suficiente $S = (S_1, \dots, S_\lambda)^T$ resulta $k \leq \lambda$. Esto es que no hay otra estadística suficiente de dimensión menor.

Teorema 2.4 (Huzurbazar (1976, ref (29))).

Sea una familia de distribuciones con d.d.d. $\delta_X(x | \theta)$ dependiente de un parámetro vectorial $\theta \in \Omega$ (intervalo no degenerado de R^k). Si se asume que el rango de variación de la v.a. X (el conjunto \mathbb{X} de (2.16)) no depende del parámetro θ , y existen las derivadas de 2o orden $\partial^2 f / \partial \theta_j \partial x_i$ ($i, j = 1, 2, \dots, k$).

Entonces

- i) la familia $\{f_X(x|\theta)\}$ admite una estadística mínima de dim = v para el parámetro θ , para cualquier $\theta \in \Theta$ y toda muestra aleatoria de largo $n > v$ si y sólo si:

$$f_X(x|\theta) = e^{\left[\sum_{j=1}^v U_j(\theta) V_j(x) + A(x) + B(\theta) \right]} \quad (a \leq x \leq b) \quad (2.26)$$

donde:

- U_j son funciones de θ linealmente independientes $j = 1, \dots, v$
- B es función sólo de θ
- V_j son funciones sólo de X , linealmente independientes $j = 1, \dots, v$
- A es función sólo de x
- $a < b$ (finitos o no)

- ii) Si se cumple que $f_X(x|\theta)$ es de la forma (2.26), entonces la estadística $T = \{T_1, \dots, T_v\}^T$ es suficiente y mínima para el parámetro θ

Donde

$$T_j = \sum_{i=1}^n V_j(x_i) \quad ; \quad j = 1, 2, \dots, v \quad (2.27)$$

Este teorema señala la importancia de la familia exponencial de distribuciones.

Debe sin embargo tenerse cuidado, éstas no son las únicas familias que admiten estadísticas mínimas con $S < n$, para sus parámetros. Otras distribuciones que no cumplen las hipótesis del teorema (por ejemplo X independiente de θ), admiten estadísticas mínimas de dimensión $< n$. En la tabla 1 se dan estadísticas mínimas suficientes para tres distribuciones importantes. Otros resultados de carácter más general que el teorema 2.4 se pueden ver en Huzurbazar (1976) [29].


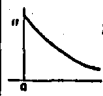
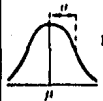
TIPO DE DISTRIBUCION	FORMA DE LA DENSIDAD DE DISTRIBUCION	PARÁMETRO θ	ESTADÍSTICA SUFICIENTE PARA θ
 Uniforme	$f(x (\mu, \lambda a)^T) = \frac{1}{\lambda a} \mathbb{I}_{[\mu - a, \mu + a]}$	$\begin{bmatrix} \mu \\ a \end{bmatrix}$	$\left[\begin{matrix} \sum_{i=1}^n X_i \\ \max X_i \end{matrix} \right]_{i=1, \dots, n}$
 Exponencial	$f(x (a, \sigma)^T) = \frac{1}{\sigma} e^{-\frac{x-a}{\sigma}} \quad \begin{matrix} (a < x < \infty) \\ (-\infty < a < \infty) \\ (\sigma > 0) \end{matrix}$	$\begin{bmatrix} a \\ \sigma \end{bmatrix}$	$\left[\begin{matrix} \sum_{i=1}^n X_i \\ \sum_{i=1}^n \frac{1}{X_i} \end{matrix} \right]_{i=1, \dots, n}$
 Normal	$f(x (\mu, \sigma^2)^T) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < \infty)$	$\begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$	$\left[\begin{matrix} \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i^2 \end{matrix} \right]_{i=1, \dots, n}$

TABLA 2.1

Estamos en condiciones ahora de proponer un resultado general sobre la varianza mínima de un estimador insesgado. Este resultado es una generalización al caso de varios parámetros

del conocido resultado de Cramer-Rao, con dos generalizaciones posteriores: una de Rao-Blackwell, y una de Rao.

Teorema 2.5. Generalización de la cota de Cramer-Rao. (Ref [43])

Sea X_1, \dots, X_n una m.a. de una v.a. X de d.d.d. $f_X(x|\theta)$. Sea $L(x_1, \dots, x_n|\theta)$ la función de verosimilitud de la muestra. Si se cumple que:

a) Existe un estimador $\hat{\tau} = (\hat{\tau}_1, \dots, \hat{\tau}_k)^T$ insesgado, de la función $\tau(\theta) = (\tau_1(\theta), \dots, \tau_k(\theta))^T$ con matriz de covarianzas $\text{cov}(\hat{\tau}) = E\{(\hat{\tau} - \tau(\theta))(\hat{\tau} - \tau(\theta))^T\}$

b) Existe una estadística $T = (T_1, \dots, T_S)^T$ suficiente para θ .

c) Se cumple que (condiciones de regularidad):

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \int \hat{\tau}_i L(x_1, \dots, x_n|\theta) dx_1, \dots, dx_n &= \int \hat{\tau}_i \frac{\partial L(x_1, \dots, x_n|\theta)}{\partial \theta_j} dx_1, \dots, dx_n = \\ &= \frac{\partial \tau_i(\theta)}{\partial \theta_j} \quad (j = 1, \dots, k) \end{aligned}$$

Entonces:

1) Existe un estimador insesgado $\hat{\tau}^*$ de $\tau(\theta)$ que depende de la muestra sólo a través de la estadística suficiente T , tal que:

$$\text{cov}(\hat{\tau}^*) \leq \text{cov}(\hat{\tau}) \quad (2.28)$$

para cualquier estimador $\hat{\tau}$ insesgado de $\tau(\theta)$.

(decimos que dos matrices A y B son $A \leq B \Leftrightarrow B - A$ es positiva semidefinida).

2) Si se llama D a la matriz $(d_{ij}) = \left(\frac{\partial \tau_i}{\partial \theta_j} \right) \Big|_{j=1, \dots, k}$,

entonces:

$$\text{cov}(\hat{\tau}^*) \geq D D^{-1} D^T \quad (2.29)$$

Donde la matriz J es la matriz de información de Fisher:

$$J = E \left\{ \left(\frac{\partial \log L(x_1, \dots, x_n|\theta)}{\partial \theta} \right) \left(\frac{\partial \log L(x_1, \dots, x_n|\theta)}{\partial \theta} \right)^T \right\} \quad (2.30)$$

En caso que J sea singular, en (2.29) se pone la inversa generalizada de J (ver [43]).

Comentarios:

1. El teorema establece que si existe un estimador insesgado y una estadística suficiente, existe un estimador insesgado óptimo en el sentido de varianza mínima generalizada, que discutiremos enseguida. Ese estimador óptimo es función siempre de la estadística suficiente.

2. El sentido de la "optimalidad" del estimador $\hat{\tau}^*$ es como sigue:

- a) de (2.28) se deduce que los elementos diagonales de las matrices $\text{cov}(\hat{\tau}^*)$ y $\text{cov}(\hat{\tau})$ que son respectivamente $\text{var}(\hat{\tau}_i^*)$, $\text{var}(\hat{\tau}_i)$ cumplen:

$$\text{var}(\hat{\tau}_i^*) \leq \text{var}(\hat{\tau}_i), \quad i=1, \dots, k \quad (2.31)$$

luego, cada estimador $\hat{\tau}_i^*$ de $\tau_i(\theta)$ (escalar) es de varianza mínima respecto a cualquier otro estimador $\hat{\tau}_i$ de $\tau_i(\theta)$, si este último es insesgado. De aquí se desprende que los estimadores $\hat{\tau}_i^*$ son los estimadores insesgados de varianza mínima para las funciones $\tau_i(\theta)$.

- b) Existe una cota mínima para las varianzas $\text{var}(\hat{\tau}_i^*)$, y está dada a partir de (2.29) por:

$$\text{var}(\hat{\tau}_i^*) \geq \sum_{\ell=1}^k \sum_{m=1}^k \frac{\partial \tau_i}{\partial \theta_\ell} \frac{\partial \tau_i}{\partial \theta_m} [J^{-1}]_{\ell m} \quad (2.32)$$

- c) El elipsoide de concentración del estimador $\hat{\tau}^*$ está contenido dentro del elipsoide de cualquier otro estimador insesgado $\hat{\tau}$.

- d) Hay un elipsoide mínimo de concentración que siempre está contenido en el de $\hat{\tau}^*$, y en el mejor de los

casos coincide con éste, y es al que corresponde la matriz $D J^{-1} D^T$.

- e) La varianza generalizada de Wilks: $\det(\text{cov}[\hat{\tau}^*])$ del estimador $\hat{\tau}^*$ es menor o igual que la de cualquier otro estimador $\hat{\tau}$ insesgado de $\tau(\theta)$. Hay una cota mínima para esta varianza, que es:

$$V_{W \text{ mín}}(\hat{\tau}) = \det(D J^{-1} D^T) \quad (2.33)$$

En la figura 2.5 se han dibujado elipsoides de concentración arbitrarios para los estimadores $\hat{\tau}^*$, $\hat{\tau}$, y el mínimo elipsoide de concentración.

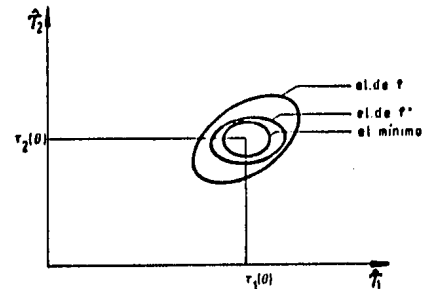


Figura (2.5)

3. El caso que más nos interesa es cuando se estima el parámetro θ . En este caso, llamando $\hat{\tau}^* = \hat{\theta}^*$; $\hat{\tau} = \hat{\theta}$ y

$\tau(\theta) = \theta$, las relaciones (2.28) a (2.33) quedan como

$$\left\{ \begin{array}{l} \text{cov}(\hat{\theta}^*) \leq \text{cov}(\hat{\theta}) \quad (2.34) \\ \text{cov}(\hat{\theta}^*) \geq J^{-1} \quad (2.35) \\ \text{var}(\hat{\theta}_i^*) \leq \text{var}(\hat{\theta}_i) \quad (2.36) \\ \text{var}(\hat{\theta}_i^*) \geq |J^{-1}|_{ii} \quad (2.37) \\ V_{W_{m \times n}}(\hat{\theta}) = \det |J^{-1}| \quad (2.38) \end{array} \right.$$

Nótese que los elementos diagonales de la matriz de información resultan en virtud de (2.37) y (2.36) cotas mínimas para las varianzas de la estimación insesgada de los parámetros.

Definición. Un estimador $\hat{\tau}^*$ insesgado de $\tau(\theta)$ se dice de varianza mínima si para cualquier estimador $\hat{\tau}$ insesgado de $\tau(\theta)$ se cumple

$$\text{cov}(\hat{\tau}^*) \leq \text{cov}(\hat{\tau}) \quad (2.39)$$

Así, el teorema 5 lo que establece es que si existe un estimador insesgado de $\tau(\theta)$ y una estadística suficiente de θ , el estimador de varianza mínima siempre es una función de la estadística suficiente.

Definición. Si el estimador $\hat{\tau}^*$ de varianza mínima de cierta

función $\tau(\theta)$ alcanza la cota del teorema 2.5, es decir si $\text{cov}(\hat{\tau}^*) = \theta J^{-1} \theta^t$ entonces $\hat{\tau}^*$ se llama de cota mínima de varianza.

Obviamente es muy importante poder construir un estimador insesgado de cota mínima de varianza. De la demostración del teorema 5 se puede deducir el siguiente teorema, que veremos sólo para el caso de estimadores del parámetro θ .

Teorema 2.6. Existencia del estimador de cota mínima de varianza, ver ref. [26]

Sujeto a las condiciones de regularidad del teorema 2.5, existe un estimador $\hat{\theta}^*$ insesgado de θ , de cota mínima de varianza si y solo si se puede expresar la $\frac{\partial \log L(x_1, \dots, x_n | \theta)}{\partial \theta}$ en la forma:

$$\left[\frac{\partial \log L(x_1, \dots, x_n | \theta)}{\partial \theta} \right]^T = A(\theta) [\hat{\theta}^* - \theta] \quad (2.40)$$

Si (2.40) es cierta, la matriz $A(\theta)$ es la de información de Fisher:

$$A(\theta) = J \quad (2.41)$$

Eficiencia de un estimador. Un estimador es eficiente si es el de cota mínima de varianza. Para grandes muestras, se define la eficiencia de cualquier estimador (respecto del esti

mejor eficiente), como el inverso de la razón del número de muestras del estimador en cuestión y del estimador eficiente, necesaria para que las varianzas de ambos sean iguales. Es decir para que sean igualmente precisos.

Cálculo de la matriz de información de Fisher

La matriz de información de Fisher es, según (2.30):

$$J = E \left[\left(\frac{\partial \log L(x_1, \dots, x_n | \theta)}{\partial \theta} \right)^T \left(\frac{\partial \log L(x_1, \dots, x_n | \theta)}{\partial \theta} \right) \right]$$

bajo condiciones de regularidad que permiten intercambio de derivada respecto θ e integración en el dominio \mathfrak{X} de las muestras (X_1, \dots, X_n) , y existencia de derivadas segundas de la función de verosimilitud respecto los parámetros θ_i , se puede demostrar que: (llamando $L = L(X_1, \dots, X_n | \theta)$)

$$J = \left(E \left[\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right] \right) \quad i, j = 1, \dots, k \quad (2.42)$$

$$J = - \left(E \left[\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} \right] \right) \quad i, j = 1, \dots, k \quad (2.43)$$

La expresión (2.43) será en muchos casos más fácil de manipular que las expresiones anteriores.

Ejemplo 4: Ejemplificaremos algunos de los conceptos últimos siguiendo el mismo caso de los ejemplos 1, 2 y 3: X_1, \dots, X_n es una m.a. de X con d.d.d. normal $N(\mu, \sigma^2)$.

Habíamos visto que
$$L(x_1, \dots, x_n | \theta) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \quad (2.44)$$

$T = \left[\begin{array}{c} \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i \end{array} \right]$ es una estadística suficiente para $\theta = \left[\begin{array}{c} \mu \\ \sigma^2 \end{array} \right]$

Calculemos la matriz de información de Fisher J :

$$\log L(x_1, \dots, x_n | \mu, \sigma^2)^T = -\frac{1}{2} n \log 2\pi - \frac{1}{2} n \log \sigma^2 - \frac{n}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \quad (2.45)$$

derivando (2.45):

$$\left\{ \begin{array}{l} \frac{\partial^2 \log L}{\partial \mu^2} = -\frac{n}{\sigma^2} ; \quad \frac{\partial^2 \log L}{\partial \mu \partial (\sigma^2)} = -\frac{n}{\sigma^4} \frac{\sum (x_i - \mu)}{\sigma^2} \\ \frac{\partial^2 \log L}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{\sum (x_i - \mu)^2}{\sigma^6} \end{array} \right.$$

luego

$$= - E \begin{bmatrix} \frac{\partial^2 \log L}{\partial \mu^2} & \frac{\partial^2 \log L}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 \log L}{\partial \mu \partial \sigma^2} & \frac{\partial^2 \log L}{\partial (\sigma^2)^2} \end{bmatrix} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix} \quad (2.46)$$

Si queremos estimar el parámetro θ con un estimador $\hat{\theta}$, la cota para la matriz de covarianzas de $\hat{\theta}$ es J^{-1} según el teorema 2.5 (ver (2.35)).

$$\text{luego} \quad \text{cov}(\hat{\theta}) \geq J^{-1} = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix} \quad (2.47)$$

El estimador $\hat{\theta}^*$ propuesto en el ejemplo 1 es:

$$\hat{\theta}^* = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n X_i \\ \frac{1}{n-1} \sum_{i=1}^n (X_i - \frac{1}{n} \sum_{i=1}^n X_i)^2 \end{bmatrix} = \begin{bmatrix} \bar{X} \\ S^2 \end{bmatrix} \quad (2.48)$$

Es claro que este estimador es insesgado y está en función de la estadística suficiente. Su matriz de covarianzas es:

$$\text{cov}(\hat{\theta}^*) = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n-1} \end{bmatrix} \quad (2.49)$$

Se puede demostrar que para cualquier otro estimador insesgado $\hat{\theta}$ de θ es $\text{cov}(\hat{\theta}^*) \leq \text{cov}(\hat{\theta})$, luego es el estimador de varianza mínima pero no alcanza la cota de mínima varianza del teorema 5. (comparar (2.49) y (2.47)).

Habiendo visto algunas propiedades de los estimadores puntuales de parámetros pasaremos a ver en los dos párrafos que siguen, dos métodos clásicos de estimación: Máxima verosimilitud y Mínimos Cuadrados.

Como hasta ahora, seguiremos suponiendo que se hace estimación de parámetros ó función de estos poseyendo la información "a priorística" del tipo de distribución de la muestra.

2.4 Estimación por máxima verosimilitud

Supongamos que X_1, \dots, X_n es un conjunto de variables aleatorias, con función de verosimilitud $L(x_1, \dots, x_n | \theta)$. Recordemos que la función de verosimilitud es la densidad de distribución conjunta de X_1, \dots, X_n , para cada $\theta \in \Theta$.

Supongamos por un momento que el parámetro es conocido $\theta = \theta_0$. La probabilidad de que la observación resulte en un intervalo alrededor de un valor particular x_1^0, \dots, x_n^0 es:

$$P\{x_1^0 < X_1 < x_1^0 + dx_1, \dots, x_n^0 < X_n < x_n^0 + dx_n\} = L(x_1^0, \dots, x_n^0 | \theta) dx_1, \dots, dx_n$$

En la figura (2.6) se ha dibujado $l(x|\theta)$, para $n = 1$, la probabilidad mencionada está representada por el área sombreada.

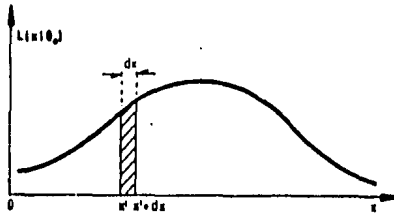


Figura (2.6)

Obviamente el valor más probable alrededor del cual se encuentra la observación x_1, \dots, x_n es aquél de mayor densidad, es decir: para el cual $l(x_1, \dots, x_n | \theta_0)$ es máximo.

Ahora procederemos al revés. El parámetro θ es desconocido y la observación x_1, \dots, x_n de las variables aleatorias X_1, \dots, X_n es conocida. Conociendo la forma de $l(x_1, \dots, x_n | \theta)$, la función de verosimilitud es una familia de funciones parametrizada en θ . La elección de un valor de θ en particular determina una función. Parece una elección ó estimación razonable de θ , la de considerar como su valor aquel $\hat{\theta}$ que haga que la observación x_1, \dots, x_n sea más probable en el sentido que de todas las funciones de la familia, $l(x_1, \dots, x_n | \hat{\theta})$ sea la que tenga el valor más alto en la observación x_1, \dots, x_n .

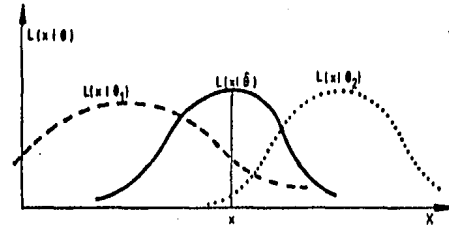


Figura (2.7)

Esto, por supuesto, no es más que una justificación heurística. Definiremos en seguida el estimador de máxima verosimilitud y se mostrarán algunos resultados generales sobre las propiedades de este método de estimación.

Históricamente se atribuye el método a Fisher, pero una primera utilización del método se puede encontrar en Gauss (1809), arribando al método de mínimos cuadrados (Método que veremos en la siguiente sección) a partir de máxima verosimilitud para una función de verosimilitud de una m.a. de una distribución normal $N(0, \frac{1}{h/2})$.

Desarrollaremos ahora las conclusiones que (se derivan) siguen de esta ley. Es evidente, que (es necesario) para que el producto:

$$\Pi = h^u \Pi^{-\frac{1}{2} u} e^{-hh(vu + v^1 v^1 + v^n v^n + \dots)}$$

Sea un máximo, que la suma

$$v + v'v' + v''v'' + \text{etc.},$$

Se vuelva un mínimo. Entonces, eso será el más probable sistema de valores de las cantidades desconocidas $p, q, r, s,$ etc., en el cual la suma de los cuadrados de las diferencias entre los valores observados y computados de las funciones $V, V', V'',$ etc., es un mínimo, si el mismo grado de precisión se presume en todas las observaciones" (Gauss, (1809), [23]).

Definición. Sea $L(\theta) = L(x_1, \dots, x_n | \theta)$ la función de verosimilitud de las v.a. X_1, \dots, X_n , en la observación x_1, \dots, x_n , dependiente de un parámetro $\theta = (\theta_1, \dots, \theta_k)^T$ desconocido; $\theta \in \Theta$ conjunto de valores posibles de θ , $\theta \in \mathbb{R}^k$.

Se define $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$ como el estimador de máxima verosimilitud de θ , si se cumple que

$$L(x_1, \dots, x_n | \hat{\theta}) = \sup_{\theta \in \Theta} L(x_1, \dots, x_n | \theta) \quad (2.50)$$

Nótese que $\hat{\theta}$ es una función de las v.a. X_1, \dots, X_n ; $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$, y es a su vez una variable aleatoria.

Ecuaciones de máxima verosimilitud

Con esta definición no es seguro que el supremo sea alcanzado en un punto interior de Θ .

En general va a ser más sencillo trabajar con el logaritmo de la función $L(\theta)$ que con ella misma. Maximizar $L(\theta)$ equivale a maximizar $\log L(\theta)$ en la suposición cierta que $L(\theta) > 0$; supongamos ahora que Θ es cerrado y $\log L(\theta)$ es diferenciable respecto θ . La búsqueda de máximos se restringe a la frontera y el interior de Θ . Si el máximo está en el interior, este debe ser un punto estacionario con matriz Hessiana negativa semi-definida.

$$\text{(condición necesaria)} \quad \left. \frac{\partial \log L(x_1, \dots, x_n | \theta)}{\partial \theta_i} \right|_{\theta = \hat{\theta}} = 0 \quad i=1, 2, \dots, k \quad (2.51)$$

$$\text{(condición suficiente para ser máximo relativo)} \quad \left\{ \frac{\partial^2 \log L(x_1, \dots, x_n | \theta)}{\partial \theta_i \partial \theta_j} \right\}_{\theta = \hat{\theta}} < 0 \quad i, j=1, 2, \dots, k \quad (2.52)$$

Incluso si se asume que el supremo se alcanza en un punto interior de Θ , las ecuaciones de máxima verosimilitud (2.51) y (2.52) no aseguran que el estimador de máxima verosimilitud sea el único máximo relativo en Θ . Las ecuaciones (2.51) y (2.52) son necesarias; $\hat{\theta}$ debe cumplirlas, pero pueden haber otros valores de θ que las cumplan y que no sean el que corresponda al máximo absoluto.

Cuando existen estadísticas suficientes mínimas de dim $S < n$ pueden asegurarse condiciones de suficiencia:

Máxima verosimilitud cuando existen estadísticas suficientes:

Supondremos que el estimador de máxima verosimilitud es interior a θ y $\log L(\theta)$ es diferenciable.

- a) Si existe una estadística suficiente T entonces el estimador de máxima verosimilitud es función de la estadística suficiente.

En efecto:

Por el criterio de factorización (teorema 2.3),

$$L(x_1, \dots, x_n | \theta) = g(T|\theta) h(x_1, \dots, x_n)$$

luego

$$\log L(x_1, \dots, x_n | \theta) = \log g(T|\theta) + \log h(x_1, \dots, x_n)$$

como $\log L(x_1, \dots, x_n | \theta)$ es diferenciable respecto θ , $\log g(T|\theta)$ también lo es, y la condición necesaria (2.51) queda:

$$\left. \frac{\partial \log g(T|\theta)}{\partial \theta_j} \right|_{\theta = \hat{\theta}} = 0 \quad j = 1, \dots, k \quad (2.53)$$

$\hat{\theta}$ es función de X_1, \dots, X_n solo a través de la estadística suficiente T .

Unicidad

- b) Si existe para θ una estadística mínima suficiente de dim $v > v$ para todo $n > v$, entonces las condiciones (2.51) y (2.52) (ecuaciones de máxima verosimilitud) son suficientes: hay un único máximo relativo de $L(x_1, \dots, x_n | \theta)$ que es máximo absoluto y corresponde al estimador máximo verosímil.

No veremos la demostración, pero se basa en que en esas condiciones, por el teorema 2.4, la función de verosimilitud debe ser de la familia exponencial (la X del teorema 2.4 es multivariada: $X = (X_1, \dots, X_n)$), luego, conocida la forma es fácil ver que hay un solo extremo relativo y ese corresponde a un máximo.

Hay otras propiedades que poseen los estimadores de máxima verosimilitud, en los siguientes teoremas se encuentran algunas de ellas:

Teorema 2.7. Principio de invariancia

Sea $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)^T$ el estimador de máxima verosimilitud de $\theta = (\theta_1, \dots, \theta_k)^T$.

Sea $\tau(\theta)$ una transformación $\tau: \theta \rightarrow \mathbb{R}^S$ (con $1 \leq S \leq k$), entonces el estimador de máxima verosimilitud de $\tau(\theta)$ es $\tau(\hat{\theta})$.

De acuerdo a esta propiedad, para estimar el valor de cualquier función conocida de los parámetros alcanza con estimar el valor de los parámetros.

Teorema 2.8. Optimalidad del estimador de máxima verosimilitud.

Siempre que exista un estimador insesgado de cota mínima de varianza del parámetro θ , este estimador es además el estimador de máxima verosimilitud.

Este teorema muestra que para encontrar estimadores insesgados de cota mínima de varianza, alcanza con investigar el estimador de máxima verosimilitud. El problema es que en general no hay estimadores insesgados de cota mínima de varianza.

Más aún, el estimador de máxima verosimilitud por lo general no es insesgado.

Las propiedades ya mencionadas del estimador de máxima verosimilitud son muy importantes, pero quizá sean más interesantes en nuestras aplicaciones posteriores las propiedades asintóticas para $n \rightarrow \infty$ es decir cuando el número de observaciones crece:

Teorema 2.9. Consistencia

Sea X_1, \dots, X_n una muestra aleatoria de una v.a. X con d.d.d. $f_X(x|\theta)$.

Sea $\hat{\theta}_n$ el estimador de máxima verosimilitud de θ para la muestra de largo n ; consideremos muestras crecientes, es decir $n \rightarrow \infty$, entonces $\hat{\theta}_n$ converge a θ con probabilidad 1, es decir:

$$\hat{\theta}_n \xrightarrow{\text{prob } 1} \theta$$

(La prueba se puede ver en [43])

Teorema 2.10. Normalidad asintótica

Supongamos X_1, \dots, X_n una m.a. de una v.a. X con d.d.d. $f_X(x|\theta)$. Entonces, para $n \rightarrow \infty$ el estimador $\hat{\theta}_n$ de máxima verosimilitud de θ converge en distribución a una variable aleatoria (multivariada por supuesto), de media θ y matriz de covarianzas J^{-1} donde J es la matriz de información de Fisher. Esto es:

$$\hat{\theta}_n \xrightarrow{\text{en dist.}} N(\theta, J^{-1}) \quad (2.54)$$

otra forma útil de poner este resultado es que

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{\text{en dist.}} N(0, J^{-1}) \quad (2.55)$$

donde $\bar{X} = \frac{1}{n} \sum X_i$ es el valor promedio por muestra de la matriz de información de Fisher. (luego $\bar{X}^{-1} = n \bar{X}^{-1}$).

Estas dos últimas propiedades hacen que el estimador de máxima verosimilitud sea muy atractivo. La propiedad de consistencia asegura que para muestras crecientes, para grandes muestras, el estimador converge al parámetro que se está estimando.

La última propiedad establece en que forma se da esa convergencia. Asintóticamente ($n \rightarrow \infty$) el estimador está normalmente distribuido. Se asegura que tiende a ser insesgado y aún cuando para pequeñas muestras no sea de cota mínima de varianza, asintóticamente posee siempre las propiedades óptimas de ser insesgado y tender a la cota mínima de varianza del teorema 2.5 (cota de Rao).

En este sentido es asintóticamente tan bueno como puede llegar a serlo un estimador, por ese se dice que es el mejor estimador normalmente distribuido.

Ejemplo 5. Siguiendo el caso de los ejemplos anteriores, para X_1, \dots, X_n una m.s. de distribución normal $N(\mu, \sigma^2)$:

El logaritmo de la función de verosimilitud era:

$$\log L(x_1, \dots, x_n | \mu, \sigma^2)^T = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{n}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Luego para encontrar el estimador de máxima verosimilitud (como la dist. es normal el teorema de unicidad asegura que la solución es única, ver ec.2.53):

$$\frac{\partial \log L}{\partial \mu} \Big|_0 = 0 \Rightarrow \frac{n}{\sigma^2} \left(\frac{\sum x_i}{n} - \hat{\mu} \right) = 0 \quad \therefore \hat{\mu} = \frac{\sum x_i}{n}$$

$$\frac{\partial \log L}{\partial \sigma^2} \Big|_0 = 0 \Rightarrow -\frac{n}{2\sigma^2} + \frac{\sum (x_i - \hat{\mu})^2}{2(\sigma^2)^2} = 0 \quad \therefore \hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \hat{\mu})^2$$

entonces el estimador de máxima verosimilitud de $\theta = (\mu, \sigma^2)^T$ es:

$$\hat{\theta} = \begin{bmatrix} \frac{n}{\sum x_i} \\ \frac{1}{n} \sum_{i=1}^n (x_i - \frac{\sum x_i}{n})^2 \end{bmatrix}$$

recordando el ejemplo 1 este estimador no es insesgado, sin embargo es consistente y \therefore asintóticamente insesgado. En virtud del teorema 10, su matriz de covarianzas para $n \rightarrow \infty$ es la propia matriz de información (que calculamos en el ejemplo 4) luego

$$\text{cov} \left(\hat{\theta}_n \right) \rightarrow \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^2}{n} \end{bmatrix} \quad \text{para } n \rightarrow \infty$$

2.5 Estimación por mínimos cuadrados

"Pero de todos estos principios, el nuestro es el más simple; por los otros nos veríamos conducidos a los más complicados cálculos" K.F. GAUSS (1809).

2.5.2 Mínimos cuadrados y el modelo lineal

El método de estimación por mínimos cuadrados es uno de los más poderosos y utilizados. Este método tiene sentido tanto en un contexto determinístico como estocástico. Lo introducimos en forma determinística y luego estudiaremos su precisión haciendo suposiciones de carácter estadístico.

Supongamos que tenemos una magnitud \bar{y} que depende de las magnitudes x_λ a través de un modelo teórico:

$$\bar{y} = \theta_1 x_1 + \dots + \theta_p x_p \quad (2.56)$$

Supongamos ahora que se desconocen los parámetros θ_λ , $\lambda = 1, \dots, p$, pero se conocen los datos " x_λ ", y observaciones " y " (mediciones) de la variable \bar{y} , que modelamos en la forma:

$$y = \theta_1 x_1 + \dots + \theta_p x_p + \varepsilon \quad (2.57)$$

Donde ε representa un error. Si para distintos datos hacemos

muchas observaciones y_λ , $\lambda = 1 \dots N$:

$$y_\lambda = \theta_1 x_\lambda^1 + \dots + \theta_p x_\lambda^p + \varepsilon_\lambda; \quad \lambda = 1 \dots N$$

Llamando:

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}; \quad \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix}; \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}; \quad x_\lambda = \begin{bmatrix} x_\lambda^1 \\ \vdots \\ x_\lambda^p \end{bmatrix}; \quad X = \begin{bmatrix} x_1^1 & \dots & x_1^p \\ \vdots & \vdots & \vdots \\ x_N^1 & \dots & x_N^p \end{bmatrix}$$

Se puede poner en forma vectorial:

$$Y = X\theta + \varepsilon \quad (2.58)$$

al que llamaremos el modelo lineal (por razones obvias). La matriz X es de datos, el vector θ es de parámetros, ε de errores, y Y de observaciones.

Conocidos los datos X y las observaciones Y , se trata de estimar el parámetro θ . Sea $\tilde{\theta}$ el valor estimado de θ , nótese que de acuerdo a (2.56) el vector predicho por el modelo teórico es:

$$\tilde{y} = X\tilde{\theta} \quad (2.59)$$

Llamaremos residuos de la estimación la diferencia entre el valor observado de Y y el valor estimado:

$$e = Y - \tilde{Y} = Y - X\tilde{\theta} \quad (2.60)$$

Definimos el estimador $\hat{\theta}$ de mínimos cuadrados de θ a aquel estimador que minimice la suma de los cuadrados de los residuos.

Si llamamos $S(\tilde{\theta})$ a esa suma para el estimador $\tilde{\theta}$, siendo:

$$e = (e_1, \dots, e_N)^T$$

entonces

$$S(\tilde{\theta}) = \sum_{i=1}^N e_i^2 = e^T e \quad (2.61)$$

$$S(\tilde{\theta}) = (Y - X\tilde{\theta})^T (Y - X\tilde{\theta}) \quad (2.62)$$

Luego el estimador $\hat{\theta}$ de mínimos cuadrados es tal que

$$S(\hat{\theta}) = \min_{\tilde{\theta} \in \mathbb{R}^p} S(\tilde{\theta}) \quad (2.63)$$

Supongamos que $\theta \in \mathbb{R}^p$, entonces para encontrar $\hat{\theta}$ derivamos $S(\tilde{\theta})$ respecto $\tilde{\theta}$:

$$\left. \frac{\partial S(\tilde{\theta})}{\partial \tilde{\theta}} \right|_{\tilde{\theta} = \hat{\theta}} = 0 \quad \therefore \quad (X^T X)\hat{\theta} = X^T Y \quad (2.64)$$

Las ecuaciones (2.64) se llaman ecuaciones normales, más adelante justificaremos su nombre. Siendo $S(\tilde{\theta})$ una forma cuadrática, el $\hat{\theta}$ de (2.64) debe ser forzosamente un mínimo. Si $X^T X$ es no singular:

$$\hat{\theta} = (X^T X)^{-1} X^T Y \quad (2.65)$$

Luego este es el estimador de mínimos cuadrados para el modelo lineal (2.50). El método ha sido utilizado desde 1795 cuando fue introducido por K.F. Gauss. Ya vimos en la sección anterior como arriba a este método por consideraciones probabilísticas, pero Gauss va mucho más allá.

"En conclusión, el principio de que la suma de cuadrados de las diferencias entre los valores observados y computados debe ser mínima puede, en la siguiente manera, ser considerado independientemente del cálculo de probabilidades.

Cuando el número de cantidades desconocidas es igual al número de cantidades observadas que dependen de ellas, las anteriores pueden ser determinadas exactamente para satisfacer las últimas. Pero cuando el número de las primeras es inferior al de las últimas, un acuerdo absoluto no puede ser obtenido a menos que las observaciones tengan una exactitud absoluta. En este caso debe tenerse cuidado para establecer el mejor acuerdo posible, o disminuir las diferencias tanto como sea posible.....

... Notando las diferencias entre las observaciones y el cálculo como Δ , Δ' , Δ'' , etc. la primera condición será satisfecha no solo si $\Delta\Delta + \Delta'\Delta' + \Delta''\Delta'' + \text{etc.}$ es un mínimo (que es nuestro principio), sino también si $\Delta^4 + \Delta'^4 + \Delta''^4 + \text{etc.}$ o $\Delta^6 + \Delta'^6 + \Delta''^6 + \text{etc.}$ o en general, si la suma de alguna de las potencias de exponente par se vuelve un mínimo. Pero de todos estos principios el nuestro es el más simple; por los otros nos veríamos envueltos en los cálculos más complicados" K.F. GAUSS (1809) [23].

2.5.2 Interpretación geométrica

Como $X^T \hat{X}0 = X^T y$, se deduce que:

$$X^T (y - \hat{X}0) = 0 \quad (5.66)$$

$$\therefore X^T \epsilon = 0 \quad (5.67)$$

Es decir que el vector de residuos es ortogonal al espacio generado por las columnas de X (de datos).

Esto se debe a que siendo $\hat{y} = X\hat{X}0$ una combinación lineal de las columnas de X , el estimador de mínimos cuadrados es aquel donde $\epsilon^T \epsilon$ es mínimo, es decir que la norma 2 del vector de residuos sea mínima.

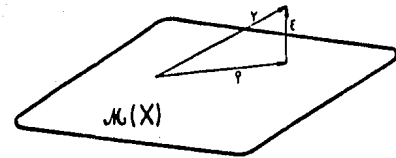


Figura (2.8)

Naturalmente \hat{y} es de todas las combinaciones lineales de columnas de X la que mejor aproxima el vector de observaciones y , entonces \hat{y} es la proyección ortogonal de y sobre el espacio generado por las columnas de X .

Por eso el vector de residuos es ortogonal al espacio $\mathcal{M}(X)$ engendrado por las columnas de X .

El vector y de observaciones se descompone así en dos componentes:

$\hat{y} \in \mathcal{M}(X)$, y $\epsilon \in \mathcal{M}^\perp(X)$ (complemento ortogonal de $\mathcal{M}(X)$):

$$y = \hat{y} + \epsilon$$

A esto deben las ecuaciones (2.64) su nombre de ecuaciones nor

males.

2.5.3 Enfoque estadístico

a) Errores independientes

Consideremos el modelo lineal (2.50). Formalmente lo mantendremos como antes, pero con algunas suposiciones del tipo estadístico.

Suponemos que las observaciones Y son generadas por dicho modelo, donde el vector e es un vector de variables aleatorias $e = [e_1, \dots, e_N]^T$, la matriz X es de datos y es independiente del vector e , el vector θ es un vector de parámetros desconocidos.

Teorema 2.11. (El estimador de mínimos cuadrados es insesgado).

Si el vector e tiene media cero ($E\{e\} = 0$), el estimador de mínimos cuadrados es insesgado.

En efecto:

$$\hat{\theta} = (X^T X)^{-1} X^T Y \quad \text{por (2.65)}$$

$$\hat{\theta} = (X^T X)^{-1} X^T (X\theta + e) \quad \text{por el modelo}$$

$$\therefore \hat{\theta} = \theta + (X^T X)^{-1} X^T e$$

de donde:

$$E\{\hat{\theta}\} = \theta + E\{(X^T X)^{-1} X^T e\} = \theta + (X^T X)^{-1} X^T E\{e\} = \theta$$

$$\therefore E\{\hat{\theta}\} = \theta \quad (2.66)$$

Teorema 2.12. (Covarianza del estimador de mínimos cuadrados)

Si el vector de errores tiene media cero y matriz de covarianzas $\text{cov}\{e\} = E\{e e^T\} = \sigma^2 I$ (donde I es la matriz identidad) entonces:

$$1) \quad \text{Cov}\{\hat{\theta}\} = E\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\} = \sigma^2 (X^T X)^{-1}$$

2) Cualquier otro estimador insesgado $\tilde{\theta}$ lineal en las observaciones Y tiene matriz de covarianzas $\text{cov}\{\tilde{\theta}\}$ tal que

$$\text{cov}\{\tilde{\theta}\} \geq \text{cov}\{\hat{\theta}\} \quad (2.69)$$

En efecto:

$$\begin{aligned} E\{(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T\} &= E\{(X^T X)^{-1} X^T e + e^T X (X^T X)^{-1}\} \\ &= (X^T X)^{-1} X^T E\{e e^T\} X (X^T X)^{-1} \\ &= (X^T X)^{-1} X^T \sigma^2 I X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} \quad \text{lo que muestra la primo-} \end{aligned}$$

ra parte.

Para la segunda parte:

Sea $\tilde{\theta} = CV$; para que sea insesgado $E(\tilde{\theta}) = \theta$. $\therefore E(CV) = \theta$.

\therefore (2.70) $CX_0 = \theta$ para cualquier θ , luego $CX = I$ (2.71)

$$\begin{aligned} E[(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T] &= E[(CV - \theta)(CV - \theta)^T] \\ &= E[(CV - CX_0)(CV - CX_0)^T] \text{ por (2.70)} \\ &= CE[(V - X_0)(V - X_0)^T]C^T \\ &= \sigma^2 C C^T \end{aligned}$$

Si definimos $D = C(X^T X)^{-1}$, es $DD^T \geq 0$

$$\therefore DD^T = CC^T - (X^T X)^{-1} X^T C^T - CX(X^T X)^{-1} + (X^T X)^{-1}$$

$$\therefore DD^T = CC^T - (X^T X)^{-1} \quad (\text{por 2.71})$$

como $DD^T \geq 0 \rightarrow CC^T \geq (X^T X)^{-1}$ que demuestra la segunda parte.

De acuerdo a estos teoremas, cuando el error ϵ tiene media cero y covarianza $\sigma^2 I$, el estimador de mínimos cuadrados es, de todos los estimadores lineales insesgados aquel que tiene varianza mínima (en el sentido de (2.69)).

Se le denomina entonces el MEJOR ESTIMADOR LINEAL INSESADO

(en la literatura inglesa: BLUE).

Debe notarse que no se puede saber si es el estimador de mínima varianza pues no se ha hecho ninguna suposición aún sobre el tipo de densidad de distribución.

Aún cuando se puede asumir la independencia de los errores ϵ_i (y \therefore su matriz de covarianzas es diagonal), difícilmente se conoce de antemano su varianza σ^2 . Esta, de acuerdo a (2.68) es necesaria para saber con qué precisión se estima el parámetro θ .

Por esta razón es importante conocer un buen estimador de σ^2 , a eso se dirige el siguiente

Teorema 2.13. (Estimador de σ^2).

Si ϵ tiene media cero y matriz de covarianzas $\sigma^2 I$, entonces un estimador insesgado de σ^2 es:

$$R^2 = \frac{S(\hat{\theta})}{N-p} \quad (2.72)$$

Donde, como antes $S(\hat{\theta}) = \epsilon^T \epsilon$; N es el no. de observaciones y p es el número de parámetros.

La demostración puede verse en [3], [37], [43].

b) Hipótesis de normalidad

Supongamos ahora que los residuos ε_i no solo tienen matriz de covarianzas $\sigma^2 I$ sino que además están normalmente distribuidas. Con esta suposición se pueden demostrar algunas propiedades de mínimos cuadrados.

Teorema 2.14. (Varianza mínima)

Sea el modelo lineal (2.58), con el vector ε normal multivariado $N(0, \sigma^2 I)$, esto es, que la variable aleatoria Y es normal multivariada $N(X0, \sigma^2 I)$. Entonces el estimador insesgado de varianza mínima es el de mínimos cuadrados.

En efecto:

Calculemos la cota de varianza mínima, para ello debe calcularse la matriz de información de Fisher J . La función de verosimilitud es:

$$L(Y|0, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}N} e^{-\frac{1}{2\sigma^2} (Y-X0)^T (Y-X0)} \quad (2.73)$$

$$\frac{\partial \log L}{\partial \theta} = -\frac{1}{\sigma^2} (Y - X0)^T X \quad (2.74)$$

$$\frac{\partial^2 \log L}{\partial \theta^2} = -\frac{N}{\sigma^2} + \frac{1}{\sigma^4} (Y - X0)^T (Y - X0) \quad (2.75)$$

luego

$$J = E \left[\begin{array}{cc} \frac{\partial \log L}{\partial \theta} \frac{\partial \log L}{\partial \theta} & \left(\frac{\partial \log L}{\partial \theta} \frac{\partial \log L}{\partial \sigma^2} \right) \\ \left(\frac{\partial \log L}{\partial \sigma^2} \frac{\partial \log L}{\partial \theta} \right) & \left(\frac{\partial \log L}{\partial \sigma^2} \frac{\partial \log L}{\partial \sigma^2} \right) \end{array} \right] = \begin{vmatrix} (X^T X) \frac{1}{\sigma^2} & 0 \\ 0 & E \left(\frac{\log L}{\partial \sigma^2} \frac{\partial \log L}{\partial \sigma^2} \right) \end{vmatrix}$$

$$\text{luego debe ser } \text{cov}(\hat{\theta}) \geq \sigma^2 (X^T X)^{-1} \quad (2.76)$$

como ya vimos que $\text{cov}(\hat{\theta}) = \sigma^2 (X^T X)^{-1}$, entonces este estimador es el de varianza mínima. Más aún, de (2.76) se deduce que es de cota mínima de varianza.

Corolario. En las hipótesis del teorema anterior, mínimos cuadrados y máxima verosimilitud conducen al mismo estimador.

Esto se puede demostrar de dos formas: o bien calculando el estimador de máxima verosimilitud y observando que coincide con el de mínimos cuadrados, o tomando en cuenta que en virtud del teorema anterior mínimos cuadrados conduce a un estimador eficiente e insesgado. Por el teorema 2.8 este debe ser además el estimador de máxima verosimilitud.

Para calcular el estimador de máxima verosimilitud igualemos a cero las ecuaciones (2.74) y (2.75) en $\theta = \hat{\theta}_{MV}$ y $\sigma^2 = \hat{\sigma}_{MV}^2$

luego:

$$\begin{cases} V^T X = \hat{\sigma}_{MV}^2 X^T X \text{ o } \hat{\sigma}_{MV} = (X^T X)^{-1} X^T Y & (\text{si } X^T X \text{ es no singular}) \\ \hat{\sigma}^2 = S(\hat{\sigma}_{MV})/N \end{cases} \quad (2.77)$$

en resumen $\hat{\theta}_{M.V.}$ es el de mínimos cuadrados. Esto no es nada nuevo, al maximizar la función de verosimilitud estamos minimizando el exponente de (2.73), luego el criterio de máxima verosimilitud es precisamente la minimización de la suma de los cuadrados de los residuos, en el caso de errores normales.

Esto ya fue visto en la sección anterior, es el razonamiento que sigue Gauss para llegar al principio de mínimos cuadrados por consideraciones probabilísticas.

Debe notarse además que el estimador $\hat{\sigma}_{MV}^2$ no coincide con el estimador R^2 de la ecuación (2.72). Si bien asintóticamente ambos coinciden ($N \rightarrow \infty$), el estimador $\hat{\sigma}_{MV}^2$ de máxima verosimilitud es sesgado.

Si se ha estimado el valor del parámetro θ , o sea que se tiene $\hat{\theta}$, se pueden establecer intervalos donde el valor real θ es $t\hat{\sigma}$, con cierto nivel de confianza.

Teorema 2.15. Normalidad del estimador de mínimos cuadrados.

Sea el modelo lineal (2.58), con el error ϵ normal $N(0, \sigma^2 I)$, entonces el estimador de mínimos cuadrados $\hat{\theta} = (X^T X)^{-1} X^T Y$ está normalmente distribuido con media θ y matriz de covarianzas $\text{cov}(\hat{\theta}) = \sigma^2 (X^T X)^{-1}$.

Demostración:

La media y la varianza ya fueron calculadas antes (teoremas 2.11 y 2.12), lo que queda es demostrar normalidad. Como $\hat{\theta}$ es una función lineal de Y que es normal, entonces también es normal.

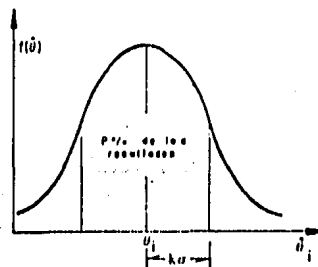


Figura (2.9)

De acuerdo a este teorema, para el estimador $\hat{\theta}_i$ del parámetro θ_i se sabe que su distribución es normal con media θ_i y varianza σ_i^2 donde

$$\sigma_{ii}^2 = \sigma^2 a_{ii} \quad (2.74)$$

Con a_{ii} el i -ésimo elemento diagonal de $(X^T X)^{-1}$. Para una probabilidad P , sabemos que podemos construir un intervalo $\theta_i \pm K \sigma_{ii}$ donde $\hat{\theta}_i$ se encuentre con probabilidad P . Por ejemplo si $P = 68\%$ entonces $K = 1$.

En general:

Si $\hat{\theta}_i$ es normal $N(\theta_i, \sigma^2 a_{ii})$ entonces $Z = \frac{\hat{\theta}_i - \theta_i}{\sigma \sqrt{a_{ii}}}$ es $N(0, 1)$. Luego, llamando K al real tal que

$$P\{|Z| \leq K\} = P\left\{-K \leq \frac{\hat{\theta}_i - \theta_i}{\sigma \sqrt{a_{ii}}} \leq K\right\} = p$$

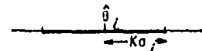
$$\therefore P\{\hat{\theta}_i - K \sigma \sqrt{a_{ii}} \leq \theta_i \leq \hat{\theta}_i + K \sigma \sqrt{a_{ii}}\} = p$$

de esta forma se construye el intervalo $\hat{\theta}_i \pm K \sigma_{ii}$ donde el valor real θ_i esté contenido con una probabilidad P , previamente determinada. Para distintos valores de P se tabularon distintos valores de K .

TABLA 2. Valores de K y P para $\hat{\theta}_i$ distribuido normalmente

P	K
68%	1
90%	1.645
95%	1.96
99%	2.576

$$P = P\{\hat{\theta}_i - K \sigma_{ii} \leq \theta_i \leq \hat{\theta}_i + K \sigma_{ii}\} \quad (2.79)$$



Este resultado obviamente es muy útil para fijar un intervalo y una probabilidad de ocurrencia de θ_i dentro de ese intervalo. La única dificultad es que en general no se conoce el valor de σ^2 , y también hay que estimarlo, por ejemplo con los dos estimadores vistos anteriormente: R^2 y $\sigma_{M.V.}^2$. En el caso que estimemos σ^2 con R^2 se emplea el siguiente

Teorema 2.16

En las mismas condiciones del teorema anterior, la distribución de la v.a. $t = \frac{(\hat{\theta}_i - \theta_i)}{\sqrt{\frac{S(\hat{\theta})}{a_{ii} (N - p)}}}$ es t de Student con $N - p$ grados de libertad.

La demostración se puede ver en [3] ó en [10].

a_{ii} - es el i -ésimo elemento de la diagonal de $(X^T X)^{-1}$
 $S(\hat{\theta})$ - es el valor de la función de costo $c^T c$ en $\theta = \hat{\theta}$

(suma de cuad. de residuos)

N - número de observaciones

P - número de parámetros.

Este teorema permite construir un intervalo alrededor del estimador $\hat{\theta}_i$, donde el valor real θ_i se encontrará con una probabilidad P determinada.

Supongamos que queremos el intervalo para P

Llamando
$$\hat{\sigma}_i = \sqrt{s_{ii} \frac{S[0]}{N-P}}$$

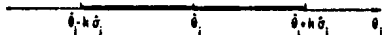


Figura (2.10)

De la distribución t de Student se puede conocer el valor K para el cual

$$P\left\{\left|\frac{\hat{\theta}_i - \theta_i}{\hat{\sigma}_i}\right| \leq K\right\} = P$$

luego
$$P(\hat{\theta}_i - K \hat{\sigma}_i \leq \theta_i \leq \hat{\theta}_i + K \hat{\sigma}_i) = P \quad (2.80)$$

y el intervalo ha sido hallado.

En la tabla III se tabulan valores de K , para un nivel de confianza P de 70%, 90%, 95%, 99% y para distintos grados de libertad $(N - P)$ de la distribución.

TABLA III. Valores de K para distintos niveles de confianza P , y grados de libertad $(N-P)$ de (2.80). Tomados de Kendall y Student [33], a su vez de Sir Roland Fisher & F. Yates).

$(N-P)$	P	70%	90%	95%	99%
20		1.064	1.725	2.086	2.845
40		1.050	1.684	2.021	2.704
60		1.046	1.671	2.000	2.660
120		1.041	1.658	1.980	2.617
∞		1.036	1.645	1.960	2.576

Debe observarse que para $N \rightarrow \infty$, la distribución t $(N-P)$ tiende a la normal $N(0,1)$, y el resultado es aproximable por el teorema 2.15. Sustituyendo σ_i^2 por $\hat{\sigma}_i^2$.

Obsérvese la coincidencia de la tabla II con la III cuando $(N - P) \rightarrow \infty$.

Pudiera ocurrir que se desconozca la distribución de los errores e_i $i = 1, \dots, N$, un resultado general para tener un intervalo de estimación de θ_i se puede obtener a partir de la desigualdad de Tchebicheff, si se conoce σ^2 .

En efecto, cualquiera sea la distribución de $\hat{\theta}_i$ sabemos que

su media es θ_{λ} y su varianza $\sigma_{\lambda}^2 = \sigma^2 a_{\lambda\lambda}$ finita donde $a_{\lambda\lambda}$ es el λ -ésimo elemento de la diagonal de $(X^T X)^{-1}$. Luego por la desigualdad de Tchebicheff:

$$P(|\hat{\theta}_{\lambda} - \theta_{\lambda}| < K \sigma_{\lambda}) \geq 1 - \frac{1}{K^2} \quad (2.81)$$

llamando $P = 1 - \frac{1}{K^2}$, se tiene que:

$$P(\hat{\theta}_{\lambda} - K \sigma_{\lambda} < \theta_{\lambda} < \hat{\theta}_{\lambda} + K \sigma_{\lambda}) \geq P, \quad P = 1 - \frac{1}{K^2} \quad (2.82)$$

Con los valores de K y P tabulados en la tabla IV se tienen intervalos de estimación "sobrados" para el parámetro θ_{λ} , es decir que la probabilidad que θ_{λ} se encuentre en ellos y mayor ó igual que P , según (2.82)

TABLA IV. Valores de K y P para la desigualdad (2.82)

P	K
70%	1.826
90%	3.162
95%	4.472
99%	10.000

Hasta ahora hemos supuesto que del modelo lineal (2.58) se buscaba encontrar θ , conocidas X y Y . En las aplicaciones del método a problemas de identificación veremos que es muy importante tener herramientas para estimar el orden del mode-

lo, es decir p . Para ello utilizaremos principalmente dos resultados, uno referido a la distribución de los residuos y el otro a las funciones de costo que resultan de la estimación de parámetros θ sujetos a restricciones.

Teorema 2.17. (Distribución de los residuos).

Sea el modelo lineal (2.58), que genera las observaciones Y , con ε normal multivariada $N(0, \sigma^2 I)$, es decir $Y = X\theta + \varepsilon$.

Entonces, si $\hat{\theta}$ es el estimador de mínimos cuadrados y ε es el vector de residuos de dicha estimación, se cumple:

- 1) $\varepsilon = Y - X\hat{\theta}$ está normalmente distribuido
- 2) ε y $\hat{\theta}$ son estocásticamente independientes

Demostración:

Como Y y $\hat{\theta}$ son v.a. normales, X constante y $\varepsilon = Y - X\hat{\theta}$ entonces ε debe ser una v.a. normal.

Para probar independencia veremos su covarianza:

$$\begin{aligned} \text{cov}(\hat{\theta}, \varepsilon) &= E\{(\hat{\theta}-\theta)\varepsilon^T\} = E\{[(X^T X)^{-1} X^T Y - \theta][Y - (X^T X)^{-1} X^T Y]^T\} \\ &= E\{[(X^T X)^{-1} X^T Y (Y - X(X^T X)^{-1} X^T Y)^T] - \theta\theta^T X^T X (X^T X)^{-1} X^T Y\} \\ &= E\{[(X^T X)^{-1} X^T Y Y^T (I - X(X^T X)^{-1} X^T) - \theta\theta^T X^T X (X^T X)^{-1} X^T Y]\} \end{aligned}$$

$$\begin{aligned} &= (X^T X)^{-1} X^T [\sigma^2 I + X \tilde{X} \tilde{X}^T] [I - X(X^T X)^{-1} X^T] \\ &= (X^T X)^{-1} [\sigma^2 I + X^T X \tilde{X} \tilde{X}^T] X^T [I - X(X^T X)^{-1} X^T] = 0 \end{aligned}$$

pues

$$X^T [I - X(X^T X)^{-1} X^T] = 0$$

luego como $\text{cov}(\hat{\theta}, \varepsilon) = 0$ y son normales, son independientes.

Mas adelante, en identificación, usaremos la normalidad de como una verificación de la estimación.

Teorema 2.18. (Restricción lineal sobre el parámetro)

Sea el modelo lineal (2.58) con error ε tal que $\text{cov}(\varepsilon) = \sigma^2 I$. El parámetro θ está sujeto a la restricción lineal $L\theta = C$ donde L es una matriz $S \times p$ de rango S . El rango de X es p . Entonces:

1) El estimador insesgado de varianza mínima $\tilde{\theta}$ sujeto a la restricción $L\tilde{\theta} = C$ está dado por:

$$\tilde{\theta} = \hat{\theta} - (X^T X)^{-1} L^T [L(X^T X)^{-1} L^T]^{-1} (L\hat{\theta} - C) \quad (2.13)$$

donde $\hat{\theta}$ es el estimador de mínimos cuadrados sin restricciones. El estimador $\tilde{\theta}$ es de mínimos cuadrados sujeto a la restricción lineal, es decir

$$R_j^2 = \frac{m \ln}{\theta} (Y - X\theta)^T (Y - X\theta) \quad \text{con } L\tilde{\theta} = C$$

2) Si el error ε es normal $N(0, \sigma^2 I)$, llamando

$$\begin{cases} R_1^2 = (Y - X\tilde{\theta})^T (Y - X\tilde{\theta}) \\ R_2^2 = (Y - X\hat{\theta})^T (Y - X\hat{\theta}) \end{cases}$$

y se cumple $L\theta = C$, entonces:

la variable aleatoria $t = \frac{R_1^2 - R_2^2}{R_2^2} \times \frac{N - P}{S} \quad (2.14)$ tiene distribución

F de Fisher: $F(S, N - p)$

La demostración puede verse en [10] cap 3, ó [3] cap 2.

Aplicaciones

Este teorema será el sostén de un método que emplearemos en identificación para determinar el orden correcto de un modelo mediante una prueba de hipótesis estadística.

Supongamos por ejemplo que tenemos un modelo lineal $Y = X\theta + \varepsilon$ con $\varepsilon \sim N(0, \sigma^2 I)$. $\theta = [\theta_1, \dots, \theta_{N_1}, \theta_{N_1+1}, \dots, \theta_{N_2}]$.

La cantidad t en este caso puede verse como una medida del de crecimiento de la función de costo al pasar de un modelo de orden n_1 a uno de orden n_2 , si t es mayor que el valor crítico elegido se rechaza el hecho de que ya hayamos superado el orden del modelo, es decir se considera que la reducción de la función de costo fue significativa.

TABLA V. Valores críticos de $F_{\{1-\alpha\}(n_2-n_1, N-n_2)}$ para la prueba F del teorema (2.18) (tomando de Mood et al [37], a su vez de M. Merrington & C.M. Thompson, Biometrika V.33).

Nivel de significancia	10%				5%				2.5%			
	20	60	120	∞	20	60	120	∞	20	60	120	∞
$n_2 - n_1$												
1	2.97	2.79	2.75	2.71	4.35	4.00	3.92	3.84	5.87	5.29	5.15	5.01
2	2.59	2.39	2.35	2.30	3.49	3.15	3.07	3.00	4.46	3.93	3.80	3.69
3	2.38	2.18	2.13	2.08	3.10	2.76	2.68	2.60	3.80	3.34	3.23	3.12
4	2.25	2.04	1.99	1.94	2.87	2.53	2.45	2.37	3.51	3.01	2.89	2.79
5	2.16	1.95	1.90	1.85	2.71	2.37	2.29	2.21	3.29	2.79	2.67	2.57
6	2.09	1.87	1.82	1.77	2.60	2.25	2.18	2.10	3.13	2.63	2.52	2.41

El siguiente teorema es útil para verificar que nuevas observaciones V_i se hayan obtenido del mismo modelo lineal que las anteriores.

Si queremos probar que el orden verdadero del modelo es n_1 , es decir que los últimos $\theta_{n_1+1}, \dots, \theta_{n_2}$ son nulos, se supone la restricción lineal: $l\theta = C$

$$n_2 - n_1 \left| \begin{array}{cccc} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & & 0 & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{array} \right| \begin{array}{c} 0_1 \\ \vdots \\ 0_{n_1+1} \\ \vdots \\ 0_{n_2} \end{array} = 0 \quad (2.85)$$

$\xleftarrow{n_1}$ $\xrightarrow{n_2 - n_1}$

Llamando R_1^2 a la función de costo calculada imponiendo $\theta_{n_1+1}, \dots, \theta_{n_2} = 0$, y R_2^2 a la calculada sin restricciones, entonces:

Bajo la hipótesis nula $n_2 > n_1 \geq n_0$ (n_0 es el verdadero orden). Resulta que

$$t = \frac{R_1^2 - R_2^2}{R_2^2} \cdot \frac{N - n_2}{n_2 - n_1} \sim F(n_2 - n_1, N - n_2) \quad (2.86)$$

Esto se puede usar como cantidad para la prueba de la hipótesis nula.

Por ejemplo para $n_2 - n_1 = 1$, $N - n_2 = 60$ $F_{0.95}(1, 60) = 4$. Luego $t > 4$ es suficiente para rechazar la hipótesis nula con un nivel de riesgo de 5% (Probabilidad de rechazo de la hipótesis nula siendo cierta).

Teorema 2.19. (Nuevas observaciones)

Sea el modelo lineal (2.58) con $\varepsilon \sim N(0, \sigma^2 I)$. Sea una partición del vector de observaciones $V = (V_1^T \vdots V_2^T)^T$ y la correspondiente partición de la matriz de datos $X = (X_1^T \vdots X_2^T)^T$.

$$\begin{bmatrix} V_1 \\ \dots \\ V_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ \dots \\ X_2 \end{bmatrix} \theta + \begin{bmatrix} e_1 \\ \dots \\ e_2 \end{bmatrix} \quad (2.47)$$

Llamando $\begin{cases} q & \text{- no. de filas de } X_1 \text{ (dim } [V_1]) \\ \lambda_1 & \text{- rango de } X_1 \\ n & \text{- no. de observaciones (dim } [V]) \\ \lambda & \text{- rango de } X \end{cases}$

Entonces la estadística

$$U = \frac{m \lambda n_0 (V_1 - X_1 \theta)^T (V_1 - X_1 \theta)}{m \lambda n_0 (V - X \theta)^T (V - X \theta)} \quad (2.48)$$

tiene distribución (i): $B\left(\frac{q - \lambda_1}{2}, \frac{n - q - \lambda + \lambda_1}{2}\right)$

(La demostración se puede ver en [43] cap. 3).

Este teorema lo utilizaremos cuando haya posibilidades de variación del modelo que genera los datos, para probar la hipótesis de que el modelo se mantiene.

Nótese que el numerador es la función de costo con las observaciones $V_1 = \{y_1, \dots, y_q\}$ y el denominador es la función de costo con las observaciones $V = \{y_1, \dots, y_q, y_{q+1}, \dots, y_n\}$.

Nos va a interesar principalmente el caso en que X_1 y X sean de rango p (no. de parámetros). $\lambda = \lambda_1 = p$; se hacen N observaciones y en la observación $N + 1$ queremos verificar la hipótesis nula: "la nueva observación fue generada por el mismo modelo". Luego $q = N$, $n = N + 1$.

Entonces $U \sim B\left(\frac{N-p}{2}, \frac{1}{2}\right)$

2. Errores dependientes. Estimación de Markov

Hasta ahora hemos tratado el caso que $\text{cov}(\varepsilon) = \sigma^2 I$. Supondremos ahora que los errores mantienen su media cero pero tienen otra matriz de covarianzas. Se puede demostrar que el estimador de mínimos cuadrados sigue siendo insesgado pero ya no es de varianza mínima.

Teorema 2.20.

Sea el modelo lineal (2.58) con errores tales que $E(\varepsilon) = 0$ y matriz de covarianzas $\text{cov}(\varepsilon) = E(\varepsilon \varepsilon^T) = \sigma^2 V$ con V definida positiva.

Entonces si $\hat{\theta}$ es el estimador de mínimos cuadrados

$\hat{\theta} = (X^T X)^{-1} X^T Y$, se cumple:

$$\begin{cases} 1) & E(\hat{\theta}) = \theta \\ 2) & \text{cov}(\hat{\theta}) = \sigma^2 (X^T X)^{-1} X^T V X (X^T X)^{-1} \end{cases}$$

Obsérvese que el caso que hemos estudiado es un caso particular de este resultado más general, cuando $V = I$.

Demostración:

$$\begin{aligned} E(\hat{\theta}) &= E[(X^T X)^{-1} X^T Y] = E[(X^T X)^{-1} X^T (X\theta + e)] = \theta + E[(X^T X)^{-1} X^T e] \\ &= \theta + (X^T X)^{-1} X^T E(e) = \theta \end{aligned}$$

con lo que se demuestra la primera parte.

$$\begin{aligned} \text{cov}(\hat{\theta}) &= E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] = E\left\{[(X^T X)^{-1} X^T Y - \theta][(X^T X)^{-1} X^T Y - \theta]^T\right\} \\ &= E\left\{[(X^T X)^{-1} X^T (Y - X\theta)][(X^T X)^{-1} X^T (Y - X\theta)]^T\right\} \\ &\quad \text{como } Y - X\theta = e \\ &= (X^T X)^{-1} X^T E(e e^T) X (X^T X)^{-1} \\ \therefore \text{cov}(\hat{\theta}) &= \sigma^2 (X^T X)^{-1} X^T V X (X^T X)^{-1} \quad (2.89) \end{aligned}$$

con lo que se demuestra la última parte.

Teorema 2.21. (Estimación de Markov).

Sea el modelo lineal (2.58) con $E(e) = 0$, $\text{cov}(e) = V$, (V def. positiva), entonces el estimador lineal insesgado de varianza mínima de θ es:

$$\hat{\theta}_M = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \quad (2.90)$$

$$\text{cov}(\hat{\theta}_M) = (X^T V^{-1} X)^{-1} \quad (2.91)$$

Demostración:

Siendo V definida positiva (y simétrica), se puede encontrar P tal que

$$V = PP^T \quad (2.92)$$

Con P no singular. Hagamos la transformación $Y_1 = P^{-1} Y$, $X_1 = P^{-1} X$. Luego el modelo lineal (2.58) puede ponerse como:

$$Y_1 = X_1 \theta + P^{-1} e \quad (2.93)$$

$$\text{cov}(P^{-1} e) = E(P^{-1} e e^T P^{-T}) = P^{-1} V P^{-T} = P^{-1} P P^T P^{-T} = I.$$

∴ para el modelo (2.93) el estimador de mínimos cuadrados da la estimación lineal óptima:

$$\hat{\theta}_M = (X_1^T X_1)^{-1} X_1^T Y_1 = (X^T P^{-1} P^{-1} X)^{-1} X^T P^{-1} P^{-1} Y = (X^T V^{-1} V)^{-1} X^T V^{-1} Y$$

Su matriz de covarianzas es:

$$\text{cov}(\hat{\theta}_M) = (X_1^T X_1)^{-1} = (X^T V^{-1} X)^{-1}$$

Observaciones:

1. El estimador $\hat{\theta}_M$ se llama estimado de Markov, se puede interpretar de acuerdo a la transformación $V = PP^T$, como el estimador de mínimos cuadrados que resulta de "filtrar" las observaciones y los datos con una matriz P^{-1} que haga que los errores filtrados tengan covarianza identidad.
2. Para hacer la estimación de Markov hay que conocer previamente la matriz V (módulo alguna constante multiplicativa).
3. Nótese que para cualquier matriz A de rango completo, el estimador

$$\hat{\theta} = (X^T A^{-1} X)^{-1} X^T A^{-1} Y \quad [2.94]$$

es un estimador lineal insesgado para θ , tengan los errores matriz de covarianzas diagonal δ no, siempre que

$$E(e) = 0.$$

2.5.4 El problema de mínimos cuadrados en general

La definición del estimador de mínimos cuadrados dada a través de las ecuaciones (2.62) y (2.63) podría haberse hecho en forma más general definiendo:

$$S'(\tilde{\theta}) = (Y - X\tilde{\theta})^T W(Y - X\tilde{\theta}) \quad [2.95]$$

el estimador de mínimos cuadrados es aquel $\hat{\theta}$ que minimiza la función costo $S(\hat{\theta})$:

$$S'(\hat{\theta}) = \min_{\tilde{\theta} \in O} S(\tilde{\theta}) \quad [2.96]$$

Definiendo el estimador de mínimos cuadrados a través de esta nueva función de costo $S'(\hat{\theta})$, se distinguen en la literatura tres casos principales:

- i) Si $W = I_n$ (matriz identidad) el estimador $\hat{\theta}$ se conoce como el estimador de mínimos cuadrados y es el caso que se estudió en la sección 2.5.
- ii) Si W es una forma definida positiva en general, el procedimiento se conoce usualmente como "mínimos cuadrados ponderados" o pesados.

Cuando W es diagonal se reduce al caso estudiado por Gauss, de distintas precisiones en observaciones independientes, siendo los elementos de la diagonal proporcionales a las varianzas de los errores correspondientes.

A veces se exige que W sea además simétrica pero esto es supérfluo ya que siguiendo el razonamiento de Strejc [20], W admite una descomposición $W = W_A + W_S$ donde W_S es simétrica y W_A es antisimétrica:

$$c^T W c = c^T W_S c + c^T W_A c \quad (2.97)$$

pero $c^T W_A c$ es siempre nulo, luego el único término que contribuye a la suma es el que corresponde a la parte simétrica de la descomposición.

El estimador $\hat{\theta}$ en este caso tiene la expresión:

$$\hat{\theta} = (X^T W X)^{-1} X^T W Y \quad (2.98)$$

Si X y e son independientes, entonces:

$$E(\hat{\theta}) = E\{(X^T W X)^{-1} X^T W (X\theta + e)\} = \theta + (X^T W X)^{-1} X^T W E(e) \quad (2.99)$$

Si además $E(e) = 0$, entonces (2.99) queda:

$$E(\hat{\theta}) = \theta \quad (2.100)$$

con lo que el estimador es insesgado. Es inmediato además que la matriz de covarianzas del estimador es:

$$\text{cov}(\hat{\theta}) = (X^T W X)^{-1} X^T W E W X (X^T W X)^{-1} \quad (2.101)$$

donde

$$E = \text{cov}(e) = E(ee^T)$$

iii) Si $W^{-1} = E = \text{cov}(e)$, la estimación se llama de Markov, y es la que se estudió en el párrafo 2.5.3.c.

Puede demostrarse que de todas las posibles elecciones de W , la elección $W^{-1} = E$ corresponde a la estimación lineal de varianza mínima. Además por (5.98) y (5.101)

$$\hat{\theta} = (X^T E^{-1} X)^{-1} X^T E^{-1} Y \quad (2.102)$$

$$\text{cov}(\hat{\theta}) = (X^T E^{-1} X)^{-1} \quad (2.103)$$

3.1 Procesos estocásticos

3.1.1 Definiciones sobre la caracterización de procesos estocásticos

Veamos algunas definiciones para unificar la notación:

Consideremos dos procesos estocásticos $x(t)$, $y(t)$ donde $t \in T$, T es el conjunto de índices.

El índice t estará asociado en nuestras aplicaciones, a la medida del tiempo.

Si $T = \{R\}$ (reales) o $T = \{R^+\}$ (reales positivos y cero), diremos que se trata de procesos en tiempo continuo.

Si $T = \{Z\}$ (enteros) o $T = \{N\}$ (naturales), diremos que se trata de procesos en tiempo discreto.

Se define:

Media o valor esperado $\mu_x(t)$ de un proceso estocástico $x(t)$ como:

$$\mu_x(t) = E\{x(t)\} \quad (3.1)$$

donde $E(\cdot)$ denota esperanza matemática.

CAPITULO 3.

En este capítulo se discute la simulación de señales de ruido blanco, y se exponen brevemente los fundamentos de métodos clásicos de identificación no paramétrica de sistemas continuos.

La discusión sobre la simulación de ruido blanco cubre dos puntos:

- i) generar las señales que serán usadas en los capítulos 5, 6 y 7 para aplicar los algoritmos de identificación paramétrica sobre sistemas simulados
- ii) introducir la notación y algunos conceptos sobre procesos estocásticos que serán utilizados en los siguientes capítulos, tanto en la introducción como en el estudio de la convergencia de los algoritmos de identificación mencionados.

Función de correlación

$$r_{xy}(t_1, t_2) = E\{x(t_1)y(t_2)\} \quad (3.2)$$

Función de covarianza

$$C_{xy}(t_1, t_2) = E\{(x(t_1) - \mu_x(t_1))(y(t_2) - \mu_y(t_2))\} \quad (3.3)$$

De estas definiciones se desprende que:

$$R_{xy}(t_1, t_2) = C_{xy}(t_1, t_2) + \mu_x(t_1)\mu_y(t_2) \quad (3.4)$$

Si $x(t) = y(t)$ las últimas dos funciones definidas se denominan de autocorrelación y autocovarianza respectivamente.

Estacionariedad. - Un proceso estocástico es estacionario o estrictamente estacionario, si su caracterización probabilística es invariante a cualquier desplazamiento del origen del conjunto de índices $\{T\}$.

Un proceso estocástico se dice estacionario en sentido amplio o débilmente estacionario si sus momentos son invariantes a cualquier desplazamiento del origen del conjunto de índices $\{T\}$.

La segunda definición se introduce porque es más verificable en la práctica.

Nótese que todo proceso estacionario es débilmente estacionario.

rio.

Si $x(t)$ y $y(t)$ son estacionarios, entonces se definen las funciones anteriores como sigue:

$$\left\{ \begin{array}{l} \mu_x(t) = \mu_x = E\{x(t)\} \text{ (constante)} \end{array} \right. \quad (3.5)$$

$$\left\{ \begin{array}{l} R_{xy}(\tau) = E\{x(t)y(t+\tau)\} \end{array} \right. \quad (3.6)$$

$$\left\{ \begin{array}{l} C_{xy}(\tau) = E\{x(t) - \mu_x\}(y(t+\tau) - \mu_y) \end{array} \right. \quad (3.7)$$

Observación:

Hemos dado las definiciones anteriores para unificar notación. Se eligieron por estar extendidas en el campo de control y procesamiento de señales.

Sin embargo, en la literatura estadística, los términos utilizados tienen otra definición formal, por ejemplo la función de correlación se define como $\rho_{xy}(\tau)$:

$$\rho_{xy}(\tau) = \frac{R_{xy}(\tau)}{\sqrt{R_{xx}(0)R_{yy}(0)}} \quad (3.8)$$

cumpliéndose

$$0 \leq |\rho_{xy}(\tau)| \leq 1 \quad \forall \tau$$

Obviamente corresponde a la misma función normalizada.

Procesos ergódicos

Si $x(t)$, $y(t)$ son estacionarios y ergódicos, se cumplen las siguientes igualdades.

En tiempo continuo

$$\left\{ \begin{aligned} \mu_x &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} x(t) dt \\ R_{xy}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} x(t)y(t+\tau) dt \\ C_{xy}(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} [x(t) - \mu_x][y(t+\tau) - \mu_y] dt \end{aligned} \right. \quad (3.9)$$

En tiempo discreto:

$$\left\{ \begin{aligned} \mu_x &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{k=-N}^N x(k) \\ R_{xy}(\tau) &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{k=-N}^N x(k)y(k+\tau) \\ C_{xy}(\tau) &= \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{k=-N}^N [x(k) - \mu_x][y(k+\tau) - \mu_y] \end{aligned} \right. \quad (3.10)$$

La hipótesis de ergodicidad es sumamente poderosa pues esencialmente afirma que toda la caracterización probabilística del proceso puede ser obtenida a partir de cualquiera de sus realizaciones.

En segundo lugar, sugiere una herramienta más práctica para conocer μ, R, C : no se necesitan conocer las densidades de pro-

babilidad conjunta sino promediar a lo largo de cualquier realización.

Las igualdades (3.9) y (3.10) pueden obtenerse a partir de la hipótesis de ergodicidad y la ley de los grandes números (ver apéndice B).

Extensión.- A veces se extienden los conceptos de covarianza y correlación a partir de las ecuaciones (3.9) y (3.10), aún al caso de procesos $x(t)$, $y(t)$ que no se sabe si son estacionarios.

En este caso, los límites de los lados derechos de (3.9) y (3.10) se llaman media, correlación y covarianza empíricas.

3.1.2 Estimación de las funciones de correlación

Aun cuando las ecuaciones (3.9) y (3.10) nos acercan un poco más al cálculo de las funciones de correlación y covarianza, subsisten dos problemas:

El primero es trivial, sólo un cambio de notación. Para computar en líneas las funciones $R_{xy}(\tau)$, $C_{xy}(\tau)$, conviene tener las expresiones en función de valores presentes y pasados, y no en función de valores presentes y futuros. Cambiaremos los productos de la forma $x(t)x(t+\tau)$ por productos de la forma $x(t-\tau)x(t)$, con un cambio trivial de variables.

El segundo problema es grave: hay que promediar a lo largo de un período infinito!

Lo más que podemos hacer entonces, será "estimar" las funciones R_{xy} y C_{xy} con promedios en un intervalo finito.

De ahora en adelante sólo consideramos la función de correlación, el cálculo de la covarianza sólo exige centrar los procesos restando su media.

Definimos estos estimadores como sigue:

$$\hat{R}_{xy}(\tau) = \frac{1}{T-\tau} \int_0^{T-\tau} x(t-\tau)y(t)dt \quad (3.11)$$

para procesos en tiempo continuo, con datos en $-\tau \leq t \leq T-\tau$

$$\hat{R}_{xy}(k) = \frac{1}{N-k} \sum_{i=0}^{N-k-1} x(i-k)y(i) \quad (3.12)$$

para procesos en tiempo discreto, con datos en $-k \leq i \leq N-k-1$.

Estas estimaciones son aproximadas. Cuando los procesos son periódicos, de período T_0 y el intervalo de promediación es ℓT_0 con ℓ entero y $\ell \geq 1$, entonces los estimadores coinciden con las funciones que queremos estimar.

Cabe preguntarse que tan precisas son las estimaciones. Esto depende de las distribuciones de $x(t)$, $y(t)$.

Si se asume que $x(t)$, $y(t)$ son gaussianas, con media cero, entonces Bendat y Piersol (1966) demuestran una expresión para la varianza de los estimadores en el caso continuo:

$$\text{Var}(\hat{R}_{xy}(\tau)) = \frac{1}{T-\tau} \int_{-\infty}^{+\infty} [R_{xx}(u)R_{yy}(u) + R_{xy}(u+\tau)R_{xy}(u-\tau)] du \quad \dots (3.13)$$

para $x(t) = y(t)$:

$$\text{Var}(\hat{R}_{xx}(\tau)) = \frac{1}{T-\tau} \int_{-\infty}^{+\infty} [R_{xx}^2(u) + R_{xx}(u-\tau)R_{xx}(u+\tau)] du \quad (3.14)$$

si $\tau = 0$

$$\text{Var}(\hat{R}_{xx}(0)) = \frac{1}{T} \int_0^{\infty} R_{xx}^2(u) du \quad (3.15)$$

Nótese que en todos los casos la varianza es inversamente proporcional a $T-\tau$.

Para procesos discretos, Bartlett (10) demuestra, a partir de un resultado no publicado de Daniell, que si $x(t)$ es gaussiano, con $E\{x^2(t)\} = 1$ y $\mu_x = 0$, entonces

$$\text{Var}(\hat{R}_{xx}(k)) = \frac{1}{N-k} \sum_{\lambda=-\infty}^{\infty} [R_{xx}^2(\lambda) + R_{xx}(\lambda-k)R_{xx}(\lambda+k) + 2R_{xx}^2(k)R_{xx}^2(\lambda) - 4R_{xx}(k)R_{xx}(\lambda)R_{xx}(\lambda-k)] \quad (3.16)$$

Si existe q , tal que $R_{xx}(\lambda) = 0 \quad \forall \lambda \geq q$, entonces:

$$\text{Var}(\hat{R}_{xx}(k)) = \frac{1}{N-k} \left[1 + \frac{q}{1-k} R_{xx}^2(k) \right], \quad \forall k \geq q \quad (3.17)$$

3.1.3 Densidad espectral

Se define como densidad espectral del proceso estacionario $x(t)$, la transformada de Fourier de su función de autocorrelación R_{xx} .

Llamando ϕ_{xx} a la densidad espectral de $x(t)$:

$$\phi_{xx}(j\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-j\omega\tau} d\tau \quad (3.18)$$

O sea:

$$\phi_{xx}(j\omega) = \{ \mathcal{L}[R_{xx}(\tau)] \}_{s=j\omega} \quad (3.19)$$

Donde \mathcal{L} es la transformada de Laplace bilateral.

Se define la densidad espectral cruzada de dos procesos conjuntamente estacionarios $x(t)$, $y(t)$ a la transformada de Fourier de su función de correlación $R_{xy}(\tau)$.

$$\phi_{xy}(j\omega) = \int_{-\infty}^{\infty} R_{xy}(\tau) e^{-j\omega\tau} d\tau \quad (3.20)$$

$$\phi_{xy}(j\omega) = \{ \mathcal{L}[R_{xy}(\tau)] \}_{s=j\omega} \quad (3.21)$$

Para $x(t)$ estacionario en tiempo discreto:

$$\phi_{xx}(j\omega) = \sum_{\tau=-\infty}^{\infty} R_{xx}(\tau) e^{-j\omega\tau} \quad (3.22)$$

$$\phi_{xx}(j\omega) = \{ Z[R_{xx}(\tau)] \}_{z=e^{j\omega}} \quad (3.23)$$

donde Z es la transformación $-z$ bilateral.

Análogamente si $x(t)$, $y(t)$ en tiempo discreto son conjuntamente estacionarios, entonces:

$$\phi_{xy}(j\omega) = \sum_{\tau=-\infty}^{\infty} R_{xy}(\tau) e^{-j\omega\tau} \quad (3.24)$$

$$\phi_{xy}(j\omega) = \{ Z[R_{xy}(\tau)] \}_{z=e^{j\omega}} \quad (3.25)$$

La justificación de la denominación "densidad espectral" está vinculada a la propiedad (3.74).

Una discusión más amplia puede encontrarse en las referencias (25), (27) y (40).

3.2 Ruido blanco y aproximaciones

3.2.1 Definición del ruido blanco

Llamaremos ruido blanco a cualquier proceso estacionario en sentido amplio, tal que su densidad espectral sea constante.

Es decir, si $x(t)$ es un proceso estacionario con

$$\phi_{xx}(j\omega) = c \quad (\forall \omega) \quad (3.26)$$

entonces

$$x(t) = \text{Ruido blanco}$$

Caso discreto

Si $x(t)$ es un proceso estacionario en tiempo discreto ($t \in \mathbb{Z}$), calculamos la autocorrelación de $x(t)$ como la transformada inversa de Fourier de la densidad espectral $\phi_{xx}(j\omega)$.

$$R_{xx}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_{xx}(j\omega) e^{j\omega k} d\omega = \frac{c}{k\pi} \text{sen } k\pi \quad (3.27)$$

$$\therefore R_{xx}(k) = \begin{cases} c & k=0 \\ 0 & k \neq 0 \end{cases} \quad (3.28)$$

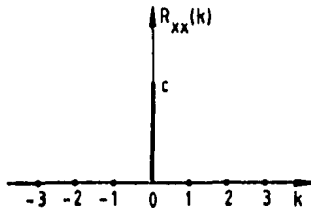


Figura (3.1)

Esto es que los valores del proceso $x(t)$ en distintos tiempos, son variables aleatorias no correlacionadas entre sí.

El inverso también vale. Si $x(t)$ es no correlacionado, es decir que $R_{xx}(k)$ tiene la forma (3.28), entonces $\phi_{xx}(j\omega)$ es de la forma (3.26), luego es ruido blanco.

Ruido blanco en tiempo discreto consiste por lo tanto de un proceso estacionario formado por una sucesión de variables aleatorias no correlacionadas.

Debe observarse que si $x(t)$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas, es un proceso de ruido blanco. El inverso no es cierto.

Caso continuo

Calculamos de la misma forma que antes la función de autocorrelación de $x(t)$ si $t \in \mathbb{R}$

$$R_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_{xx}(j\omega) e^{j\omega\tau} d\omega = c \delta(\tau) \quad (3.29)$$

Análogamente al caso discreto, la señal $x(t)$ no está correlacionada consigo misma para $\tau \neq 0$ (instantes distintos del tiempo). Aparece un problema: su varianza no es finita, por lo tanto no sólo es una señal de energía no finita sino también de potencia no finita (en analogía al caso eléctrico).

En efecto:

$$\text{Var}(x(t)) = R_{xx}(0) = c \delta(t) \Big|_{t=0} \quad (3.30)$$

Este problema surge de la condición exigida de espectro constante para un ancho de banda no finito.

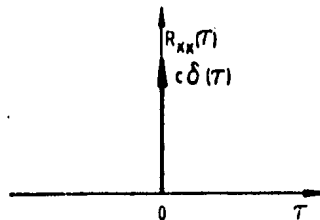


Figura (3.2)

Más aún, se puede demostrar (ver (42)) que si se exige que un proceso sea continuo (en el sentido de media cuadrática), de varianza acotada, media cero y de valores no correlacionados es decir

$$E\{x(t) \cdot x(s)\} = 0 \quad \forall t \neq s,$$

entonces $E\{x^2(t)\} = 0$

Es decir, es un proceso nulo con probabilidad 1 (casi seguramente).

Veremos cómo lograr buenas aproximaciones al ruido blanco en tiempo continuo.

3.2.2 Ruido continuo de banda limitada

Una forma de evitar el problema del párrafo anterior (varianza no acotada) consiste en construir un proceso de espectro constante c , pero en un intervalo finito:

$$\phi_{xx}(f\omega) = \begin{cases} c & \text{si } |\omega| \leq \Omega \\ 0 & \text{si } |\omega| > \Omega \end{cases} \quad (3.31)$$

Calculando la autocorrelación:

$$R_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c e^{j\omega\tau} d\omega = \frac{c}{2\pi} \int_{-\Omega}^{\Omega} e^{j\omega\tau} d\omega$$

$$R_{xx}(\tau) = \frac{c}{\pi\tau} \operatorname{senn} \Omega\tau \quad (3.32)$$

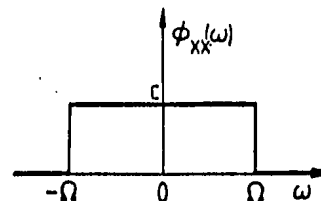


Figura (3.3)

Se puede observar que

- Eligiendo Ω lo suficientemente grande puede hacerse que la correlación entre $x(t)$ y $x(s)$, $t \neq s$ sea arbitrariamente pequeña.

Sin embargo, dado $|t-s|$ no existe valor de Ω acotado, tal que dicha correlación sea nula.

- Si $\Omega \rightarrow \infty$, se puede demostrar en forma sencilla, que la autocorrelación tiende a la del ruido blanco:

$$\lim_{\Omega \rightarrow \infty} R_{xx}(\Omega, \tau) = c \delta(\tau)$$

$$\Omega \rightarrow \infty$$

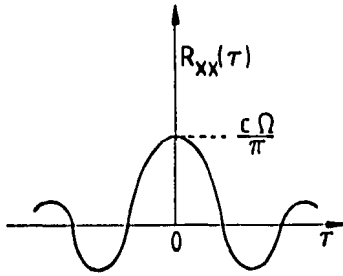


figura (3.4)

Veamos otra aproximación a ruido blanco continuo, con banda limitada.

Una forma de acotar el intervalo de constancia de ϕ_{xx} es considerar un proceso con densidad espectral (3.33)

$$\phi_{xx}(j\omega) = c \frac{\Omega^2}{\omega^2 + \Omega^2} \quad (3.33)$$

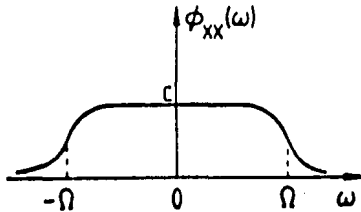


figura (3.5)

La función de autocorrelación correspondiente es:

$$R_{xx}(\tau) = c \frac{\Omega}{\pi} e^{-\Omega|\tau|}$$

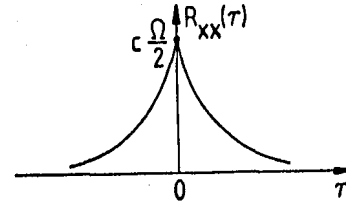


Figura (3.6)

Obviamente para esta aproximación al ruido blanco valen las mismas observaciones que para la aproximación anterior.

Aparte de dichas observaciones, esta aproximación tiene sentido físico, especialmente en circuitos eléctricos.

Considérese un proceso de ruido blanco generado por una fuente de voltaje con una f.e.m. $e(t)$, y un circuito como el de la fig (3.7), donde R y C podrían representar resistencias parásitas y capacidades parásitas (que influyen en el comportamiento en altas frecuencias).

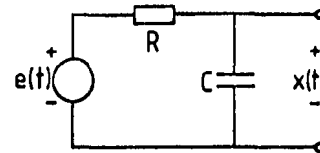


Figura (3.7)

llamando $x(t)$ al voltaje de salida,

$$RC \dot{x}(t) + x(t) = e(t)$$

tomando transformadas de Laplace:

$$X(s) = \left[\frac{1}{sRC + 1} \right] E(s) \quad (3.34)$$

Como $e(t)$ es ruido blanco, $\phi_{ee}(j\omega) = k$, usando la propiedad (3.7#)

$$\phi_{xx}(j\omega) = \left[\frac{1}{1 - j\omega RC} \cdot \frac{1}{1 + j\omega RC} \right] \phi_{ee}(j\omega)$$

$$\phi_{xx}(j\omega) = \frac{1}{1 + \omega^2 R^2 C^2} k = \frac{(1/RC)^2}{(1/RC)^2 + \omega^2} k$$

$$\therefore \phi_{xx}(j\omega) = \frac{(1/RC)^2}{(1/RC)^2 + \omega^2} k \quad (3.36)$$

Compárese con (3.33). Finalmente, sucede que por muy buenos que sean los simuladores eléctricos de ruido blanco, no se pueden despreciar los efectos de resistencia de conductores o capacidades parásitas por proximidad. Las densidades espectrales pueden considerarse (a primer aproximación) como del tipo (3.36).

3.2.3 Señal binaria aleatoria e intervalos discretos (S.B.A.)

Consideramos una señal $x(t)$ que toma valores $\pm V$ durante in-

tervalos de duración λ , tal que en cada punto $\lambda, 2\lambda, 3\lambda, \dots, k\lambda, \dots$ cambia de signo con probabilidad 1/2. Es obvio que:

$$\begin{cases} E\{x(t)\} = 0 \\ R_{xx}(0) = V^2 \end{cases}$$

Se puede demostrar (ver (25)) que su función de autocorrelación es:

$$R_{xx}(\tau) = \begin{cases} (1 - |\frac{\tau}{\lambda}|) V^2 & \text{si } |\tau| < \lambda \\ 0 & \text{si } |\tau| \geq \lambda \end{cases} \quad (3.37)$$

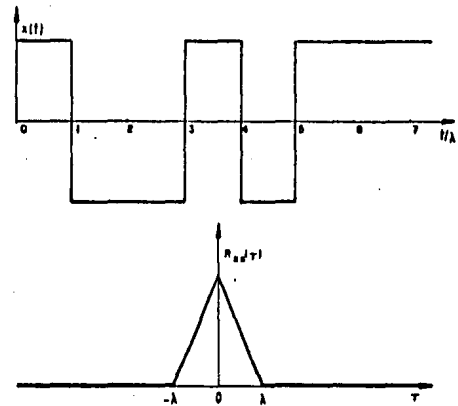


Figura (3.8)

Esta señal tiende a no ser correlacionada en tiempos distintos si $\lambda \neq 0$, y su varianza se conserva finita, siempre es

$$\text{Var}\{x(t)\} = V^2$$

3.2.4 Señales binarias pseudo-aleatorias (SBS o PRBS)

Consideremos una señal que tenga las siguientes propiedades:

- i) Toma valores $\pm V$, cambiando de uno a otro en intervalos regulares $t=0, \lambda, 2\lambda, \dots$ en forma determinística, es decir, siguiendo cierta ley, repetible.
- ii) La señal es periódica de período $N\lambda$, con N entero impar.
- iii) Cada período contiene $\frac{1}{2}(N+1)$ intervalos de un nivel, y $\frac{1}{2}(N-1)$ intervalos del otro con una función de autocorrelación, como la que se muestra en la figura.

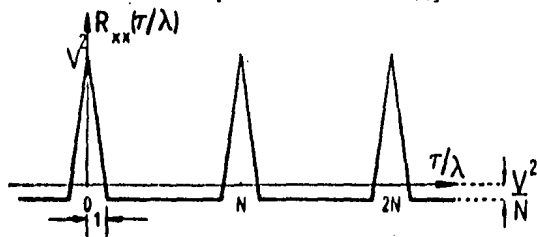


Figura (3.9)

Secuencias PRBS de máxima longitud

Son señales p.r.b.a. fácilmente generables con registros de corrimiento.

Haciendo una convención que asocie una variable binaria $y(t)$ al valor de $x(t)$, p. ej.: $y=0 \rightarrow x=V$, $y=1 \rightarrow x=-V$, estas secuencias satisfacen una ecuación en diferencias lineal (módulo 2), como sigue:

$$D^m y \oplus D^{m-1} y \oplus \dots \oplus D y \oplus y = V \quad (3.38)$$

donde el símbolo \oplus implica suma módulo 2 (o OR-EXCLUSIVO), y V es una constante binaria: 0 ó 1.

Reescribiendo (3.38)

$$(D^m \oplus \dots \oplus I)y = V \quad (3.39)$$

donde I es el operador identidad.

Si $V \neq 0$, una secuencia de orden m se llama secuencia nula, y se puede ver que es periódica. El número máximo de elementos en una secuencia nula es $2^m - 1$, y en ese caso se llama SECUENCIA NULA DE LONGITUD MÁXIMA. O sea que $N = 2^m - 1$. (MLNS).

Una ecuación del tipo (3.39) que genere una secuencia nula de longitud máxima no puede tener un polinomio $(D^m \oplus \dots \oplus I)$ que sea producto de dos polinomios de orden menor, debe ser irreductible. Tampoco puede ser un factor módulo 2 de $D^m \oplus I$ con $n < 2^m - 1$.

Por ejemplo: Una secuencia nula de longitud máxima para $m=6$, puede ser generada por $(D^6 \oplus D^5 \oplus I)y = 0$, o sea:

$y = (D^6 \oplus D^5)y$, que se implementa como se muestra en la figura.

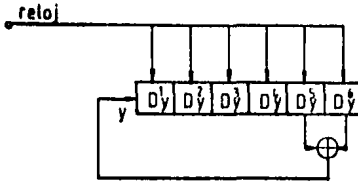


Figura (3.10)

Se presenta una tabla de cómo construir estas secuencias para distintas longitudes.

m (no. de celdas del registro)	$N=2^{m-1}$ (longitud de la secuencia)	Polinomio generador
2	3	$(D^2 \oplus D)y = y$
3	7	$(D^3 \oplus D)y = y$
4	15	$(D^4 \oplus D^3)y = y$
5	31	$(D^5 \oplus D^3)y = y$
6	63	$(D^6 \oplus D^5)y = y$
7	127	$(D^7 \oplus D^4)y = y$
8	255	$(D^8 \oplus D^4 \oplus D^3 \oplus D^2)y = y$
9	511	$(D^9 \oplus D^5)y = y$
10	1023	$(D^{10} \oplus D^7)y = y$
11	2047	$(D^{11} \oplus D^9)y = y$

La ventaja de las secuencias MLNS radica en la facilidad de su generación tanto por Hardware como por Software, pero debe

señalarse que no es la única forma de obtener secuencias binarias pseudo-aleatorias.

A modo de ejemplo, el código en PASCAL con el que fue generada la secuencia PRBS de longitud 63 usada para las pruebas es:

```

PROGRAM PRBS(INTEGER,INTEGER);
(*****
(* GENERA UNA SEÑAL PRBS/MLNS DE LONGITUD MAXIMA(MLNS) DE PERIODO 63 *)
(* DATOSIL=NUMERO TOTAL DE MUESTRAS A GENERAR *)
(* RESULTADOS EN ARCHIVO PRBS/SECUCIAL *)
(*****
VAR REGCARRAYLO: ARRAY OF INTEGER;
    I,J,L: INTEGER;
    PRBSFILE: OF INTEGER;
BEGIN N:=WRITE(PRBS);
      WRITELN('CUANTAS MUESTRAS ?')(READ(L));
      FOR I:=1 TO 6 DO REGCJ:=1;
      FOR J:=1 TO L DO
        BEGIN REGCJ:=(REGCJ#REGCJ63) MOD 3;
          FOR K:=6 DOWN TO 1 DO REGCJ:=REGCJ-K;
            PRBS[I]:=2*REGCJ-1;
              PUT(PRBS)
        END;
      LOCK(PRBS)
    END.
  
```

Este programa fue corrido, generando una secuencia de longitud 63 que se lista y se grafica a continuación:

-1	-1	-1	-1	-1	1	-1	-1
-1	-1	1	1	-1	-1	-1	1
-1	1	-1	-1	1	1	1	1
-1	1	-1	-1	-1	1	1	1
-1	-1	1	-1	-1	1	-1	1
1	-1	1	1	1	-1	1	1
-1	-1	1	1	-1	1	-1	1
-1	1	1	1	1	0	1	1

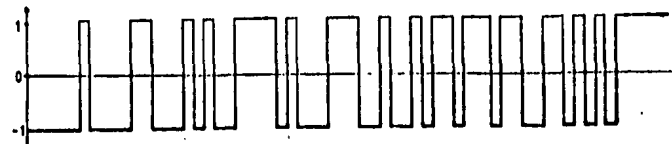


Figura (3.11)

Observaciones:

Las ventajas de las secuencias PRBS son varias:

- . Si bien son determinísticas, su función de autocorrelación aproxima muy bien la función de autocorrelación del ruido blanco. De todas las señales binarias periódicas son las que mejor aproximan al ruido blanco.
- . Por ser determinísticas son repetibles lo que nos da facilidades en la experimentación y comparación de resultados puesto que podemos repetir estas señales en distintas pruebas.
- . Como vimos, son de generación muy sencilla y rápida, tanto por software como por hardware.
- . En las aplicaciones debe elegirse el período de repetición $N=2^m-1$ con cuidado, de forma que su naturaleza periódica no afecte los resultados.

3.2.5 Señales de excitación persistente

Se dice que una señal $u(t)$, discreta $\{t=1,2,\dots\}$ es de excitación persistente de orden n , si se cumple que:

- $$\left\{ \begin{array}{l} \text{i) existe el } \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t), \text{ y es finito} \\ \text{ii) existen los: } \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t)u(t+\tau) = R_{uu}(\tau) \quad (3.40) \\ \tau=0, \dots, n-1 \text{ y son finitos} \\ \text{iii) la matriz } A=(a_{ij}) \text{ con } a_{ij}=R_{uu}[i-j] \text{ es positiva} \\ \text{definida. } i, j=1, 2, \dots, n \end{array} \right.$$

Pueden demostrarse varias propiedades de este tipo de señales muy útiles en los procesos de identificación:

- . Adelantándonos al capítulo 4, demostraremos que no existe ningún filtro de respuesta a impulso finita de orden $n-1$ tal que excitado con una señal de excitación persistente de orden n su salida sea de varianza nula; en efecto, por (4.23), si y es la salida del filtro:

$$R_{yy}(0) = \sum_{k=0}^{n-1} \sum_{\ell=0}^{n-1} h(k)h(\ell)R_{uu}(k-\ell)$$

luego por la condición iii) $\rightarrow R_{yy}(0) > 0$.

- . Puede demostrarse también que una señal de excitación persistente de orden n , debe contener en el caso determinístico, al menos n frecuencias distintas entre 0 y π .

Ejemplos de señales de excitación persistente

- . Es inmediato demostrar que una señal de ruido blanco (con cualquier densidad de distribución) es persistentemente excitante de orden n , para cualquier n finito.
- . También se demuestra que cualquier sucesión binaria pseudoaleatoria de longitud máxima n es de excitación persistente de orden n . Ver por ejemplo (4).

3.3 Simulación de ruido blanco discreto

3.3.1 Métodos de simulación de ruido blanco discreto

Simularemos procesos estocásticos estacionarios en tiempo discreto $x(i)$, $i=1,2,\dots$, con $x(i)$ no correlacionada con $x(j)$, ($i \neq j$) y las variables $x(i)$ idénticamente distribuidas, con media cero y varianza finita.

Se simularán procesos con distribución normal, uniforme y de Bernoulli (PRBS).

Simulación de distribución uniforme

Hay varios métodos de simulación de distribuciones uniformes, anteriormente fue muy usado el método de J. Von Newman de "centro de cuadrados", desplazado hoy en día por el método de generadores congruenciales, especialmente aptos para computadoras digitales.

El método consiste en generar recursivamente una sucesión de números R_n de la forma:

$$R_{n+1} = [\lambda R_n + \mu] \text{mód } P \quad (3.41)$$

donde $0 \leq R_n < P$, R_0 es un valor inicial (semilla), los números λ , μ son constantes enteras y el número P es generalmente el máximo entero representable en la máquina. En una computa-

dora binaria $P=2^b$

La sucesión normalizada $R_n^* = R_n/P$

aproxima enormemente una sucesión de v.a. independientes con distribución uniforme en $[0,1]$, con una buena elección de los parámetros μ , λ .

Cuando $\mu \neq 0$ se dice que es un "generador multiplicativo", si $\mu = 0$ el generador es "congruencial combinado".

Es claro que la sucesión R_n^* es pseudo aleatoria y periódica, obviamente el período máximo alcanzable es P , este se alcanza si se cumplen tres condiciones:

- i) El máximo común divisor de $\{\mu, P\} = 1$
- ii) $\lambda \neq 1 \pmod{q}$ para todo factor primo q de P .
- iii) $\lambda \neq 1 \pmod{4}$ si P es divisible por 4

En una computadora con representación binaria $P=2^b$, y las condiciones (i), (ii), (iii) se satisfacen si λ es de la forma:

$$\lambda = 4u + 1 \quad (3.42)$$

y además μ es impar.

Para la simulación de ruido blanco con distribución uniforme, hemos aprovechado el hecho de que la computadora utilizada para las simulaciones (Burroughs 6800) tiene un intrínseco:

RANDOM (S), que produce una distribución uniforme con un generador congruencial combinado:

$$R_{n+1} = (152587890725xR_n + 116177073375) \bmod 2^{39} \quad \dots (3.43)$$

El dato inicial es S (semilla, o R_0), y en cada llamada al intrínseco se produce un nuevo valor de R_n^a y se actualiza $S \cdot R_{n+1}$ para la siguiente llamada. Es la forma de generación de uniformes que se utiliza en el programa ALEATORIO, escrito en Pascal, y que genera ruido blanco con media cero, uniforme en $[-0.5, 0.5]$ o normal $(0, 1)$, ver apéndice D.

Puede verificarse que el generador congruencial combinado RAN DOH verifica la condición (3.42), y μ impar, por lo que su período es máximo. En la B-6800 es $P=2^{39}$.

Simulación de distribución normal

El método que usaremos fue propuesto por Box & Müller (1958) (11). Consiste en generar una sucesión $x(t)$, $t=1, 2, \dots$ de variables aleatorias independientes con distribución normal $N(0, 1)$ a partir de una sucesión de variables aleatorias independientes con distribución uniforme en $(0, 1)$.

Si U, V son dos variables aleatorias independientes con distribución uniforme en $(0, 1)$, entonces X, Y definidas como:

$$\begin{cases} X = (-2 L_n U)^{1/2} \cos 2\pi V \\ Y = (-2 L_n U)^{1/2} \sin 2\pi V \end{cases} \quad (3.44)$$

Son variables aleatorias independientes con distribución normal $N(0, 1)$. En efecto, por (3.44) se cumple que:

$$\begin{cases} U = e^{-(x^2 + y^2)/2} \\ V = (1/2\pi)(\arctan y/x) \end{cases} \quad (3.45)$$

De acuerdo con el teorema 2.1 podemos calcular la distribución conjunta de X, Y . La distribución conjunta de U, V es:

$$b_{U,V}(u, v) = I_U(u)I_V(v)$$

donde

$$I_U = \begin{cases} 1 & \text{si } 0 < u < 1 \\ 0 & \text{si } u \leq 0 \text{ o } u \geq 1 \end{cases}$$

$$I_V = \begin{cases} 1 & \text{si } 0 < v < 1 \\ 0 & \text{si } v \leq 0 \text{ o } v \geq 1 \end{cases}$$

La densidad conjunta de X, Y es entonces:

$$b_{X,Y}(x, y) = |J| I_U(u(x, y)) I_V(v(x, y)) = |J| I(x, y) \quad (3.46)$$

donde

$$I(x, y) = 1 \quad \forall x, y, \text{ y además}$$

$$J = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} -e^{-(1/2)(x^2+y^2)} x & -e^{-(1/2)(x^2+y^2)} y \\ -\frac{1}{2\pi} \frac{y}{x^2+y^2} & \frac{1}{2\pi} \frac{x}{x^2+y^2} \end{vmatrix}$$

$$f(x, y) = \frac{1}{2\pi} e^{-1/2(x^2 + y^2)} \quad (3.47)$$

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-(1/2)(x^2 + y^2)} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

... (3.48)

luego las parejas X, Y son variables aleatorias independientes con distribución normal $N(0, 1)$.

El programa ALEATORIO utiliza una sucesión de variables aleatorias uniformes $(0, 1)$, e independientes, para generar por el método de Box & Muller una sucesión de variables aleatorias normales $(0, 1)$ e independientes. Además computa la media muestral y la desviación muestral corregida para realizar las pruebas sobre la media y la varianza descritas en la sección siguiente. Su listado puede encontrarse en el apéndice D.

Simulación de señales binarias

Una forma de simular señales binarias, que adoptan valores de ± 1 , con probabilidades 0.5 y 0.5 respectivamente es tomar una señal con distribución uniforme $(0, 1)$, llamémosla $y(t)$, y generar $x(t)$ como:

$$x(t) = \begin{cases} +1 & \text{si } 0.25 \leq y < .75 \\ -1 & \text{en caso contrario} \end{cases} \quad (3.49)$$

Esta es la forma con que el programa RUIDOBINARIO (Apéndice

D) genera una señal binaria aleatoria a intervalos discretos como se definió en la sección 3.2.3 con $V=1$, $\lambda=1$.

Las señales binarias que más utilizaremos son las sucesiones PRBS de largo máximo.

Generaremos sucesiones de período 63 y 127 de acuerdo a las ecuaciones de la sección 3.2.4 con los programas PRBS63 y PRBS127 respectivamente.

El único dato que necesitan estos programas es la longitud (no. de instantes) de la señal; han sido escritos en Pascal para la computadora B6000 del C.S.C. y su listado se encuentra en el apéndice D.

3.3.2 Pruebas sobre las señales simuladas

Para verificar que las señales simuladas puedan considerarse como realizaciones de procesos estocásticos con ciertas características, se desarrollaron algunas pruebas estadísticas.

Estas mismas pruebas serán usadas junto con las técnicas de identificación para determinar la estructura de los modelos.

a) Pruebas de descorrelación

* Autocorrelación

El problema consiste en verificar que una realización de un

proceso en tiempo discreto corresponda a un proceso estacionario caracterizable como ruido blanco.

Sea $x(t)$, $t=1, 2, 3, \dots, N+k$, una realización. Si se computan las estimaciones de la función de autocorrelación:

$$\hat{r}_{xx}(\tau) = \frac{1}{N} \sum_{i=1}^N x(i)x(i+\tau), \quad \tau = 1, 2, \dots, k \quad (3.50)$$

se espera que para ruido blanco con media cero $\hat{r}_{xx}(\tau)$, $\tau=1, \dots, k$ sea bastante próxima a cero. ¿qué tan pequeña debe ser?

El siguiente resultado permite construir una respuesta a esta pregunta:

TEOREMA 3.1.

Sea $x(t)$, $t=1, 2, \dots$, una realización de ruido blanco con media cero y varianza σ^2 , es decir $r_{xx}(\tau) = 0 \forall \tau \neq 0$, $r_{xx}(0) = \sigma^2$.

Sean las estimaciones de las autocorrelaciones:

$$\hat{r}_{xx}(\tau) = \frac{1}{N} \sum_{i=1}^N x(i)x(i+\tau)$$

entonces:

$$\hat{r}_{xx}(\tau) = \frac{\hat{r}_{xx}(\tau)}{\hat{r}_{xx}(0)} \xrightarrow{\text{en ley}} \mathcal{N}(0, \frac{1}{N}) \text{ para } N \rightarrow \infty \text{ y } \tau \neq 0. \quad (3.51)$$

Demostración:

De acuerdo al teorema B.9 del apéndice B

$$\hat{r}_{xx}(\tau) \xrightarrow{\text{en ley}} \mathcal{N}(0, \frac{\sigma^2}{N}) \text{ para } N \rightarrow \infty, \tau \neq 0$$

Por otro lado $\hat{r}_{xx}(0) = \frac{1}{N} \sum_{i=1}^N x(i)^2$ es tal que $E\{x(i)^2\} = \sigma^2 < \infty$

entonces, por el teorema B.4 del apéndice B (Khinchine, Ley débil de los grandes números), se cumple que:

$$\hat{r}_{xx}(0) \xrightarrow{\text{Prob.}} \sigma^2$$

Entonces la variable aleatoria $\hat{r}_{xx}(\tau)/\hat{r}_{xx}(0)$ está en las condiciones del teorema B.1, parte 2) del apéndice B, si $\sigma \neq 0$ y resulta:

$$\hat{r}_{xx}(\tau) = \frac{\hat{r}_{xx}(\tau)}{\hat{r}_{xx}(0)} \xrightarrow{\text{en ley}} \mathcal{N}(0, \frac{1}{N}) \text{ para } N \rightarrow \infty, \sigma \neq 0, \tau \neq 0$$

que es precisamente lo que queríamos demostrar.

Hemos visto que las variables aleatorias $\hat{r}_{xx}(\tau)$, $\tau=1, 2, \dots$, tienden a distribuciones normales para $N \rightarrow \infty$. Veremos que éstas son además independientes.

TEOREMA 3.2

Sea la sucesión $x(t)$, $t=1, 2, \dots$ como en el teorema anterior, entonces para $N \rightarrow \infty$ se cumple que $E\{\hat{r}_{xx}(\tau_1) \cdot \hat{r}_{xx}(\tau_2)\} \rightarrow 0$ para

$\tau_1 \neq \tau_2$ y $\tau_1, \tau_2 \neq 0$.

En efecto:

La variable aleatoria $\hat{\lambda}_{xx}(\tau_1)\hat{\lambda}_{xx}(\tau_2) = \frac{\hat{R}_{xx}(\tau_1)\hat{R}_{xx}(\tau_2)}{\hat{R}_{xx}(0)^2}$, es

cociente de dos variables aleatorias. Supongamos que

$\hat{R}_{xx}(\tau_1), \hat{R}_{xx}(\tau_2)$ converge a alguna variable aleatoria V con distribución desconocida F , como $\hat{R}_{xx}(0) \xrightarrow{\text{prob.}} \sigma^2$, por el teorema B.2 del apéndice B, $\hat{R}_{xx}(0)^2 \rightarrow \sigma^4$ y se cumple por el teorema B.1 del apéndice B que:

$$\hat{\lambda}_{xx}(\tau_1)\hat{\lambda}_{xx}(\tau_2) \xrightarrow{\text{ley}} \frac{V}{\sigma^2} \quad (3.52)$$

$$\text{entonces } \lim_{N \rightarrow \infty} E\{\hat{\lambda}_{xx}(\tau_1)\hat{\lambda}_{xx}(\tau_2)\} = \frac{E(V)}{\sigma^2} \quad (3.53)$$

sólo queda demostrar que $E(V)=0$, esto se cumple pues recordando que V es la variable aleatoria a la que converge el producto $\hat{R}_{xx}(\tau_1)\hat{R}_{xx}(\tau_2)$:

$$E\{\hat{R}_{xx}(\tau_1)\hat{R}_{xx}(\tau_2)\} = \frac{1}{N^2} E\left\{\sum_{i=1}^N x(i)x(i+\tau_1) \sum_{j=1}^N x(j)x(j+\tau_2)\right\} =$$

$$= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E\{x(i)x(j)x(i+\tau_1)x(j+\tau_2)\} \begin{cases} 0 & \text{si } \tau_1 \neq \tau_2 \\ \sigma^4 & \text{si } \tau_1 = \tau_2 \end{cases}$$

El último caso era previsible (ver teorema anterior), pero el que interesa es $\tau_1 \neq \tau_2$.

Siendo $E\{\hat{R}_{xx}(\tau_1)\hat{R}_{xx}(\tau_2)\} = 0$ si $\tau_1 \neq \tau_2$, una constante independiente de N , también $E(V) = 0$ para $\tau_1 \neq \tau_2$, entonces:

$$\lim_{N \rightarrow \infty} E\{\hat{\lambda}_{xx}(\tau_1)\hat{\lambda}_{xx}(\tau_2)\} = \frac{E(V)}{\sigma^2} = \begin{cases} 0 & \text{si } \tau_1 \neq \tau_2 \\ \frac{1}{N} & \text{si } \tau_1 = \tau_2 \end{cases} \quad (3.54)$$

Corolario.— Como $\hat{\lambda}_{xx}(\tau)$, $\tau=1, 2, \dots$ tienden a estar normalmente distribuidas, y con covarianza nula, entonces tienden a distribuciones independientes.

La prueba: Basados en los resultados anteriores podemos proponer unas pruebas de descorrelación:

- i) Bajo la hipótesis nula: $x(k)$, $k=1, 2, \dots$ son no correlacionadas (y con media=0), vemos que las estimaciones $\hat{\lambda}_{xx}(\tau)$ se distribuyen $\mathcal{N}(0, \frac{1}{N})$ para $N \rightarrow \infty$, y son independientes.

Si se computan $\hat{\lambda}_{xx}(\tau)$, $\tau=1, 2, \dots, k$ (usaremos normalmente $k=10$),

$$P\{|\hat{\lambda}_{xx}(\tau)| \leq 2.8 \frac{1}{\sqrt{N}}\} = .99488$$

entonces

$$P\{|\hat{\lambda}_{xx}(1)| \leq 2.8 \frac{1}{\sqrt{N}}, \dots, |\hat{\lambda}_{xx}(10)| \leq 2.8 \frac{1}{\sqrt{N}}\} = (.99488)^{10} =$$

$$= .95 \quad (3.55)$$

de donde se desprende que con un nivel de confianza de

95%, todos los $|\hat{\lambda}_{xx}(\tau)|, \tau=1 \dots 10$ deben ser menores que $1.8 \frac{1}{\sqrt{N}}$.

- 11) Como $\hat{\lambda}_{xx}(\tau), \tau=1 \dots k$, son independientes, normales, de varianza $\frac{1}{N}$ (teórica), computamos la estimación de su varianza como

$$\Delta^2 = \frac{1}{k} \sum_{i=1}^k |\hat{\lambda}_{xx}(i)|^2 \quad (3.56)$$

es evidente que la cantidad χ :

$$\chi = \frac{k \Delta^2}{1/N} \quad (3.57)$$

Se distribuye como χ^2 (ji-cuadrada) con k grados de libertad (para $N \rightarrow \infty$).

Probaremos la hipótesis nula de que la varianza de $\hat{\lambda}_{xx}(\tau)$ es $1/N$ contra la alternativa $\text{var}(\hat{\lambda}_{xx}(\tau)) > 1/N$. Entonces, con un nivel de confianza de $P\%$, bajo la hipótesis nula, se cumple que:

$$\Delta^2 \leq \frac{k_p}{N} \quad (3.58)$$

y para la desviación cuadrática media:

$$\sqrt{\Delta^2} \leq \sqrt{k_p} / \sqrt{N} \quad (3.59)$$

donde la constante k_p es:

$$k_p = \frac{\chi^2_{k,p}}{k}$$

con $\chi^2_{k,p}$ el punto de probabilidad acumulada $P\%$ para la

distribución χ^2_k (ji-cuadrada con k grados de libertad). Se tabulan valores de la constante k_p para distintos niveles de confianza (P), y $k = 10$ (que es lo que usaremos normalmente).

TABLA 3.1

P	k_p	$\sqrt{k_p}$
95%	1.8307	1.35303
97.5%	2.0483	1.43119
99%	2.3209	1.52345

Para ejecutar ambas pruebas, se escribió el programa CORRELACION, en Pascal para la computadora Burroughs-6800 del C.S.C. Dicho programa toma como datos: un archivo de señales a probar, el número N de productos para la estimación de las autocorrelaciones, y el número k de estas. Computa y despliega las cantidades $\hat{R}_{xx}(\tau), \hat{\lambda}_{xx}(\tau), \tau = 1 \dots k$. Saca además un archivo para ser graficado con GRAFO/PASCALERO, donde se dibujan las cantidades $\hat{\lambda}_{xx}(\tau)$ contra el parámetro τ , trazándose las líneas de desviación cuadrática media teórica $\pm 1/\sqrt{N}$.

La figura (3.12) es un resultado típico de dicho programa.



Figura (3.12)

• Cruces por cero. (Se exige independencia)

Sea $x(t)$ $t = 1, \dots, N$ una observación de largo N de una señal estacionaria constituida por variables aleatorias independientes, con media cero y varianza finita. En lo que sigue describiremos efectos de cuantización numérica de la computadora (como lo hemos venido haciendo), que en particular provoca que en una distribución continua $P(x(t) = x) \neq 0$ para x real, por ejemplo cero.

Consideremos la variable aleatoria $y(t)$, $t = 1, \dots, N-1$ definida como:

$$y(t) = \begin{cases} 1 & \text{si } \text{Sg}[x(t)] \neq \text{Sg}[x(t+1)] \\ 0 & \text{si } \text{Sg}[x(t)] = \text{Sg}[x(t+1)] \end{cases} \quad (3.60)$$

como $x(t)$ y $x(t+1)$ son independientes, $y(t)$ tiene una distribución de Bernoulli con $p = 0.5$, cualquiera sea la distribución de $x(t)$.

Se desprende entonces que la suma de cambios de signo: $\sum_{t=1}^{N-1} y(t)$ tiene una distribución binomial $\{N-1, 0.5\}$. Esta distribución es simétrica respecto a la media $(N-1)p = (N-1)/2$, entonces bajo la hipótesis nula: " $x(t)$, $t = 1 \dots N$ son variables aleatorias independientes con media cero".

$$\left| w_N - 0.5 \right| < \frac{K_{97.5}}{N-1} \quad \text{con probabilidad } 0.95 \quad (3.61)$$

Donde $w_N = \frac{1}{N-1} \sum_{t=1}^{N-1} y(t)$ es el promedio de cambios de signo; $K_{97.5}$ es el punto de probabilidad acumulada = 97.5% de la distribución binomial $\{N-1, 0.5\}$.

Esto establece un rango de valores para w_N con probabilidad 0.95.

En la tabla 3.II se tabulan los límites inferiores de estos intervalos para distintos valores de $N-1$.

TABLA 3.11

$N-1$	$0.5 - K/N-1$
16	.227
20	.258
30	.305
40	.333
60	.363
100	.396
200	.427
400	.451
600	.460
800	.4654
1000	.469
1500	.4747
2000	.4781

Los valores correspondientes a $N - 1 = 16, \dots, 200$ fueron extraídos de "Biometric Tables for Statisticians", los que corresponden a $N - 1 = 400, \dots, 2000$ se calcularon aproximando con la función de probabilidad normal $N(0.5, \frac{1}{4N})$, a la que tienen de la distribución binomial para $N \rightarrow \infty$.

Para calcular valores intermedios se realizó una interpolación polinomial entre los valores tabulados, considerando abscisas logarítmicas para espaciar uniformemente los datos. La interpolación se hizo con el algoritmo de Neville-Aitken.

Los resultados de la interpolación se presentan en la figura (3.13).

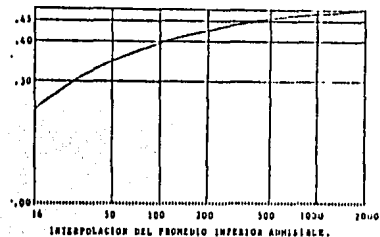


Figura (3.13)

Para efectuar esta prueba se escribió el programa CRUCES/CE-RO que toma como datos un archivo de la señal a probar y el número N de muestras a tomar en cuenta.

Cuenta el número de cambios de signo y compara este número con el rango admisible para una confiabilidad del 95%. El rango admisible lo calcula interpolando polinomialmente la tabla con el algoritmo de Neville-Aitken.

El programa fue escrito en Pascal para la Borroughs-6800 del C.S.C. su listado se encuentra en el apéndice D.

b) Pruebas sobre la distribución

* Media y varianza

Sea $x(t)$, $t = 1, 2, \dots$ una señal de ruido blanco con media cero, varianza σ^2 y momentos de cuarto orden finitos. Entonces, es fácil demostrar por el teorema (B.7) central del límite (ver apéndice B), que:

$$\left\{ \begin{array}{l} \bar{x} = \frac{1}{N} \sum_{i=1}^N x(i) \xrightarrow{\text{dist.}} \mathcal{N}(0, \sigma^2/N) \text{ para } N \rightarrow \infty \\ \sigma^2 = \frac{1}{N} \sum_{i=1}^N x(i)^2 \xrightarrow{\text{dist.}} \mathcal{N}(\sigma^2, \frac{\mu_4 - \sigma^4}{N}) \end{array} \right. \quad (3.62)$$

$$\left\{ \begin{array}{l} \bar{x} = \frac{1}{N} \sum_{i=1}^N x(i) \xrightarrow{\text{dist.}} \mathcal{N}(0, \sigma^2/N) \text{ para } N \rightarrow \infty \\ \sigma^2 = \frac{1}{N} \sum_{i=1}^N x(i)^2 \xrightarrow{\text{dist.}} \mathcal{N}(\sigma^2, \frac{\mu_4 - \sigma^4}{N}) \end{array} \right. \quad (3.63)$$

para $N \rightarrow \infty$

donde $\mu_4 = E\{x(t)^4\}$: Se puede calcular μ_4 para distintas distribuciones, por ejemplo:

$$\left. \begin{array}{l} \text{para normal } (0, \sigma^2) \quad \mu_4 = 3\sigma^4 \\ \text{para uniforme en } (a, b) \end{array} \right\} \quad (3.64)$$

$$\mu_4 = (b-a)^4/80 \quad (3.65)$$

De aquí se desprende que para $N \rightarrow \infty$, conocida la varianza σ^2 , bajo la hipótesis nula $\mu = 0$, debe ser:

$$|\bar{x}| \leq 1.96 \sigma \frac{1}{\sqrt{N}} \quad (3.66)$$

con un nivel de confiabilidad del 95%. Para el mismo nivel de confiabilidad, supuestos $\mu = 0$ y μ_4 , para $N \rightarrow \infty$ bajo la hipótesis nula $E\{x(i)^2\} = \sigma^2$ debe cumplirse:

$$|\hat{\sigma}^2 - \sigma^2| \leq 1.96 \sqrt{\frac{\mu_4 - \sigma^4}{N}} \quad (3.67)$$

En particular realizaremos estas pruebas sobre la distribución normal: $N(0, 1)$, y uniforme en $(-0.5, 0.5)$, y las desigualdades (3.66) y (3.67) quedan, para estos casos como:

$$\text{para la dist. normal } N(0, 1) : |\bar{x}| \leq 1.96/\sqrt{N}$$

$$|\hat{\sigma}^2 - 1| \leq 2.772/\sqrt{N}$$

$$\text{para la dist. uniforme en } (-0.5, 0.5) : |\bar{x}| \leq 0.5658/\sqrt{N}$$

$$|\hat{\sigma}^2 - 0.08333| \leq 0.1461/\sqrt{N}$$

Estas pruebas se incluyen en el programa ALEATORIO, que se menciona más adelante, usado para generar señales de ruido blanco con distribución uniforme o normal.

Observaciones:

En el caso de la distribución normal, la ecuación (3.62) no es asintótica, se cumple $\forall N \geq 1$, y la (3.63) podría ser sustituida en forma exacta con una distribución χ^2 ($j/2$ - cuadrada) con N grados de libertad. Para valores grandes de N la diferencia es mínima, por ejemplo para $N = 200$, los límites de la prueba (3.67) difieren en aproximadamente 4%, para valores de N mayores, la diferencia es menor aún.

Debe observarse además que en la prueba de la media se supone varianza conocida, y vice-versa. Estas pruebas solo han sido propuestas para confirmar que las señales generadas son "buenos representantes" del tipo de señal que se quiere simular. En un caso general hubiera sido más apropiada una prueba conjunta, por ejemplo con el estimador de la ecuación (2.45), y en el caso de distribución normal, recurrir a la prueba t (por la distribución t de Student) para la media, y χ^2 (ji -cuadrada) para la varianza. Las pruebas propuestas son sencillas y ad-hoc a la aplicación propuesta.

• Densidad de distribución - criterio de Pearson

Supongamos que se tiene una variable aleatoria Y , que puede tomar valores discretos y_1, \dots, y_m , con probabilidades p_1, \dots, p_m respectivamente. Dada una muestra aleatoria simple de N observaciones de dicha variable se observan frecuencias de aparición n_1, n_2, \dots, n_m para los valores x_1, \dots, x_m . ¿como podemos confirmar que dicha muestra corresponde a la distribución mencionada?

Se trata pues de determinar cuando la diferencia entre las frecuencias observadas y sus valores esperados Np_1, Np_2, \dots, Np_m es significativa.

Puede demostrarse a partir del teorema de Laplace, que la va-

riable aleatoria

$$\chi = \sum_{i=1}^m \frac{(n_i - Np_i)^2}{Np_i} \quad (3.68)$$

Se distribuye como una χ^2 (ji -cuadrada) con $m-1$ grados de libertad. Es decir

$$\chi \xrightarrow{d.i.t} \chi^2_{m-1} \quad \text{para } N \rightarrow \infty \quad (3.69)$$

Una demostración a partir de resultados de convergencia para grandes muestras puede encontrarse en Rao, C.R. [43].

Sea ahora $x(t)$, $t = 1, \dots, N$ una observación de ruido blanco con varianza finita, si se quiere verificar que $x(t)$ tiene una distribución determinada puede dividirse el rango de variación de $x(t)$ en m intervalos (celdas); si p_i , $i=1, \dots, m$ son las probabilidades asociadas a dichos intervalos, determinadas por la distribución teórica, y n_i , $i=1, \dots, m$ son las frecuencias observadas entonces la cantidad χ definida en (3.68) puede usarse como criterio pues bajo la hipótesis nula de que las probabilidades asociadas a los intervalos son p_1, \dots, p_m , entonces χ se distribuye como una χ^2_{m-1} . Para un nivel de confianza de P se propone como criterio de aceptación de la hipótesis nula:

$$\chi \leq \chi^2_{m-1, P} \quad (3.70)$$

donde $X_{m-1, P}^2$ es el punto de probabilidad acumulada de P para la distribución χ^2 con $m-1$ grados de libertad. En la tabla 3.III se tabulan algunos de estos puntos críticos para valores de $m-1$: 8, 12, 14, y $P = 0.9, 0.95, 0.99$.

TABLA 3.III. Tomado de Pearson & Hartley "Biometrika Tables for Statisticians"

$m-1$	$X_{m-1, 0.9}^2$	$X_{m-1, 0.95}^2$	$X_{m-1, 0.99}^2$
8	13.362	15.507	20.09
12	18.549	21.026	26.217
14	21.064	23.685	29.141

Para efectuar esta prueba se escribió el programa HISTO. El programa fue escrito en Pascal para la computadora Borroughs-6000 del C.S.C, los datos son un archivo con la señal $x(t)$ a probar, el número N de observaciones de dicha señal a tomar en cuenta, el extremo mínimo y máximo de los intervalos y el número de estos.

El programa cuenta el número de valores observados en cada intervalo, y aquellos que caen fuera del rango $[MIN, MAX]$. Crea un archivo, para graficar con GRAFO/PASCALERO, que contiene un histograma de la señal para los intervalos pedidos, y puede incorporar una curva normal para comparar visualmente con el histograma.

Un resultado típico de este programa se puede ver en la figura (3.14).

El listado de dicho programa se encuentra en el apéndice D.

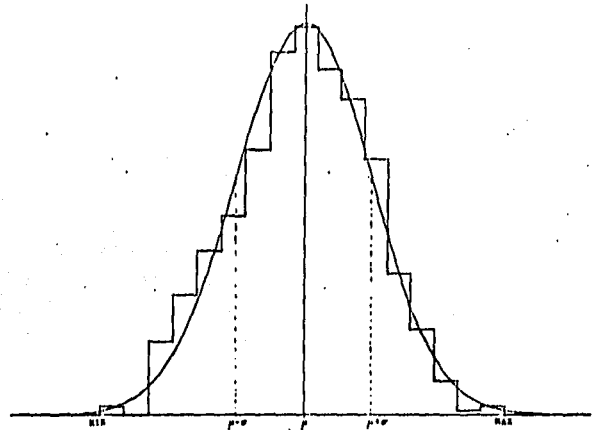


Figura (3.14)

3.3.3 Señales simuladas

Con los métodos de simulación descritos en 3.3.1 se simularon 10 señales de ruido blanco, para usarse en los capítulos 5, 6

y 7, en ejemplos de los algoritmos de identificación con datos simulados.

Las señales simuladas pasaron las pruebas de no-correlación, y distribución vistas en 3.3.2.

A continuación se listan en la tabla 3.IV las características de estas señales.

NOMBRE	TIPO DE DISTRIBUCION	LARGO	METODO DE GENERACION	PROGRAMA GENERADOR
RUIDO 1	Normal (0,1)	645	Box - Muller	ALEATORIO
RUIDO 2	Normal (0,1)	645	Box - Muller	ALEATORIO
RUIDO 3	Normal (0,1)	645	Box - Muller	ALEATORIO
RUIDO 4	Normal (0,1)	645	Box - Muller	ALEATORIO
RUIDO 5	Normal (0,1)	1127	Box - Muller	ALEATORIO
RUIDO 6	Normal (0,1)	310	Box - Muller	ALEATORIO
RUIDO 7	Uniforme en [-0.5,0.5]	645	Congruencial combinado	ALEATORIO
INT 1	Binaria aleatoria a intervalo disc.	310	A partir de uniforme	RUIDO/BINARIO
PRUS 63	Sucesión PRUS, máxima longitud	325	Ecuación en diferencias mód. 2	PRUS/63
PRUS 127	Sucesión PRUS, máxima longitud	1127	Ecuación en diferencias mód. 2	PRUS/127

TABLA 3.IV

En el apéndice C pueden encontrarse gráficos de las señales y resultados de las pruebas estadísticas.

3.4 Identificación no paramétrica de sistemas continuos

En esta sección se describen en líneas generales, algunos métodos clásicos de determinación de la respuesta a impulso y la función de transferencia de sistemas dinámicos continuos, con procesos estocásticos como señales de entrada.

3.4.1 Convolución y correlación

Sea un sistema dinámico lineal, continuo invariante en el tiempo. Si su respuesta a impulso es $h(t)$ y su función de transferencia es $H(s) = \mathcal{L}\{h(t)\}$, se cumple que para una entrada $u(t)$, su salida $y(t)$ es tal que:

$$y(t) = \int_{-\infty}^{+\infty} h(\tau) u(t-\tau) d\tau = \int_0^{\infty} h(\tau) u(t-\tau) d\tau$$

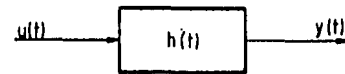


Figura (3.15)

Calculemos ahora la correlación entre la entrada y la salida, suponiendo ergodicidad del proceso:

$$R_{uy}(0) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u(t-0) y(t) dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u(t-0) \left(\int_0^{\infty} h(\tau) u(t-\tau) d\tau \right) dt$$

inviertiendo el orden de integración. (se puede hacer puesto que t y τ son independientes)

$$R_{uy}(0) = \int_0^{\infty} h(\tau) \cdot \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u(t-0) u(t-\tau) dt d\tau = \int_0^{\infty} h(\tau) R_{uu}(0-\tau) d\tau$$

$$\therefore \boxed{R_{uy}(0) = \int_0^{\infty} h(\tau) R_{uu}(0-\tau) d\tau} \quad (3.71)$$

Obsérvese que la relación (3.71) implica que si al sistema se le aplicara una entrada idéntica a la función de autocorrelación de $u(t)$, su salida sería idéntica a la función de correlación entre $u(t)$ y $y(t)$.



Figura (3.16)

Si ahora tomamos transformadas de Laplace a la ecuación (3.71) resulta:

$$\Phi_{uy}(s) = H(s) \Phi_{uu}(s), \text{ evaluando en } s = j\omega$$

$$\Phi_{uy}(j\omega) = H(j\omega) \Phi_{uu}(j\omega) \quad (3.72)$$

$$\Phi_{yu} = \overline{H(j\omega)} \Phi_{uu}(j\omega)$$

análogamente se puede demostrar que

$$\Phi_{yy}(j\omega) = H(j\omega) \Phi_{yu}(j\omega) \quad (3.73)$$

Luego, sustituyendo (3.73) en (3.72):

$$\Phi_{yy} = H(j\omega) \Phi_{yu} = H(j\omega) \cdot \overline{H(j\omega)} \Phi_{uu}$$

\therefore

$$\boxed{\Phi_{yy} = |H(j\omega)|^2 \Phi_{uu}} \quad (3.74)$$

que relaciona las densidades espectrales de la salida y entrada a un sistema lineal.

De las ecuaciones (3.71) a (3.74) se desprenden procedimientos triviales para identificar el sistema. Por ejemplo si $u(t)$ es ruido blanco unitario ($C = 1$), $R_{uy}(\tau)$ será la respuesta a impulso $h(\tau)$, y $\Phi_{uy}(j\omega)$ será $H(j\omega)$.

Pero en la práctica no es posible en general hacer esto, así

que veremos como identificar sistemas sin sacarlos de su operación normal.

3.4.2 Determinación de la respuesta a impulso

Supongamos ahora, con mayor generalidad, al sistema anterior, excitado por una entrada $u(t)$ que es superposición de dos entradas: $u_1(t)$ que representa la entrada en operación normal del sistema, y $u_2(t)$ que representa una "señal de prueba", la que elegiremos posteriormente para identificar la respuesta a impulso. Supondremos que la salida del sistema se corrompe con ruido aditivo $n(t)$. En resumen:

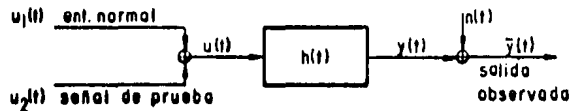


Figura (3.17)

En general se pueden hacer algunas suposiciones adicionales:

Como seleccionaremos $u_2(t)$, podemos elegirla de forma que no esté correlacionada con $u_1(t)$. Debe sin embar-

go tenerse presente que en la identificación en lazo cerrado esto no será posible en general, puesto que $u_1(t)$ dependerá dinámicamente de $\bar{y}(t)$, y por lo tanto también de $u_1(t)$.

Elegiremos $u_2(t)$ independientemente de $n(t)$, y por lo tanto no correlacionados.

En resumen:

$$\left\{ \begin{array}{l} 1) R_{u_1 u_2}(\theta) = 0 \quad \forall \theta \\ 2) R_{u_1 n}(\theta) = 0 \quad \forall \theta \end{array} \right.$$

Como el sistema es lineal, aplicaremos superposición, llamando:

$$\left\{ \begin{array}{l} y_1(t) = \int_0^{\infty} h(\tau) u_1(t - \tau) d\tau \\ y_2(t) = \int_0^{\infty} h(\tau) u_2(t - \tau) d\tau \end{array} \right.$$

$$\bar{y}(t) = y_1(t) + y_2(t) + n(t)$$

Suponiendo ergodicidad calculemos la correlación entre u_2 y \bar{y} .

$$R_{u_2 y}(0) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u_2(t-0) \bar{y}(t) dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u_2(t-0) [y_1(t) + y_2(t) + n(t)] dt$$

$$\therefore R_{u_2 y}(0) = \int_0^{\infty} h(\tau) \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u_2(t-0) u_1(t-\tau) dt d\tau +$$

$$+ \int_0^{\infty} h(\tau) \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u_2(t-0) u_2(t-\tau) dt d\tau +$$

$$+ \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T u_2(t-0) h(t) dt \quad (3.75)$$

$$\therefore R_{u_2 y}(0) = \int_0^{\infty} h(\tau) R_{u_2 u_1}(0-\tau) d\tau + \int_0^{\infty} h(\tau) R_{u_2 u_2}(0-\tau) d\tau + R_{u_2 n}(0)$$

por 1) y 2):

$$R_{u_2 y}(0) = \int_0^{\infty} h(\tau) R_{u_2 u_2}(0-\tau) d\tau$$

Identificación con ruido blanco

Si elegimos como entrada $u_2(t) = b(t)$, ruido blanco, al ser su función de autocorrelación: $R_{bb}(0) = k \delta(0)$

$$\therefore R_{bb}(0) = \int_0^{\infty} h(\tau) k \delta(0-\tau) d\tau = h(0)$$

esto es que la función de autocorrelación coincide con la respuesta a impulso. Un esquema de identificación en línea para

obtener N valores de $h(t)$, podría ser el de la figura (3.18). Donde los promediadores PROME realizan la operación:

$$\frac{1}{t} \int_0^t (\cdot) dt$$

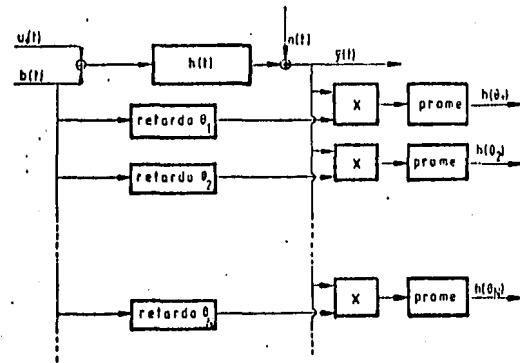


Figura (3.18)

Identificación con PRBS en tiempo finito

En la práctica no tendremos accesible para la identificación una señal $b(t)$ de ruido blanco, ni tampoco la posibilidad de observar y promediar durante un tiempo infinito. De hecho no estaremos entonces calculando las funciones de correlación si no estimándolas.

Reescribiremos la ecuación (3.11), integrando entre 0 y T.

$$\frac{1}{T} \int_0^T u_2(t-\theta) \bar{y}(t) dt = \hat{R}_{u_2 \bar{y}}$$

$$\text{luego } \hat{R}_{u_2 \bar{y}} = \int_0^T h(\tau) \hat{R}_{u_2 u_1}(0-\tau) d\tau + \int_0^T h(\tau) \hat{R}_{u_2 u_2}(0-\tau) d\tau + \hat{R}_{u_2 n}(0)$$

donde hemos supuesto que el sistema tiene respuesta a impulso de duración finita, es decir, existe un tiempo de asentamiento T_s (Settling Time) tal que las entradas ocurridas en un tiempo anterior a $t-T_s$ no afectan, o no influyen en la salida del tiempo t .

$$\begin{cases} \hat{R}_{u_2 u_1}(0-\tau) = \frac{1}{T} \int_0^T u_2(t-\theta) u_1(t-\tau) dt \\ \hat{R}_{u_2 u_2}(0-\tau) = \frac{1}{T} \int_0^T u_2(t-\theta) u_2(t-\tau) dt \\ \hat{R}_{u_2 n}(0) = \frac{1}{T} \int_0^T u_2(t-\theta) n(t) dt \end{cases}$$

Si elegimos T múltiplo del período de la señal p. a. b. s.

$\hat{R}_{u_2 u_2}(0) = \hat{R}_{u_2 u_2}(\theta)$ (ver ec. 3.11), la estimación es exacta.

$$\therefore \hat{R}_{u_2 \bar{y}}(\theta) = \int_0^T h(\tau) \hat{R}_{u_2 u_1}(0-\tau) d\tau + \int_0^T h(\tau) \hat{R}_{u_2 u_2}(0-\tau) d\tau + \hat{R}_{u_2 n}(0) \quad (3.76)$$

$$\hat{R}_{u_2 \bar{y}}(\theta) = \alpha(T, \theta) + \int_0^T h(\tau) R_{u_2 u_2}(0-\tau) d\tau + \beta(T, \theta)$$

$$\text{las cantidades } \alpha(T, \theta) = \int_0^T h(\tau) \hat{R}_{u_2 u_1}(0-\tau) d\tau$$

$$\beta(T, \theta) = \hat{R}_{u_2 n}(0)$$

tienen a cero si $T \rightarrow \infty$; más aún, la ecuación (3.13) nos da la varianza de $\hat{R}_{u_2 u_1}$ y $\hat{R}_{u_2 n}$, en el caso de que fueran señales gaussianas.

$$\text{Var} \{ \hat{R}_{u_2 u_1}(0) \} = \frac{1}{T} \int_{-\infty}^{+\infty} R_{u_2 u_2}(u) R_{u_1 u_1}(u) du$$

$$\text{Var} \{ \hat{R}_{u_2 n}(0) \} = \frac{1}{T} \int_{-\infty}^{+\infty} R_{u_2 u_2}(u) R_{nn}(u) du$$

$$\therefore \hat{R}_{u_2 \bar{y}}(0) = \int_0^T h(\tau) R_{u_2 u_2}(0-\tau) d\tau$$

Siendo u_2 una secuencia PRBS, cuando el período de transición $\lambda \rightarrow 0$ $R_{u_2 u_2}(0) \rightarrow V^2 \lambda \delta(0)$

$$\therefore \text{en este caso } R_{u_2 \bar{y}}(0) = V^2 \lambda h(0)$$

y el efecto de la integración en un período finito, y la operación normal del sistema aparece como un ruido o perturbación que disminuye de amplitud a medida que el período de integración tiende a ∞ .

$$\hat{R}_{u_2 y}(\theta) = V^2 \lambda h(\theta) \quad (3.77)$$

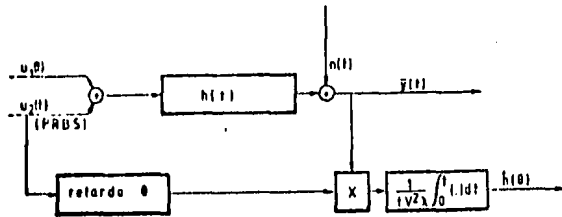


Figura (3.19)

Se puede demostrar además que si $u_1 = 0$, y $u_2(t)$ y $n(t)$ son gaussianas no correlacionadas, ergódicas y con media cero.

Siendo u_2 ruido blanco y $R_{u_2 u_2}(\tau) = R \delta(\tau)$, entonces:

$$\text{var}(\hat{h}(\tau)) = \frac{R_{nn}(0)}{T R} \quad (3.78)$$

Donde T es el período de integración.

Menos visto pues como es posible identificar con ruido blanco y con secuencias PRBS la respuesta a impulso de un sistema dinámico lineal, incluso cuando este opera normalmente.

Solo cabe agregar como consideración práctica de que el perfo

do de la señal PRBS debe ser mayor que el tiempo de asentamiento (T_d), y para considerar el sistema funcionando en estado estacionario (respecto a la señal PRBS), al menos un tiempo T_d se debe dejar pasar luego de aplicar la señal, antes de comenzar la correlación.

Los métodos de correlación aplicados a sistemas discretos se discutirán en la sección 4.2.2.

3.4.3 Respuesta en frecuencia y análisis espectral.

Reescribiremos la ecuación (3.72) para un sistema lineal excitado por un proceso estacionario:

$$\Phi_{uy}(j\omega) = H(j\omega) \Phi_{uu}(j\omega)$$

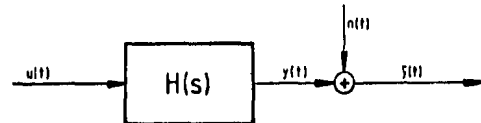


Figura (3.20)

Si la salida del sistema es perturbada con ruido aditivo $n(t)$ el espectro cruzado de u , y es:

$$\phi_{uy} = \phi_{uy} + \phi_{un} \quad \therefore \quad \phi_{uy}(j\omega) = H(j\omega) \phi_{uu}(j\omega) + \phi_{un}(j\omega) \quad (3.79)$$

Si $u(t)$ y $n(t)$ no están correlacionadas, entonces $\phi_{un}(j\omega) = 0$, en cuyo caso podríamos determinar $H(j\omega)$

$$H(j\omega) = \frac{\phi_{uy}(j\omega)}{\phi_{uu}(j\omega)} \quad (3.80)$$

En particular si $u(t)$ es ruido blanco con densidad C (constante), entonces

$$H(j\omega) = \phi_{uy}(j\omega)/C$$

La determinación de la respuesta en frecuencia podría hacerse en forma eficiente con un esquema de transformada Rápida de Fourier (TRF).

Esto es porque puede demostrarse [27] que:

$$\phi_{uy}(j\omega) = U(j\omega) V^*(j\omega)$$

$$\phi_{uu}(j\omega) = U(j\omega) U^*(j\omega)$$

Entonces se procedería primero al cálculo de

$$\begin{cases} V(j\omega) = F(y(t)) \\ U(j\omega) = F(u(t)) \end{cases}$$

$$\text{Luego} \quad H(j\omega) = \frac{\phi_{uy}}{\phi_{uu}}$$

Usando como excitación una señal de tipo PRBS, lo primero que debemos tener en cuenta es su espectro de potencias, que diferirá obviamente del ruido blanco ideal.

Por ejemplo:

Si $x(t)$ es binaria aleatoria a intervalos discretos (sección 3.2.3)

$$\begin{aligned} \phi_{xx}(j\omega) &= 2 \int_0^{\infty} V^2 \left(1 - \frac{\tau}{\lambda}\right) e^{-j\omega\tau} d\tau \\ \phi_{xx}(j\omega) &= V^2 \lambda \frac{\text{Sen}^2(\omega\lambda/\tau)}{(\omega\lambda/\tau)^2} \end{aligned} \quad (3.81)$$

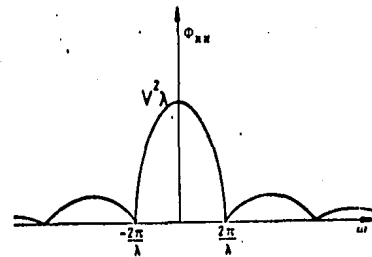


Figura (3.21)

Otra limitación a la clase de modelos radica en la posibilidad de construir para ellos algoritmos robustos de identificación. Por último que en este trabajo no se propone estudiar en forma general la identificación de modelos sino restringirnos a alguna clase particular de ellos.

Nos limitaremos a descripciones de entrada-salida de sistemas dinámicos lineales en tiempo discreto, causales, de orden finito. Los sistemas de

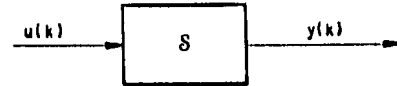


Fig 4.1

esta forma pueden describirse desde el punto de vista de entrada-salida por una ecuación en diferencias de orden finito como la 4.1

$$y(k) + a_1 y(k-1) + \dots + a_n y(k-n) = b_0 u(k) + b_1 u(k-1) + \dots + b_n u(k-n) \quad (4.1)$$

Donde: k es el índice de muestreo $k=0, 1, 2, \dots$

u(k) es la entrada al sistema en el instante k

$$(u(k) \in \mathbb{R} \forall k)$$

y(k) es la salida del sistema en el instante k

$$(y(k) \in \mathbb{R} \forall k)$$

a_i, b_i son en general funciones de k. Si son constantes el sistema es invariante en el tiempo.

Hay varias razones que justifican la elección de esta clase de modelos:

- Una gran cantidad de situaciones reales pueden ser aproximadas de

4. MODELOS PARA LA IDENTIFICACION

4.1 Sistemas lineales discretos

4.1.1 Introducción

El propósito de este capítulo es encontrar una descripción matemática de los sistemas que se identificarán, que incorpore su contexto o medio ambiente, el que supondremos incide sobre los sistemas a través de perturbaciones. De lo que se trata entonces es de caracterizar el sistema y su medio ambiente a través de modelos de la dinámica del sistema y las perturbaciones.

La primera condición que deben cumplir estos modelos es su utilidad. Si se propone una identificación con propósitos de control, los modelos deberán ser adecuados para este propósito.

esta forma.

- Se pueden desarrollar como se verá, algoritmos de identificación robustos con buenas propiedades de convergencia e implementables en computadoras digitales.
- El propio hecho de trabajar con computadoras digitales implica una discretización del proceso aún cuando este sea continuo. En muchos casos los cálculos pueden ser realizados entre instantes de muestreo consiguiendo una implementación en línea.
- En caso que se tengan sistemas con partes continuas y se quiera la evolución de las variables durante el intervalo de muestreo hay técnicas, por ejemplo la transformación-z completa (o "extendida") (1), que pueden resolver el problema.
- El trabajar con señales de ruido blanco en tiempo discreto evita algunos problemas técnicos que surgen en tiempo continuo.
- Este tipo de modelos es adecuado para propósitos de control acompañando técnicas del control clásico y estocástico.
- Aún cuando esta descripción es del tipo una entrada - una salida, los algoritmos de identificación pueden extenderse a representaciones del tipo: varias entradas varias salidas.
- Como veremos inmediatamente pueden hacerse extensiones a alguna clase de sistemas no-lineales

Sistemas intrínsecamente lineales (ver ref (3)).

Si bien la identificación se hará sobre sistemas descriptibles por ecuaciones en diferencias del tipo (4.1), debe observarse que la propiedad fundamental es la linealidad en los parámetros más que en los datos. Más

aún, en una amplia clase de sistemas no lineales una transformación en los datos y en los parámetros habilita una relación lineal entre los nuevos datos y parámetros del tipo (4.1) sobre la que se podrá aplicar una técnica de identificación.

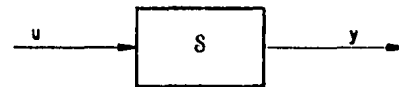


fig 4.2

Veamos un ejemplo: sea el sistema S descriptible de la forma

$$y(k) = \frac{a_2 u(k) + u(k-1)}{a_1 u(k) u(k-1)}$$

con los datos $u(k)$, $y(k)$. Obviamente no es lineal, pero si hacemos

$$\frac{1}{u(k)} = u'(k), \quad \frac{a_2}{a_1} = \beta_1, \quad \frac{1}{a_1} = \beta_2 \text{ entonces (4.2) se transforma en}$$

$$y(k) = \frac{\beta_2}{\beta_1} \frac{1}{u(k-1)} + \frac{1}{\beta_1} \frac{1}{u(k)} = \beta_1 u'(k-1) + \beta_2 u'(k) \quad (4.3)$$

que es del tipo de (4.1) en los parámetros β_1 , y los datos $u'(k)$, $y(k)$.

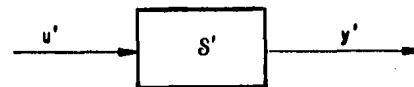


fig 4.3

se obtuvo entonces una transformación $u \rightarrow u'$, y la nueva relación entre u' y y es lineal.

Generalizando un poco: sea un sistema en general de la forma

$$y(k) = f(u(k), \dots, u(k-n), y(k-1), \dots, y(k-n), \theta)$$

donde θ es un vector de m parámetros constantes:

Si existen transformaciones T_1, T_2, T_3 tales que:

- $T_1: R \rightarrow R$ es $y'(k) = T_1(y(k))$ es unívoca.
- $T_2: R \rightarrow R$ es unívoca y $u'(k) = T_2(u(k))$
- $T_3: R^n \rightarrow R^{n+1}$ unívoca y $T_3(\theta) = \{a_1, \dots, a_n, b_0, b_1, \dots, b_n\}$

y se cumple que:

$$y'(k) + a_1 y'(k-1) + \dots + a_n y'(k-n) = b_0 u'(k) + \dots + b_n u'(k-n)$$

entonces diremos que el sistema S' es intrínsecamente lineal. Nótese que la invertibilidad de T_3 aseguraría la relación uno a uno entre los sistemas S y los S' .

De aquí se desprende que las técnicas de identificación para sistemas lineales pueden utilizarse en una amplia clase de sistemas no lineales conociendo las transformaciones apropiadas en los datos y parámetros.

4.1.2 Respuesta a pulso y función de transferencia

No haremos aquí una deducción rigurosa de las representaciones de los sistemas discretos ni las propiedades de la transformada Z , se recordarán algunas definiciones y se introducirá la notación que seguiremos. Un estudio detallado puede verse en las referencias (1), (14), (34).

Consideraremos sistemas S que siguen una ley del tipo (4.1) veremos algunas representaciones equivalentes.

Definiendo la delta de Kronecker $\delta(k) = \begin{cases} 0 & \text{si } k \neq 0 \\ 1 & \text{si } k = 0 \end{cases}$

Llamaremos sucesión de peso o respuesta a pulso a la respuesta del sistema cuando es excitado por una entrada $\delta(k)$. Diferenciar para una entrada $u(k)$ cualquiera es:

$$u(k) = \sum_{s=-\infty}^{\infty} u(s) \delta(k-s) \quad (\text{por definición de } \delta k)$$

$$y(k) = \sum_{s=-\infty}^{\infty} u(s) h(k-s) \quad (\text{por superposición})$$

$$y(k) = \sum_{s=0}^k u(s) h(k-s) \quad (\text{por causalidad, suponiendo además } u(k) \text{ causal, es decir } u(k) = 0 \text{ si } k < 0).$$

Si $u(k)$ no es causal el límite inferior del índice de sumación es $-\infty$.

Haciendo el cambio $k-s = n$ se obtienen dos expresiones para entradas no causales en general:

$$\boxed{y(k) = \sum_{n=-\infty}^k u(n) h(k-n)} \quad \boxed{y(k) = \sum_{n=0}^{\infty} u(k-n) h(n)} \quad (4.4)$$

Esta representación se llama de suma de convolución y es análoga a la integral de convolución para sistemas continuos. Su desventaja práctica radica en que en el cálculo hay que almacenar toda la sucesión $h(k)$, al menos los más significativos, y las entradas anteriores. Para entradas causales el cálculo de $y(k)$ requiere k productos y $k-1$ sumas.

Cuando a partir de algún N natural $h(k) = 0 \quad \forall k > N$, el sistema se denomina comúnmente filtro FIR (de respuesta a impulso finita).

Cuando las condiciones iniciales son nulas es útil la representación del sistema en el campo complejo a través de la transformada Z , que tiene además una cómoda interpretación operacional.

Recordando que la transformada Z (o de Dirichlet) unilateral de una sucesión $h(k)$ es

$$Z\{h(k)\} = \sum_{k=0}^{\infty} h(k)z^{-k} \quad (4.5)$$

Puede verse que para un sistema S con sucesión de peso $h(k)$, al aplicar la propiedad de convolución en el tiempo

$$Y(z) = H(z)U(z) \quad (4.6)$$

$$\text{Donde } \begin{cases} U(z) = Z\{u(k)\} \\ H(z) = Z\{h(k)\} \\ Y(z) = Z\{y(k)\} \end{cases}$$

Se pueden introducir las condiciones iniciales con una modificación de (4.6) pero se pierde la extremada sencillez algebraica de esta relación.

4.1.3 Interpretación operacional de z^{-1}

Cuando el sistema es del tipo que nosotros trataremos es decir que se rige por una ley como (4.1), se puede ver que su función de transferencia

$H(z) = \sum_{k=0}^{\infty} h(k)z^{-k}$ converge, cuando es estable, a una función racional en z en el exterior del círculo unitario.

En efecto, sea el sistema

$$y(k) + a_1 y(k-1) + \dots + a_n y(k-n) = b_0 u(k) + \dots + b_n u(k-n) \quad (4.7)$$

Si $u(k) = y(k) = 0 \quad \forall k < 0$, se puede aplicar la propiedad del corrimiento a la derecha

$$Z\{y(k-1)\} = z^{-1} Y(z) \quad \text{con} \quad Y(z) = Z\{y(k)\}$$

Tomando transformada Z en (4.7):

$$Y(z) (1 + a_1 z^{-1} + \dots + a_n z^{-n}) = U(z) (b_0 + b_1 z^{-1} + \dots + b_n z^{-n}) \quad (4.8)$$

llamando
$$\begin{cases} A(z^{-1}) = 1 + a_1 z^{-1} + \dots + a_n z^{-n} \\ B(z^{-1}) = b_0 + b_1 z^{-1} + \dots + b_n z^{-n} \end{cases}$$

Otro
$$H(z) = \frac{Y(z)}{U(z)} \quad \text{por (4.6)}$$

$$H(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_n z^{-n}}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} \quad (4.9)$$

de otra forma
$$H(z) = \frac{b_0 z^n + \dots + b_n}{z^n + \dots + a_n} = \frac{B'(z)}{A'(z)} \quad (4.10)$$

Definiendo
$$\begin{cases} A'(z) = z^n + a_1 z^{n-1} + \dots + a_n = z^n A(z^{-1}) \\ B'(z) = b_0 z^n + b_1 z^{n-1} + \dots + b_n = z^n B(z^{-1}) \end{cases}$$

La estabilidad del sistema es si y sólo si los polos de $A'(z)$ son interiores a la circulo unitario o lo que es lo mismo, los polos de $A(z)$ son exteriores a él. Consideraremos sólo sistemas estables.

Cuando el sistema tiene condiciones iniciales nulas, la aplicación de la propiedad de corrimiento a la derecha (retraso):

$$Z\{f(k-n)\} = z^{-n} Z\{f(k)\}$$

permite interpretar el factor z^{-1} (ver (4.9)) como un operador de retraso unitario. En este sentido la notación se presta a que haya un poco de

confusión, al mezclar señales en el tiempo con transformadas-z que tienen sentido en el dominio complejo z. Por ejemplo el esquema de la figura (4.4) es claro si queremos representar un bloque que retrasa la señal $x(t)$

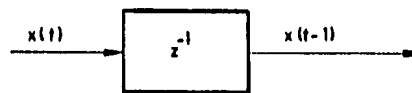


fig 4.4

en una unidad de tiempo, si convenimos en utilizar el símbolo z^{-1} para representar el retraso unitario. Pero no tiene sentido alguno una señal $z^{-1} x(t-1)$ en el plano complejo z.

Para evitar esto se introduce frecuentemente en la literatura el operador q^{-1} de retraso unitario, diferenciándolo de la variable compleja z^{-1} . La figura (4.5) representa el mismo sistema que el de la figura (4.4). Aunque

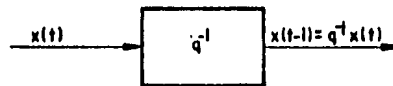


fig 4.5

las expresiones que relacionan las señales en el tiempo y sus transformadas son análogas al manipular algebraicamente los operadores, su interpretación es bien distinta.

$$\begin{cases} Y(z) = H(z) U(z) \\ y(t) = H(q^{-1})u(t) \end{cases} \quad (4.11)$$

En efecto, toda la deducción de (4.9) es válida sustituyendo z^{-1} por q^{-1} , aún cuando las condiciones iniciales no son nulas: la relación (4.11) es más general, y la escribiremos en forma extendida. Para el sistema (4.1) se cumple en el tiempo que:

$$y(t) = \frac{b_0 + b_1 q^{-1} + \dots + b_n q^{-n}}{1 + a_1 q^{-1} + \dots + a_n q^{-n}} u(t) \quad (4.12)$$

4.1.4 Causalidad. Notación

Hemos visto que cualquier sistema de la forma (4.1) se puede representar de la forma (4.9) ó (4.12). Planteemos la pregunta inversa generalizándola: ¿Dada una función racional $H(z)$ de la forma (4.13), es posible realizar un sistema dinámico del tipo (4.1) tal que su función de transferencia sea (4.13)?

$$H(z) = \frac{b'_0 z^m + \dots + b'_m}{z^n + a'_1 z^{n-1} + \dots + a'_n} \quad (4.13)$$

Es trivial deducir que para ello debe ser $m \leq n$, puesto que de lo contrario el sistema debería ser no causal y por lo tanto no realizable.

La forma (4.9), canónica para las descripciones de entrada salida de sistemas discretos de orden n presenta en la literatura referida a identi-

cación algunas variantes. Estas variantes provienen de adecuar los modelos, para hacer explícita o incorporar información que ya se tenga, a priori, sobre la estructura del modelo. Por ejemplo se puede saber que existen retrasos en la respuesta, y se conoce el valor de ese retraso, o puede conocerse que los últimos coeficientes b_j o a_j de la ecuación (4.12) son nulos, reduciendo el número de parámetros que se necesitan para identificar el sistema.

Se emplea muy frecuentemente una forma del tipo:

$$H(z) = z^{-k} \frac{\beta_0 + \beta_1 z^{-1} + \dots + \beta_m z^{-m}}{1 + \alpha_1 z^{-1} + \dots + \alpha_l z^{-l}} \quad (4.14)$$

Es fácil ver que la equivalencia entre (4.14) y (4.9) implica:

$$\begin{cases} n = \max(m + k, l) \\ a_i = \alpha_i \quad 1 \leq i, a_i = 0 \quad i > l \\ b_i = 0 \quad i < k, b_i = \beta_{i-k} \quad k \leq i \leq k + m, b_i = 0 \quad i > k + m \end{cases}$$

En la relación (4.14) destaca un retraso de k unidades de tiempo ($k \geq 0$).

Es conveniente ver que por la propiedad del desplazamiento a la derecha,

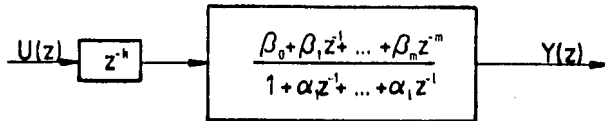


Fig 4.6

de la transformada— τ , en la descripción del sistema por función de transferencia, los retrasos introducen factores racionales en $H(z)$.

Esta es una ventaja, comparando los factores trascendentes que aparecen al describir sistemas en tiempo continuo con transformada de Laplace.

4.2 Modelado de las perturbaciones

En esta sección se estudiará la parte del modelo que corresponde a las perturbaciones estocásticas.

Para esto se estudiará primero cómo responde un sistema discreto del tipo (4.1) a una entrada $u(t)$ aleatoria.

Posteriormente veremos que siempre se puede obtener un proceso estocástico estacionario de espectro racional dado, como resultado de excitar un sistema (4.1) de parámetros adecuados con ruido blanco discreto.

4.2.1 Respuesta de sistemas lineales discretos a procesos estocásticos

Sea un sistema S asintóticamente estable, regido por una ley de recurrencia del tipo (4.1).

Sea $h(k)$, $k=0, 1, 2, \dots$ su respuesta a pulso y $H(z)$ su función de transferencia.

Consideremos dos procesos estocásticos:

- i) $u(t)$ proceso estocástico de segundo orden.
- ii) $v(t)$ proceso estocástico de segundo orden, débilmente estacionario.

Como resultado de excitar S con $u(t)$, $v(t)$, se obtienen las salidas $y_u(t)$, $y_v(t)$ respectivamente.

Estas salidas son procesos estocásticos. Además y_v es estacionaria.

Recordando que:

$$\mu_u(t) = E\{u(t)\} \quad \mu_v = E\{v(t)\} \quad (4.15)$$

$$R_{uu}(t, s) = E\{u(t)u(s)\} \quad R_{vv}(\tau) = E\{v(t)v(t + \tau)\} \quad (4.16)$$

$$R_{uy}(t, s) = E\{u(t)y(s)\} \quad R_{vy}(\tau) = E\{v(t)y(t + \tau)\} \quad (4.17)$$

$$\phi_{vy}(j\omega) = \sum_{n=-\infty}^{\infty} e^{-jn\omega} R_{vy}(n) \quad (4.18)$$

Se pueden demostrar (ver por ejemplo la referencia (4)), las relaciones que se resumen en la siguiente tabla.

Función	caso general, $u(t)$	entrada estacionaria, $v(t)$
Salida	$y_u(t) = \sum_{k=0}^{\infty} h(k) u(t-k)$	$y_v(t) = \sum_{k=0}^{\infty} h(k) v(t-k)$
Media de la Salida	$\mu_y = \sum_{k=0}^{\infty} h(k) \mu_u(t-k)$	$\mu_y = \mu_v \sum_{k=0}^{\infty} h(k)$ $\mu_y = \mu_v H(z)$
Autocorrelación de la Salida	$R_{yy}(s, t) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h(k) h(l) R_{uu}(s-k, t-l)$	$\Phi_{yy}(z) = H(z) \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h(k) h(l) R_{uu}(z^{-k}, z^{-l})$ $\Phi_{yy}(j\omega) = H(e^{j\omega}) H(e^{-j\omega}) \Phi_{uu}(j\omega)$
Correlación entrada-salida	$R_{uy}(s, t) = \sum_{k=0}^{\infty} h(k) R_{uu}(s, t-k)$	$\Phi_{uy}(z) = \sum_{k=0}^{\infty} h(k) R_{uu}(z^{-k}, z)$ $\Phi_{uy}(j\omega) = H(e^{j\omega}) \Phi_{uu}(j\omega)$

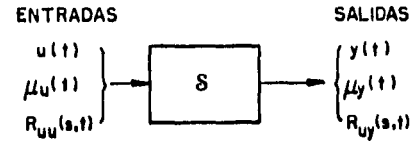


fig 4.7

Debe observarse un hecho importante: la relación entrada-salida (que finalmente es la (4.1), no solo la guardan las señales y_u y y_v sino también las medias: μ_u , μ_y y también las correlaciones R_{uu} , R_{uy} .

Es decir que si fijamos la variable s (en el caso general, puesto que para el caso estacionario es innecesario) entonces $\mu_y(t)$ y $R_{uy}(s, t)$ son las salidas que resultan de aplicar al sistema S las entradas $\mu_u(t)$ y $R_{uu}(s, t)$ respectivamente; ver figura (4.7).

Otros resultados de importancia son (4.24) y (4.27).

La ecuación (4.24) relaciona las densidades espectrales de potencia de los procesos de entrada y salida a través de un factor de ganancia que es precisamente la ganancia de potencia del sistema:

$$A_p = H(e^{-j\omega}) H(e^{j\omega})$$

La ecuación (4.25) relaciona la densidad espectral de potencia de la entrada con la transformada de Fourier de la correlación cruzada de los procesos de entrada y salida, cuando la entrada es estacionaria.

Volvamos en seguida una aplicación de estos resultados.

4.2.2 Una aplicación: Identificación no paramétrica de sistemas discretos por correlación

Las ecuaciones (4.24) y (4.27) muestran la posibilidad de calcular la respuesta a pulso y la función de transferencia respectivamente, de un sistema.

Hasta para ello excitar el sistema considerado con una señal u de ruido blanco.

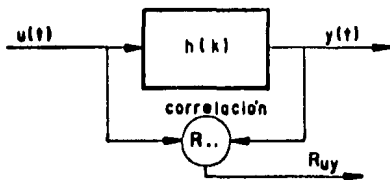


fig 4.8

Su función de correlación es $R_{uu}(\tau) = c\delta(\tau)$ (con la δ de Kronecker).

Su densidad espectral es $\phi_{uu}(j\omega) = c$

por (4.24)

$$R_{uy}(\tau) = \int_{-\infty}^{\infty} h(\beta) c \delta(\tau - \beta) d\beta = c h(\tau)$$

$$\therefore h(\tau) = \frac{R_{uy}(\tau)}{c} \quad (4.28)$$

de esta forma se calcula la sucesión de respuesta a pulso. Para el cálculo de la función de transferencia se sigue de (4.27).

$$\begin{aligned} \phi_{uy}(j\omega) &= H(e^{+j\omega}) c \\ \therefore H(e^{+j\omega}) &= \frac{1}{c} \phi_{uy}(j\omega) \end{aligned} \quad (4.29)$$

Lo cual implica calcular primero la densidad espectral cruzada de u y y .

4.2.3 El teorema de factorización espectral

Hasta aquí se han visto algunos resultados sobre la señal de salida que resulta de excitar un sistema dinámico con un proceso estocástico. Se plantea el problema inverso; ¿puede ser representado un proceso estocástico como salida de un sistema lineal, con una excitación particular?

Supongamos un sistema asintóticamente estable, excitado con ruido blanco, entonces

$$\phi_{yy}(j\omega) = H(e^{-j\omega}) H(e^{+j\omega}) \quad (4.30)$$

Nuestro problema se plantea ahora en encontrar un sistema asintóticamente estable cuya función de transferencia sea tal que se cumpla (4.30), conocida $\phi_{yy}(j\omega)$. Este es el problema de factorización espectral.

TEOREMA 4.1. (Factorización espectral)

Sea un proceso estocástico estacionario de densidad espectral $\phi(j\omega)$, racional en $\cos \omega$, entonces existe una función racional $H(z)$ con todos sus polos interiores al círculo unitario y sus ceros dentro o en el círculo unitario tal que:

$$\phi(j\omega) = H(e^{-j\omega}) H(e^{j\omega}) = |H(e^{j\omega})|^2 \quad (4.31)$$

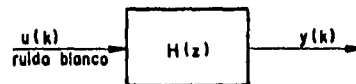
La prueba puede encontrarse en la referencia (4).

Corolario. TEOREMA 4.2. (De representación)

Dada una densidad espectral $\phi(j\omega)$, racional en $\cos \omega$ (o $e^{j\omega}$), entonces existe un sistema dinámico lineal, asintóticamente estable, tal que su salida es un proceso estocástico estacionario de densidad espectral $\phi(j\omega)$ si la entrada es ruido blanco discreto.

La demostración es inmediata: por el teorema 4.1 existe $H(z)$ tal que $|H(e^{j\omega})|^2 = \phi(j\omega)$. Luego alcanza con tomar un sistema cuya función de transferencia sea $H(z)$, la estabilidad está asegurada por el teorema 4.1.

Este teorema asegura la representación de cualquier proceso estacionario con densidad espectral racional, como resultado de un sistema dinámico



$$\phi_{yy}(j\omega) = H(e^{j\omega}) H(e^{-j\omega})$$

fig 4.9

lineal excitado con ruido blanco. Esta representación es válida para las estadísticas de segundo orden, no se puede extender a la función de distribución de los procesos. Dependiendo de la distribución de la señal de entrada, lo será la de la señal de salida.

Por ejemplo, es fácil demostrar que si $u(k)$ es un proceso de ruido blanco gaussiano, $y(k)$ también será gaussiano, con media:

$$E\{y} = H(1) \mu\}$$

y varianza

$$E\{(y-\mu)^2\} = \sigma^2 \left[\sum_{i=1}^{\infty} h_i^2 \right], \text{ donde } \sigma^2 = \text{var}(u(t)) \text{ y } h_i, i = 1, 2, \dots$$

es la sucesión de peso del sistema.

4.2.4 Procesos AR, IIA y ARMA

En la literatura de series de tiempo se ha extendido mucho la notación que distingue algunos casos particulares de la representación del párrafo anterior.

Consideremos un proceso estacionario $v(t)$:
$$v(t) = \frac{C(q^{-1})}{D(q^{-1})} e(t) \quad (4.32)$$

donde $e(t)$ es ruido blanco de media cero y varianza 1, y los polinomios C y D son:

$$C(q^{-1}) = C_0 + C_1 q^{-1} + \dots + C_r q^{-r}$$

$$D(q^{-1}) = 1 + d_1 q^{-1} + \dots + d_r q^{-r}$$

en el tiempo, (4.32) es:

$$v(t) = -d_1 v(t-1) - \dots - d_r v(t-r) + C_0 e(t) + C_1 e(t-1) + \dots + C_r e(t-r) \quad (4.33)$$

si $d_1 = d_2 = \dots = d_r = 0$, se obtiene un proceso de la forma:

$$v(t) = C_0 e(t) + C_1 e(t-1) + \dots + C_r e(t-r) \quad (4.34)$$

conocido como proceso de promedios móviles o M. A. (por moving-average).

Si fuera $C_1 = C_2 = \dots = C_r = 0$, se tendría un proceso de la forma:

$$v(t) = -d_1 v(t-1) - \dots - d_r v(t-r) + C_0 e(t) \quad (4.35)$$

conocido como proceso autorregresivo, o A. R. (por auto-regresivo).

Este tipo de procesos fue introducido en 1927 por Yule en un trabajo sobre análisis de series de manchas solares.

El caso general de la expresión (4.33) se denomina autorregresivo de promedios móviles, ARMA.

Para el modelo ARMA, pasando al plano complejo:

$$Y(z) = \frac{C^*(z)}{D^*(z)} U(z) \quad (4.36)$$

donde:

$$U(z) = Z\{u(t)\}$$

$$Y(z) = Z\{y(t)\}$$

$$\begin{cases} C^*(z) = C_0 z^r + C_1 z^{r-1} + \dots + C_r \\ D^*(z) = z^r + d_1 z^{r-1} + \dots + d_r \end{cases}$$

4.3 Forma canónica para el sistema y perturbaciones

4.3.1 El modelo ARMAX.

Una forma canónica para la descripción de entrada-salida de un sistema discreto de orden n , es:

$$y_1(k) = \frac{B_1(q^{-1})}{A_1(q^{-1})} U(k) \quad (4.37)$$

Donde A., B. son polinomios de grado n en el operador q^{-1} ; retardo unitario

$$A_1(q^{-1}) = 1 + a'_{11}q^{-1} + a'_{12}q^{-2} + \dots + a'_{1m}q^{-m}$$

$$B_1(q^{-1}) = b'_0 + b'_{11}q^{-1} + b'_{12}q^{-2} + \dots + b'_{1m}q^{-m}$$

para la ecuación (4.37), tomando transformada z se obtiene la función de transferencia del sistema:

$$H(z) = \frac{Y(z)}{U(z)} = \frac{z^m B_1(z^{-1})}{z^m A_1(z^{-1})}$$

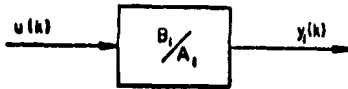


fig 4.10

Como se explicó al principio, trataremos de encontrar un modelo para la relación entrada-salida de un sistema con perturbaciones estocásticas aditivas. Por aditividad, la salida del sistema es:

$$y(k) = y_1(k) + v(k) \quad (4.38)$$

donde $y_1(k)$, descrito por la ecuación (4.37) es la parte determinística del sistema, y $v(k)$ es la perturbación aditiva. Suponiendo que $v(k)$ es un proceso estacionario de espectro racional, puede modelarse, de acuerdo al teorema

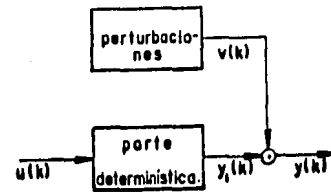


fig 4.11

4.2 como:

$$v(k) = \frac{C_1(q^{-1})}{A_2(q^{-1})} e(k) \quad (4.39)$$

donde C_1, A_1 , son polinomios en el operador q^{-1} , $e(k)$ es ruido blanco con media cero y varianza 1

$$C_1(q^{-1}) = c_0 + c_1 q^{-1} + c_2 q^{-2} + \dots + c_1^l q^{-l}$$

$$A_2(q^{-1}) = 1 + a_1^l q^{-1} + a_2^l q^{-2} + \dots + a_2^l q^{-l}$$

Si $v(k)$ es gaussiana, $e(k)$ también puede suponerse gaussiana.

Por multiplicación:

$$y(k) = \frac{B_1(q^{-1})}{A_1(q^{-1})} u(k) + \frac{C_1(q^{-1})}{A_2(q^{-1})} e(k) \quad (4.40)$$

$$y(k) = \frac{A_2(q^{-1})B_1(q^{-1})}{A_1(q^{-1})A_2(q^{-1})} u(k) + \frac{A_1(q^{-1})C_1(q^{-1})}{A_1(q^{-1})A_2(q^{-1})} e(k) \quad (4.41)$$

de donde:

$$y(k) = \frac{B(q^{-1})}{A(q^{-1})} u(k) + \frac{C(q^{-1})}{A(q^{-1})} e(k) \quad (4.42)$$

$$\text{con } A(q^{-1}) = A_1(q^{-1}) A_2(q^{-1}) = 1 + a_1 q^{-1} + \dots + a_n q^{-n}$$

$$B(q^{-1}) = A_2(q^{-1}) B_1(q^{-1}) = b_0 + b_1 q^{-1} + \dots + b_n q^{-n} \quad (4.43)$$

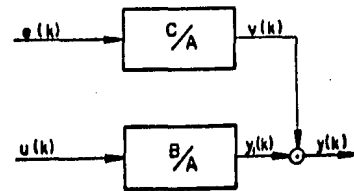
$$C(q^{-1}) = A_1(q^{-1}) C_1(q^{-1}) = c_0 + c_1 q^{-1} + \dots + c_n q^{-n}$$

$$n = m + l$$

La misma relación que (4.32) en forma explícita de ecuación en diferencias: (4.43)

$$y(k) = -a_1 y(k-1) - \dots - a_n y(k-n) + b_0 u(k) + \dots + b_n u(k-n) + c_0 e(k) + c_1 e(k-1) + \dots + c_n e(k-n)$$

En forma de diagrama de bloques, el sistema se representaría como en la figura (4.12).



Modelado de sistema y perturbaciones

Fig 4.12

Si la entrada es nula, se trata visiblemente de un proceso ARMA. En la literatura de series de tiempo se denomina usualmente a este modelo como ARMAX, esto es un modelo ARMA excitado además con una variable exógena $u(k)$.

Tomando transformada Z en la ecuación (4.42):

$$Y(z) = \frac{z^n B(z^{-1})}{z^n A(z^{-1})} U(z) + \frac{z^n C(z^{-1})}{z^n A(z^{-1})} E(z) \quad \forall z \neq 0 \quad (4.44)$$

La estabilidad depende de la posición de los polos de $z^n A(z^{-1})$, pero de (4.44) resulta que

$$z^n A(z^{-1}) = z^m A_1(z^{-1}) z^l A_2(z^{-1}) \quad (4.45)$$

el primer factor de (4.45) tiene sus ceros dentro del círculo unitario si el sistema (4.37) es estable, el segundo factor los tiene también dentro del círculo unitario por el Teorema 4.1 (de factorización espectral); se deduce que $z^n A(z^{-1})$ tiene todos sus ceros dentro del círculo unitario, con lo que se asegura la estabilidad asintótica.

Es bueno señalar aquí, que no hay pérdida de generalidad al asumir que los grados de los polinomios son iguales, puesto que en caso contrario pueden ser igualados a cero algunos coeficientes.

A pesar de esto en muchas circunstancias conviene incorporar la información que se tenga sobre la estructura del modelo, proveniente por ejemplo de consideraciones físicas. En estos casos se puede "afinar" el modelo al usar, por ejemplo retrasos (tiempo muerto) entre la entrada y salida, coeficientes nulos en el numerador o denominador, etc.

En ese sentido se puede elegir otra forma general para el modelo del sistema y perturbaciones:

$$y(t) = q^{-k} \frac{\beta_0 + \beta_1 q^{-1} + \dots + \beta_m q^{-m}}{1 + \alpha_1 q^{-1} + \dots + \alpha_n q^{-n}} u(t) + \frac{\gamma_0 + \gamma_1 q^{-1} + \dots + \gamma_r q^{-r}}{1 + \alpha_1 q^{-1} + \dots + \alpha_n q^{-n}} e(t) \quad (4.46)$$

Esta forma canónica (4.46) se relaciona con (4.42) en forma trivial.

4.3.2 Relación señal a ruido

Supongamos un sistema como (4.42), donde $u(t)$ es un proceso estacionario con media cero. En este caso, en analogía con los sistemas eléctricos, puede definirse la potencia de $u(t)$ (cualquiera sea su naturaleza) como

$$\text{Pot} \{u(t)\} = E \{u^2(t)\} = \text{var} \{u(t)\} \quad (4.47)$$

Definimos la relación señal a ruido del sistema (4.42) como:

$$S/N = \frac{\text{Pot} \{y_1(t)\}}{\text{Pot} \{v(t)\}} = \frac{E \{y_1^2(t)\}}{E \{v^2(t)\}} \quad (4.48)$$

es decir como la razón de la variancia de la parte determinística y la variancia de la perturbación del sistema.

Esta magnitud es muy importante en la cuantificación del ruido, puesto que casi siempre, más que el valor exacto de la variancia del ruido, nos interesa qué tan grande es ésta respecto a la parte determinística de la señal de salida.

Cálculo de varianzas

Para determinar la cantidad S/N hay que conocer $E\{y^2(t)\}$ y $E\{v^2(t)\}$, casos particulares de salidas de sistemas excitados por procesos estacionarios.

Consideremos el caso general:

$$\text{Sea } y(t) = \frac{B(q^{-1})}{A(q^{-1})} e(t) \quad (4.49)$$

con $e(t)$: ruido blanco de variancia 1.

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \dots + a_n q^{-n} \\ B(q^{-1}) &= b_0 + b_1 q^{-1} + \dots + b_n q^{-n} \end{aligned}$$

El problema es calcular $E\{y^2(t)\} = R_{yy}(0)$

De acuerdo a (4.23):

$$R_{yy}(0) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h(k) h(l) R_{ee}(k-l) \quad (4.50)$$

donde $h(i)$, $i = 1, 2, \dots$ es la sucesión de peso del sistema (4.49); como $e(k)$ es ruido blanco con $\text{var}\{e(k)\} = 1$, la ecuación (4.50) queda:

$$R_{yy}(0) = \sum_{k=0}^{\infty} h^2(k) \quad (4.51)$$

Esta es una ecuación que permite el cálculo de $R_{yy}(0)$.

Otra forma de cálculo es recordando que

$$R_{yy}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_{yy}(j\omega) e^{-j\omega\tau} d\omega \quad (4.52)$$

por el teorema de factorización espectral:

$$\phi_{yy}(j\omega) = \frac{B^*(e^{j\omega})}{A^*(e^{j\omega})} \frac{B(e^{-j\omega})}{A(e^{-j\omega})} \quad (4.53)$$

donde $\begin{cases} B^*(z) = z^n B(z^{-1}) \\ A^*(z) = z^n A(z^{-1}) \end{cases}$

Luego, para $\tau = 0$ (4.52) queda:

$$R_{yy}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{B^*(e^{j\omega})}{A^*(e^{j\omega})} \right| \cdot \left| \frac{B(e^{-j\omega})}{A(e^{-j\omega})} \right| d\omega \quad (4.54)$$

haciendo el cambio $z = e^{j\omega}$; $dz = je^{j\omega} d\omega$ $d\omega = \frac{1}{j} z^{-1} dz$

luego la integral (4.54)

$$R_{yy}(0) = \frac{1}{2\pi j} \oint \left[\frac{B^*(z)}{A^*(z)} \cdot \frac{B^*(z^{-1})}{A^*(z^{-1})} \right] \frac{1}{z} dz \quad (4.55)$$

c. u.

tomada sobre la circunferencia unitaria.

Esta integral puede ser evaluada por cálculo de residuos. Un método alternativo muy eficiente y general, básicamente propuesto por J. Nekolný en 1937 (38), consiste en evaluar (4.55) en forma recursiva.

Describiremos el algoritmo, la prueba puede ser consultada en (4), capítulo 5.

Llamemos:

$$\begin{cases} i_k^a = a_0^k z^k + a_1^k z^{k-1} + \dots + a_k^k & , k = n, n-1, \dots, 0 \\ i_k^b = b_0^k z^k + b_1^k z^{k-1} + \dots + b_k^k & , k = n, n-1, \dots, 0 \end{cases} \quad (4.56)$$

calculados con

$$\begin{cases} a_1^{k-1} = a_1^k - a_k a_{k-1}^k & ; a_k = a_k^k / a_0^k \\ b_1^{k-1} = b_1^k - b_k a_{k-1}^k & ; b_k = b_k^k / a_0^k \end{cases} \quad (4.57)$$

a partir de

$$\begin{cases} A_n^a(z) \equiv A^a(z) \\ B_n^b(z) \equiv B^b(z) \end{cases}$$

llamando ahora

$$I_k = \frac{1}{2\pi j} \oint_{C, u.} \left(\frac{B_k^b(z) B_k^a(z^{-1})}{A_k^b(z) A_k^a(z^{-1})} \right) \frac{1}{z} dz \quad (4.58)$$

resulta que

$$I_k = \frac{1}{a_0^k} \sum_{i=0}^k \frac{(b_i^a)^2}{a_i^k}$$

Como lo que interesa es sólo el cálculo de I_n , el algoritmo se resume así:

1). Calcular los coeficientes de $A_k^a, B_k^b, k = n-1, \dots, 0$, a partir

de $A_n^a \equiv A^a, B_n^b \equiv B^b$, en forma recursiva, con las ecuaciones

$$(4.57).$$

2). Evaluar I_n con (4.59).

Puede demostrarse además que si $a_0 > 0$, entonces el sistema es estable si y sólo si $a_k^k > 0, k = n-1, \dots, 0$. Esto hace que el algoritmo indique además si el sistema propuesto es estable.

Se implantó este algoritmo con el programa VARINZA/SALIDA, que puede encontrarse en el apéndice D.

Este programa fue escrito en lenguaje PASCAL para la computadora Burroughs-6800 del C. S. C. de la UNAM.

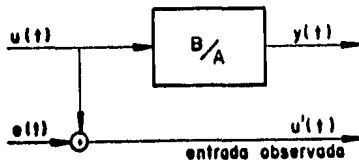
4.3.3 Casos particulares

Algunas situaciones especialmente importantes en la práctica, están contempladas como casos particulares de la forma canónica (4.42). Veamos

dos de ellas.

1) observaciones con ruido de medición en la entrada

Supongamos un sistema S de orden n , donde la medición de la señal de entrada está corruptida por el ruido blanco (por ejemplo, en el instrumento de medición).



Sistema con ruido de medición en la entrada

fig 4.13

Busquemos una relación entre las señales de entrada y salida medidas:

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})} u(t) \quad (\text{sistema})$$

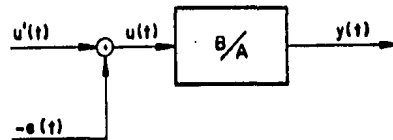
$$u'(t) = u(t) + e(t) \quad (\text{medición ruidosa})$$

donde $e(t)$ es una señal de ruido blanco, media cero, que perturba la medición de la entrada.

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})} u'(t) - u(t)$$

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})} u'(t) - \frac{B(q^{-1})}{A(q^{-1})} e(t) \quad (4.60)$$

(4.60) es una forma canónica para sistemas donde la observación de la entrada está perturbada con ruido blanco.



Sistema con ruido en la entrada

fig 4.14

Puede arribarse a la misma ecuación suponiendo que la señal de entrada al sistema está perturbada en forma aditiva con una señal de ruido blanco = $-e(t)$.

Si la perturbación a la entrada no es ruido blanco, pero tiene espectro racional, de acuerdo al teorema 4.1 existen $C(q^{-1})$, $D(q^{-1})$ tales que

$$v(t) = \frac{C(q^{-1})}{D(q^{-1})} e(t), \quad e(t) = \text{ruido blanco, media cero, var} = 1,$$

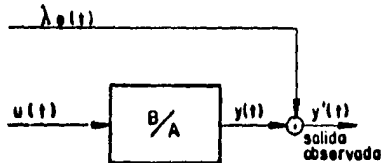
y la señal $v(t)$ tiene el mismo espectro que la perturbación.

Entonces el sistema puede ser modelado como:

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})} \frac{D(q^{-1})}{D(q^{-1})} u'(t) = \frac{C(q^{-1})}{D(q^{-1})} \frac{A(q^{-1})}{A(q^{-1})} e(t)$$

ii) sistema con error de medición a la salida

Supongamos ahora el mismo sistema dinámico discreto que antes, pero ahora hay errores en la observación de la salida. Suponemos que estos errores



Sistema con errores de medición a la salida

fig 4.15

no están correlacionados: son ruido blanco con media cero.

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})} u(t) \quad (\text{sistema})$$

$$y'(t) = y(t) + \lambda e(t) \quad (\text{observaciones})$$

$$y'(t) = \frac{B(q^{-1})}{A(q^{-1})} u(t) + \lambda \frac{A(q^{-1})}{A(q^{-1})} e(t)$$

donde $e(t)$ es ruido blanco de media cero, varianza 1, y λ representa la desviación cuadrática media de los errores.

Conclusiones

Partiendo de la base de que los modelos a utilizar en la identificación sean útiles a los propósitos de la identificación, se ha podido modelar en forma de ecuación en diferencias un sistema dinámico lineal en tiempo discreto, con perturbaciones aditivas de espectro racional.

Esta clase de modelos es el conjunto \mathcal{M} , de los modelos sobre el que se restringirá la búsqueda de uno particular $M \in \mathcal{M}$, tal que ajuste en forma óptima los datos observados según algún criterio C que es precisamente el tema de los métodos de identificación; en resumen la elección del criterio de optimalidad y la forma particular del modelo (un subconjunto $\mathcal{M}' \subset \mathcal{M}$) es lo que distinguirá un método de otro.

La suposición de que los sistemas son invariantes en el tiempo, y que los procesos estocásticos involucrados son estacionarios, es muchas veces injusta. Sin embargo, ésta puede ser una buena aproximación en algunos casos; en otros, cuando las variaciones de los parámetros sean suficientemente lentas respecto al tiempo de muestreo, podrán modificarse los algoritmos de identificación, para "seguir" las variaciones de los parámetros.

Por último hay que destacar que en el modulado para la identificación es importante incorporar todo el conocimiento que exista "a priori" sobre la estructura del modelo: orden, retrasos, valores de parámetros conocidos.

En general se asume una distribución Gaussiana de las perturbaciones, esto está refrendado porque comúnmente las perturbaciones son resultado de muchas contribuciones simultáneas de magnitudes aleatorias donde no hay una que pese más que el conjunto de todas.

Como se verá la distribución normal en las perturbaciones tiene sus ventajas. Aún cuando no sea así, el promediar intervalos de medición y recurrir al teorema del límite central arroja resultados similares.

Se tienen observaciones de la entrada y la salida: $\{u\}_N = u(1), \dots, u(1-m), \{y\}_N = y(1) \dots y(1-n)$ con $N > m+n$, y se quiere obtener el valor de los parámetros del sistema, es decir de los coeficientes de los polinomios A y B :

$$A(q^{-1}) = 1 + a_1 q^{-1} + \dots + a_n q^{-n} \quad (5.3)$$

$$B(q^{-1}) = b_0 + b_1 q^{-1} + \dots + b_m q^{-m} \quad (5.4)$$

De acuerdo a (5.2) las observaciones se pueden arreglar en forma vectorial. Si convenimos en llamar:

$$Y_N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \quad X_N = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix} \quad \begin{cases} x_i^T = (-y(i-1) \dots -y(i-n), u(i), \dots, \\ \quad \quad \quad u(i-m)); \quad i=1 \dots N \\ e_N^T = (e(1), \dots, e(N)) \\ \theta^T = (a_1, \dots, a_n, b_0, \dots, b_m) \end{cases} \quad (5.5)$$

entonces

$$Y_N = X_N \theta + e_N \quad (5.6)$$

Debo notarse que el modelo (5.2) ó (5.6) es un modelo de error de predicción, es decir de la forma

$$y(t) = E(y(t) | \{y\}_{t-1}, \{u\}_t, t, \theta) + e_t \quad (5.7)$$

donde $e(t)$ es el proceso de innovaciones, y

$$E(y(t) | \{y\}_{t-1}, \{u\}_t, t, \theta) = -a_1 y(t-1) - \dots - a_n y(t-n) + b_0 u(t) + \dots + b_m u(t-m) \quad (5.8)$$

CAPÍTULO 5

5.1 El método y propiedades de convergencia

5.1.1 Formulación del problema

Sea un sistema dinámico lineal e invariante en tiempo discreto, con perturbaciones aditivas de espectro racional. De acuerdo a lo visto en el capítulo 4, esto se puede poner en la forma

$$A(q^{-1}) y(t) = B(q^{-1}) u(t) + C(q^{-1}) e(t) \quad (5.1)$$

Supongamos por ahora que el polinomio $C(q^{-1}) = 1$, es decir que

$$A(q^{-1}) y(t) = B(q^{-1}) u(t) + e(t) \quad (5.2)$$

donde $e(t)$ es una sucesión centrada de ruido descorrelacionado, es decir un proceso estocástico tal que $E(e(t)) = 0$; $E(e(t)e(t+k)) = 0$ para $k = \pm 1, \pm 2, \dots$

Caso el sistema es invariante, las expresiones (5.7) y (5.8) no dependen explícitamente de t .

La linealidad del sistema permite poner el vector de observaciones y_N como en (5.6) en un modelo vectorial lineal. Observando (5.8) y (5.2) trataremos de ajustar un modelo de la forma

$$\hat{A}(q^{-1}) y(t) = \hat{b}(q^{-1}) u(t) \quad (5.9)$$

a los datos observados. Un criterio natural de ajuste es minimizar las discrepancias entre las salidas del modelo ajustado y las observaciones en el sentido de mínimos cuadrados. Es decir, si, como en el capítulo 2 llamamos residuos de la estimación a:

$$e_N = y_N - x_N \hat{\theta}_N \quad ; \quad e_N^T = (e_N(1), \dots, e_N(N)) \quad (5.10)$$

donde $\hat{\theta}_N$ es el estimado de θ con datos desde $t = 1$ hasta $t = N$; y definimos la función de costo

$$S_N(\hat{\theta}_N) = e_N^T e_N = \sum_{i=1}^N (e_N(i)^2) \quad (5.11)$$

entonces el estimador de mínimos cuadrados $\hat{\theta}_N$, dadas las observaciones $(u)_N$ $(y)_N$ es aquél que minimiza la función de costo $S_N(\hat{\theta})$

$$\hat{\theta}_N = \min_{\theta} S_N(\theta) \quad (5.12)$$

En el capítulo (2) se derivó la expresión (2.65) para el estimador:

$$\hat{\theta}_N = (X_N^T X_N)^{-1} X_N^T y_N \quad (5.13)$$

En resumen se ha completado el esquema de identificación por el método de mínimos cuadrados como sigue: dadas las observaciones $(u)_N$ $(y)_N$, se selecciona dentro de la clase de modelos (5.9) aquel que ajusta mejor los datos según el criterio de mínimos cuadrados (5.11). Debe notarse que el criterio tiene sentido en varios aspectos:

- i) Minimiza la suma de cuadrados de las diferencias entre las salidas calculadas y observadas.
- ii) Estas diferencias, desde el momento que el modelo es de errores de predicción, son realmente las diferencias entre el valor real observado y el valor predicho a un paso con los datos anteriores.
- iii) En aplicaciones de control, en particular en control lineal con criterio cuadrático, es natural minimizar en media cuadrática el error de predicción.

5.1.2 Convergencia

En la sección (2.5) se estudiaron muchas propiedades del estimador de mínimos cuadrados, pero los resultados obtenidos no pueden ser generalizados directamente. Más aún, algunos de éstos (como por ejemplo el teorema 20 y la observación 3 al teorema 21) serán contradictorios.

Esto se debe a que por la forma de construir $Y_N, X_N, 0, \epsilon, u_N$ (ver (5.5)) se viola la hipótesis fundamental de la sección (2.5) de que ϵ y u son independientes. En efecto, X_N contiene como elementos, salidas del sistema que dependen de errores que son elementos de u_N .

En cualquier caso probamos que la mayor parte de los resultados de la sección (2.5) se cumplen asintóticamente, conforme $N \rightarrow \infty$.

Estudiaríamos varios resultados importantes referidos a la convergencia asintótica de la estimación de mínimos cuadrados en este caso. El primero es una extensión de Åström, (3), a un teorema de Munn y Wald para estimación de parámetros de procesos autorregresivos. Otros resultados más generales, aplicables a modelos de error de proyección aunque no tenga la estructura (5.2) se derivan de resultados de Teoría de Martingalas. Un tercer enfoque se verá en el Capítulo 7 con un resultado debido a L. Ljung et al. que demuestra la convergencia de una clase general de algoritmos, uno de los cuales es mínimos cuadrados.

Al ser los resultados de carácter asintótico, al final del capítulo se ejemplifican simulaciones para observar qué sucede para muestras finitas. Una manera de seguir este capítulo es leer luego de cada sección el párrafo correspondiente de simulaciones de la segunda parte del capítulo.

TEOREMA 5.1

Sea un sistema como (5.2) con A y B coprimos y observaciones $(u)_N, (y)_N$ de las entradas y salidas. Si

- 1) Los errores $u(t)$ son independientes e igualmente distribuidos, $E\{u(t)\} = 0$, todos los momentos de $u(t)$ existen y son finitos y $E\{u(t)^2\} = \sigma^2$
- 2) las raíces de $A'(z)$ están dentro del círculo unitario con $A'(z) = z^n + a_1 z^{n-1} + \dots + a_n$
- 3) Los límites: $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t)$ y $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t)u(t+r) = R_u(r)$ y la matriz $A = \{a_{ij}\}$ con $a_{ij} = R_u(i-j)$ es positiva definida (es decir $u(t)$ es una excitación persistente) entonces:

- a) el estimador de mínimos cuadrados \hat{O}_N converge en media cuadrática al valor real 0 para $N \rightarrow \infty$
- b) El estimador \hat{O}_N tiende en ley a una distribución normal con

$$E(\hat{O}_N) = 0 \text{ y } \text{cov}(\hat{O}_N) = \sigma^2 R^{-1}/N \quad (5.14)$$

con

$$R = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i x_i^T \quad (5.15)$$

la demostración puede verse en (3). Nótese que la condición 2 implica que el sistema no perturbado es asintóticamente estable, y la condición 3 implica que la señal de entrada en

"consistentemente excitante".

TEOREMA 5.2

Sea el sistema (5.2). Si se cumple:

- i) A y B son coprimos, y $A'(z)$ (como en el teorema) tiene sus ceros dentro del círculo unitario.
- ii) $e(t)$ es una secuencia de variables aleatorias independientes con media cero, y varianza σ^2 .
- iii) $y(t)$ es un proceso estacionario; entonces bajo débiles condiciones de regularidad (), se cumple que:
 - a) El estimador de mínimos cuadrados $\hat{\theta}_N$ es fuertemente consistente es decir $\hat{\theta}_N \xrightarrow{p} \theta$ si $N \rightarrow \infty$

- b) $\sqrt{N}(\hat{\theta}_N - \theta)$ tiende asintóticamente ($N \rightarrow \infty$) a estar normalmente distribuido

$$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{\text{en ley}} N(0, P) \quad (5.16)$$

donde

$$P = \sigma^2 E \{x_i x_i^T\}^{-1} \quad i = 1, 2, \dots, N$$

Un esbozo de la demostración puede verse en (26)

En la práctica para $N \rightarrow \infty$ se puede aplicar la ley de los grandes números y considerar que el promedio de $x_i x_i^T$ converge a su valor esperado

$$P = \sigma^2 \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N [x_i x_i^T]^{-1} = \sigma^2 \lim_{N \rightarrow \infty} \frac{1}{N} (X_N^T X_N)^{-1}$$

De aquí se deduce que para grandes muestras el estimador $\hat{\theta}$ se distribuye normalmente con media θ y matriz de covarianzas

$$\text{cov}(\hat{\theta}_N) = \sigma^2 (X_N^T X_N)^{-1} \quad (5.17)$$

Hemos hecho hincapié en que A y B sean coprimos. Veremos un enfoque más general:

Resultados generales de convergencia

Consideraremos sistemas como (5.2):

$$A_0(q^{-1}) y(t) = B_0(q^{-1}) u(t) + e(t) \quad (5.18)$$

donde

$$\begin{aligned} A_0(q^{-1}) &= 1 + a_1 q^{-1} + \dots + a_n q^{-n} \\ B_0(q^{-1}) &= b_1 q^{-1} + \dots + b_m q^{-m} \end{aligned} \quad (5.19)$$

Nótese que por (5.19) hemos excluido sistemas con término de transmisión directa. Recordando la discusión del capítulo anterior no es una limitación muy importante para nuestras aplicaciones.

$$\theta^T = (a_1, \dots, a_n, b_1, \dots, b_m), \quad \theta \in \mathbb{R}^{m+n} \quad (5.20)$$

Si llamamos:

$$D_T = \{ \theta' \in \mathbb{R}^{m+n}; \frac{B_0^T(z)}{A_0^T(z)} = \frac{B_0(z)}{A_0(z)} \text{ para casi toda } z \} \quad (5.21)$$

se observa que D_T es el conjunto de parámetros que definen sistemas que generan los mismos datos. Desde el punto de vista de la identificación será entonces imposible distinguir entre dos sistemas cuyos parámetros pertenecen a D_T . Corresponde al caso de cancelación de polos y ceros.

Definición

Un sistema es identificable bajo el criterio de mínimos cuadrados si siendo $\hat{\theta}_N$ el estimador de mínimos cuadrados $\hat{\theta}_N \xrightarrow{\text{prob } 1} \theta'$ D_T para $N \rightarrow \infty$.

Definición:

Un sistema es de parámetros identificables bajo el criterio de mínimos cuadrados si es identificable y D_T consta de un solo elemento.

Ejemplo

Si se pretende identificar un sistema con grados de A y B n_0 y m_0 respectivamente, con parámetros $\theta \in \mathbb{R}^{m+n}$ con $m > m_0$, $n > n_0$ no puede asegurarse que haya un solo modelo dentro de la clase considerada, que puede generar los datos medidos.

TEOREMA 5.3

Sea un sistema como (5.18), tal que:

- 1) $E\{e(t)\} = 0$; $E\{e^2(t)\} < c < \infty$
- 2) $u(t)$ es independiente de $e(t+k)$, $k = 1, 2, \dots$
- 3) $\lim_{N \rightarrow \infty} \sup \frac{1}{N} \sum_{t=1}^N E\{y^2(t) + u^2(t)\} < \infty$ con probabilidad 1
- 4) $\frac{1}{N} \sum_{t=1}^N x_i x_i^T > 0$ con probabilidad 1

(5.22)

Entonces el sistema en cuestión es identificable, es decir

$$\hat{\theta}_N \xrightarrow{\text{prob } 1} \theta' \in D^T \text{ para } N \rightarrow \infty$$

la demostración puede verse en (26) y (35).

Observaciones

- * Debe notarse lo poderoso de este resultado. Se admite que la entrada se genera por realimentación (lineal ó no), siempre que se cumplan 2, 3 y 4. La restricción 2 no significa problema por lo general, mientras se pueda suponer que el proceso que genera $e(t)$ es independiente del que genera $u(t)$
- * La restricción 3 es bastante amplia. Puede demostrarse que si la entrada $u(t)$ es generada como

$$u(t) = f_c(y(t) \dots y(0); u(t-1) \dots u(0)) + v_t(t) + w(t)$$

donde f puede ser no lineal y variable en el tiempo, $u_p(t)$ es independiente de las entradas anteriores y

$$\lim_{N \rightarrow \infty} \sup \frac{1}{N} \sum_{t=1}^N |u_p(t)|^2 < \infty$$

$w(t)$ es una secuencia de ruido blanco filtrado, independiente de $e(t)$ y 4° momentos finitos. Si f_t es tal que el sistema en lazo cerrado es exponencialmente estable (ver diferencia en seguida), entonces se cumple la restricción 3.

Esta observación es importante por ejemplo en generación de entradas a través de un lazo de control autoadaptivo.

- La restricción 4 es importante ya que de no cumplirse, en general el estimador $\hat{\theta}_N \xrightarrow{\text{prob}} \theta_1$ con $\theta_1 \in D_1$ donde hemos llamado D_1 al conjunto

$$D_1 = \{ \theta_1 \mid \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N E \{ (A_0(q^{-1}) - A_{01}(q^{-1})) y(t) - (B_{01}(q^{-1}) - B_0(q^{-1})) u(t) \}^2 = 0 \} \quad (5.23)$$

Este conjunto depende fuertemente de los datos, puede contener el parámetro 0 del sistema que genera los datos, así como de otros sistemas que para esa entrada en particular, genera la misma salida. En tal caso el sistema no es identificable para esos datos, puede ser por ejemplo que la entrada no contenga un conjunto suficiente de frecuencias para excitar todos los modos del sistema.

Si se impone una restricción sobre la clase de señales de entrada, puede asegurarse la identificabilidad del sistema. Por ejemplo: si la entrada se genera por:

$$u(t) = F(q^{-1}) y(t) + u_p(t)$$

donde $u_p(t)$ no depende de $e(t)$ y el sistema en lazo cerrado es exponencialmente estable, entonces alcanza que $u_p(t)$ sea persistentemente excitante de orden suficientemente alto.

Definición

Sea un sistema como (5.18) donde $u(t)$ está determinado por realimentación (lineal o no) de la salida. Si llamamos $y_N^0(t)$ y $u_N^0(t)$ a la salida y entrada del sistema en t , inicializado en $t=-N$ con condiciones iniciales nulas; $C(Y_{t-N})$ es alguna función escalar de las condiciones iniciales en $t=-N$; Y_{t-N} , tal que $E\{C(Y_{t-N})^2\} < K < \infty$, entonces si se cumple:

$$|y(t) - y_N^0(t)| \leq C(Y_{t-N}) \lambda^N; \quad |u(t) - u_N^0(t)| < C(Y_{t-N}) \lambda^N \quad (5.24)$$

para algún $\lambda < 1$, entonces se dice que el sistema en lazo cerrado es exponencialmente estable.

teorema

Consideramos el sistema (5.18), con $e(t)$ una sucesión de variables aleatorias con media cero, varianza y momentos de 4° orden finitos. La entrada está generada como

$$u(t) = \Gamma(q^{-1}) y(t) + u_R(t)$$

donde $u_R(t)$ es independiente de $e(t)$, y es persistentemente excitante o cumple

$$\delta I < \frac{1}{N} \sum_{\tau=1}^N ((u_R(\tau) \dots u_R(\tau-M)) (u_R(\tau) \dots u_R(\tau-M))^T) < \frac{1}{\delta} I \quad N \gg N_0 \quad (5.25)$$

para algún N .

Γ es tal que el sistema en lazo cerrado es exponencialmente estable, entonces

$$\hat{\theta}_N \xrightarrow{\text{prob } 1} \theta \in D_T \text{ para } N \rightarrow \infty$$

(si D_T es no vacío).

TEOREMA 5.4

Sea el sistema (5.2) con $u(t)$ y $y(t)$ procesos estacionarios en sentido amplio con momentos de 4° orden finito. Entonces si $\hat{\theta}_N$ es el estimador de mínimos cuadrados de θ y llamando

$$\hat{\sigma}_N^2 = \frac{1}{N} (Y_N - X_N \hat{\theta}_N)^T (Y_N - X_N \hat{\theta}_N) \quad (5.26)$$

$$S_N^* = \frac{1}{N-p} (Y_N - X_N \hat{\theta}_N)^T (Y_N - X_N \hat{\theta}_N) \quad (p = \dim \theta) \quad (5.27)$$

Se cumple que S_N y $\hat{\sigma}_N^2$ son estimadores débilmente constantes de σ^2 .

$$S_N^* \xrightarrow{\text{prob}} \sigma^2 \quad \left. \vphantom{S_N^*} \right\} \text{ para } N \rightarrow \infty \quad (5.28)$$

$$\hat{\sigma}_N^2 \xrightarrow{\text{prob}} \sigma^2 \quad \left. \vphantom{\hat{\sigma}_N^2} \right\} \text{ para } N \rightarrow \infty \quad (5.29)$$

5.7 Mínimos cuadrados y métodos de correlación

Con la notación de (5.5) el estimador de mínimos cuadrados cumple las ecuaciones normales

$$X_N^T X_N \hat{Q}_N = X_N^T Y_N \quad (5.30)$$

de acuerdo a la expresión de X_N y de Y_N :

$$\begin{array}{c}
 \begin{array}{ccc}
 \sum_{i=0}^{N-1} y^2(i) & \sum_{i=0}^{N-1} y(i)y(i-1) \dots \sum_{i=0}^{N-1} y(i)y(i-n+1) & \left| \begin{array}{ccc}
 \sum_{i=0}^{N-1} y(i)u(i+1) & \sum_{i=0}^{N-1} y(i)u(i) & \dots \sum_{i=0}^{N-1} y(i)u(i-m+1) \\
 \vdots & \vdots & \vdots \\
 \sum_{i=1}^{N-2} y^2(i) & \dots \sum_{i=1}^{N-2} y(i)y(i-n+2) & \left| \begin{array}{ccc}
 \sum_{i=1}^{N-2} y(i)u(i+2) & \sum_{i=1}^{N-2} y(i)u(i+1) & \dots \sum_{i=1}^{N-2} y(i)u(i-m+2) \\
 \vdots & \vdots & \vdots \\
 \sum_{i=1-n}^{N-n} y^2(i) & \dots \sum_{i=1-n}^{N-n} y(i)y(i+n) & \left| \begin{array}{ccc}
 \sum_{i=1-n}^{N-n} y(i)u(i+n) & \sum_{i=1-n}^{N-n} y(i)u(i+n-1) & \dots \sum_{i=1-n}^{N-n} y(i)u(i-m+n) \\
 \vdots & \vdots & \vdots \\
 \vdots & \vdots & \vdots \\
 \sum_{i=1}^N u^2(i) & \sum_{i=1}^N u(i)u(i-1) \dots \sum_{i=1}^N u(i)u(i-m) & \\
 \vdots & \vdots & \vdots \\
 \sum_{i=0}^{N-1} u^2(i) & \dots \sum_{i=0}^{N-1} u(i)u(i-m+1) & \\
 \vdots & \vdots & \vdots \\
 \sum_{i=1-m}^{N-m} u^2(i) & &
 \end{array} \right.
 \end{array} \right.
 \end{array}
 \end{array}
 \end{array}$$

... (5.31)

Se han exhibido sólo algunos elementos, la matriz es simétrica.

$$X_N^T Y_N = \begin{bmatrix}
 \sum_{i=0}^{N-1} Y(i) y(i+1) \\
 \vdots \\
 \sum_{i=0}^{N-n} Y(i) y(i+n) \\
 \vdots \\
 \sum_{i=1}^N y(i) u(i) \\
 \vdots \\
 \sum_{i=1-m}^{N-m} u(i) y(i+m)
 \end{bmatrix} \quad (5.32)$$

En los métodos de correlación se computan las cantidades

$$\hat{R}_{uu}(k) = \frac{1}{N-k} \sum_{i=1}^{N-k} u(i) u(i+k) \quad \hat{R}_{yu}(k) = \frac{1}{N-k} \sum_{i=1}^{N-k} y(i) u(i+k)$$

$$\hat{R}_{yy}(k) = \frac{1}{N-k} \sum_{i=1}^{N-k} y(i) y(i+k) \quad \hat{R}_{uy}(k) = \frac{1}{N-k} \sum_{i=1}^{N-k} u(i) y(i+k)$$

(5.33)

De acuerdo a estas definiciones, y despreciando las diferencias introducidas por los primeros y últimos sumandos:

5.2 Mínimos cuadrados y métodos de correlación

Ox la notación de (5.5) el estimador de mínimos cuadrados cumple las ecuaciones normales

$$X_N^T X_N \hat{O}_N = X_N^T Y_N \quad (5.30)$$

de acuerdo a la expresión de X_N y de Y_N :

$$\begin{array}{c}
 \begin{array}{ccc}
 \begin{array}{c} N-1 \\ \vdots \\ N-1 \\ \vdots \\ N-n \\ \vdots \\ N-n \end{array} & \begin{array}{c} N-1 \\ \vdots \\ N-2 \\ \vdots \\ N-n \end{array} & \begin{array}{c} N-1 \\ \vdots \\ N-2 \\ \vdots \\ N-n \end{array} \\
 \begin{array}{c} \sum_{i=0}^{N-1} y^2(i) \\ \vdots \\ \sum_{i=1}^{N-1} y^2(i) \\ \vdots \\ \sum_{i=1-n}^N y^2(i) \end{array} & \begin{array}{c} \sum_{i=0}^{N-1} y(i)y(i-1) \dots \sum_{i=0}^{N-1} y(i)y(i-n+1) \\ \vdots \\ \sum_{i=1}^{N-2} y(i)y(i-n+2) \\ \vdots \\ \sum_{i=1-n}^N y(i)y(i+n) \end{array} & \begin{array}{c} \sum_{i=0}^{N-1} y(i)u(i+1) - \sum_{i=0}^{N-1} y(i)u(i) \dots \sum_{i=0}^{N-1} y(i)u(i-m+1) \\ \vdots \\ \sum_{i=1}^{N-2} y(i)u(i+2) - \sum_{i=1}^{N-2} y(i)u(i+1) \dots \sum_{i=1}^{N-2} y(i)u(i-m+2) \\ \vdots \\ \sum_{i=1-n}^N y(i)u(i+n) - \sum_{i=1-n}^N y(i)u(i+n-1) \dots \sum_{i=1-n}^N y(i)u(i-m+n) \end{array} \\
 \hline
 \begin{array}{ccc}
 \sum_{i=1}^N u^2(i) & \sum_{i=1}^N u(i)u(i-1) \dots \sum_{i=1}^N u(i)u(i-m) & \\
 \sum_{i=0}^{N-1} u^2(i) & \dots \sum_{i=0}^{N-1} u(i)u(i-m+1) & \\
 \vdots & \vdots & \\
 \sum_{i=1-m}^{N-m} u^2(i) & &
 \end{array} \\
 \hline
 \begin{array}{ccc}
 n & & m+1 \\
 \leftarrow & & \rightarrow
 \end{array} \\
 \dots(5.31)
 \end{array}$$

Se han exhibido sólo algunos elementos, la matriz es simétrica.

$$X_N^T Y_N = \begin{bmatrix} \sum_{i=0}^{N-1} y(i)y(i+1) \\ \vdots \\ \sum_{i=0}^{N-n} y(i)y(i+n) \\ \vdots \\ \sum_{i=1}^N y(i)u(i) \\ \vdots \\ \sum_{i=1-m}^{N-m} u(i)y(i+m) \end{bmatrix} \quad (5.32)$$

En los métodos de correlación se computan las cantidades

$$\begin{aligned}
 \hat{R}_{uu}(k) &= \frac{1}{N-k} \sum_{i=1}^{N-k} u(i)u(i+k) & \hat{R}_{yu}(k) &= \frac{1}{N-k} \sum_{i=1}^{N-k} y(i)u(i+k) \\
 \hat{R}_{yy}(k) &= \frac{1}{N-k} \sum_{i=1}^{N-k} y(i)y(i+k) & \hat{R}_{uy}(k) &= \frac{1}{N-k} \sum_{i=1}^{N-k} u(i)y(i+k)
 \end{aligned} \quad (5.33)$$

De acuerdo a estas definiciones, y despreciando las diferencias introducidas por los primeros y últimos sumandos:

$$\begin{bmatrix} \hat{R}_{yy}(0) & \hat{R}_{yy}(1) & \dots & \hat{R}_{yy}(n-1) & \hat{R}_{yu}(1) & -\hat{R}_{uy}(0) & \dots & \hat{R}_{uy}(m-1) \\ \hat{R}_{yy}(1) & \hat{R}_{yy}(0) & \dots & \hat{R}_{yy}(n-2) & \hat{R}_{yu}(2) & -\hat{R}_{yu}(1) & \dots & \hat{R}_{uy}(m-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{yy}(n-1) & \hat{R}_{yy}(n-2) & \dots & \hat{R}_{yy}(0) & \hat{R}_{yu}(n) & -\hat{R}_{yu}(n-1) & \dots & \hat{R}_{uy}(m-n) \\ \hline & & & & \hat{R}_{uu}(0) & \hat{R}_{uu}(1) & \dots & \hat{R}_{uu}(m) \\ & & & & \hat{R}_{uu}(1) & \hat{R}_{uu}(0) & \dots & \hat{R}_{uu}(m-1) \\ & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & \hat{R}_{uu}(m) & \hat{R}_{uu}(m-1) & \dots & \hat{R}_{uu}(0) \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_n \\ \hat{b}_0 \\ \hat{b}_1 \\ \vdots \\ \hat{b}_m \end{bmatrix} = \begin{bmatrix} \hat{R}_{yy}(1) \\ \hat{R}_{yy}(2) \\ \vdots \\ \hat{R}_{yy}(n) \\ \hat{R}_{uy}(0) \\ \hat{R}_{uy}(1) \\ \vdots \\ \hat{R}_{uy}(m) \end{bmatrix}$$

... (5.34)

Nótese que los elementos definidos en (5.33), son las correlaciones estadísticas si los procesos son estacionarios y ergódicos.

Si se hace un análisis por correlación para estimar la respuesta a impulso, las cantidades \hat{R}_{uu} , \hat{R}_{yy} , \hat{R}_{yu} , \hat{R}_{uy} se pueden utilizar para el cálculo del estimador \hat{C} de mínimos cuadrados.

5.3 Errores correlacionados

Estudiaríamos qué es lo que sucede si se admiten modelos como (5.1), o como (5.2) con $e(t)$ correlacionados. Supongamos $u(t)$ correlacionados. Si los procesos $u(t)$, $e(t)$, $y(t)$ son tales que las cantidades definidas en (5.3) tienen límite para $N \rightarrow \infty$, la ecuación (5.34) se cumple asintóticamente con los elementos R iguales a las correlaciones muestrales. (Más aún, si los procesos son estacionarios y ergódicos entonces las R son las correlaciones estadísticas).

Supondremos que dichos límites existen, sin exigir, sin embargo la restricción fuerte de ergodicidad.

Del modelo (5.2)

$$y(t) = -a_1 y(t-1) \dots - a_n y(t-n) + b_0 u(t) + \dots + b_m u(t-m) + u(t)$$

... multiplicando sucesivamente por $y(t-1) \dots y(t-n)$, $u(t) \dots u(t-m)$:

$$\begin{cases} y(t)N(t-1) + \dots + a_1 y(t-1)N(t-1) + \dots + a_n y(t-n)N(t-1) + b_0 u(t)N(t-1) + \dots + b_m u(t-m)N(t-1) + e(t)N(t-1) \\ y(t)N(t-2) + \dots + a_1 y(t-1)N(t-2) + \dots + a_n y(t-n)N(t-2) + b_0 u(t)N(t-2) + \dots + b_m u(t-m)N(t-2) + e(t)N(t-2) \\ \vdots \\ y(t)N(t-n) + \dots + a_1 y(t-1)N(t-n) + \dots + a_n y(t-n)N(t-n) + b_0 u(t)N(t-n) + \dots + b_m u(t-m)N(t-n) + e(t)N(t-n) \quad (5.35) \\ y(t)N(t) + \dots + a_1 y(t-1)N(t) + \dots + a_n y(t-n)N(t) + b_0 u(t)N(t) + \dots + b_m u(t-m)N(t) + e(t)N(t) \\ \vdots \\ y(t)N(t-m) + \dots + a_1 y(t-1)N(t-m) + \dots + a_n y(t-n)N(t-m) + b_0 u(t)N(t-m) + \dots + b_m u(t-m)N(t-m) + e(t)N(t-m) \end{cases}$$

Si repetimos el sistema de ecuaciones (5.35) para $t = 1, 2, \dots, N$ y sumamos las ecuaciones correspondientes dividiendo entre N y definimos

$$R_{yy}(\ell) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N+\ell-1} v(k) u(k+\ell)$$

(si los procesos son ergódicos alcanza con definir $R_{yy}(\ell) = E\{v(k)u(k+\ell)\}$) que son las correlaciones muestrales

$$\begin{bmatrix}
 R_{yy}(0) & R_{yy}(1) & \dots & R_{yy}(n-1) & -R_{yu}(1) & -R_{yu}(2) & \dots & -R_{yu}(n-1) \\
 R_{yy}(1) & R_{yy}(0) & \dots & R_{yy}(n-2) & -R_{yu}(2) & -R_{yu}(1) & \dots & -R_{yu}(n-2) \\
 \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 R_{yy}(n-1) & R_{yy}(n-2) & \dots & R_{yy}(0) & -R_{yu}(n-1) & -R_{yu}(n-2) & \dots & -R_{yu}(n-1) \\
 \hline
 & & & & R_{uu}(0) & R_{uu}(1) & \dots & R_{uu}(n) \\
 & & & & R_{uu}(1) & R_{uu}(0) & \dots & R_{uu}(n-1) \\
 & & & & \vdots & \vdots & \ddots & \vdots \\
 & & & & R_{uu}(n) & R_{uu}(n-1) & \dots & R_{uu}(0)
 \end{bmatrix}
 \begin{bmatrix}
 e_1 \\
 e_2 \\
 \vdots \\
 e_n \\
 b_1 \\
 b_2 \\
 \vdots \\
 b_n
 \end{bmatrix}
 =
 \begin{bmatrix}
 R_{yy}(1) + R_{yu}(1) \\
 R_{yy}(2) + R_{yu}(2) \\
 \vdots \\
 R_{yy}(n) + R_{yu}(n) \\
 R_{uu}(0) - R_{uu}(0) \\
 R_{uu}(1) - R_{uu}(1) \\
 \vdots \\
 R_{uu}(n) - R_{uu}(n)
 \end{bmatrix}
 \quad (5.34)$$

La ecuación (5.36) es la que cumplen los parámetros y la ecuación (5.34) es la que cumplen los estimadores (asintóticamente), compárense.

Observemos que asintóticamente ($N \rightarrow \infty$) los términos R de (5.34) convergen a los R de (5.36). Ambas igualdades son equivalentes si:

- 1) $R_{yu}(k) = 0 \quad k = 1, 2, \dots, n$
- 2) $R_{uu}(k) = 0 \quad k = 0, 1, 2, \dots, m$

es decir, si las entradas y salidas no están correlacionadas con los errores posteriores, considerando adelantos $1 \dots m$ y $1 \dots n$ respectivamente. Desarrollando $y(t)$:

$$y(t) = \sum_{i=0}^m a_i u(t-i) + \sum_{i=0}^n \gamma_i e(t-i)$$

para que $R_{yu}(k) = 0$ no deben estar correlacionados los errores $e(t)$ entre sí y además $R_{uu}(k) = 0, k = 0, 1, 2, \dots$

Por eso es que si $e(t)$ es un proceso no caracterizable como ruido blanco, el estimador de mínimos cuadrados puede dar lugar a una estimación sesgada.

Es una gran diferencia con el caso estudiado en el Capítulo 2, donde vimos que aún en el caso de errores correlacionados el estimador de mínimos cuadrados era insesgado (perdía su propiedad de varianza mínima). Vemos que en la estimación de sistemas dinámicos lineales no se pueden extender el Teorema 20 y la observación 3 del Teorema 21 del Capítulo 2.

Sesgo de la estimación

Como

$$Y_N = X_N^T 0 + e_N$$

$$X_N^T X_N^T 0 = X_N^T Y_N - X_N^T e_N \quad (5.37)$$

$$E\left\{\frac{1}{N} X_N^T X_N^T 0\right\} = E\left\{\frac{1}{N} X_N^T Y_N\right\} - E\left\{\frac{1}{N} X_N^T e_N\right\} \quad (5.38)$$

(como 0 es constante)

$$E\left\{\frac{1}{N} X_N^T X_N^T 0\right\} = E\left\{\frac{1}{N} X_N^T Y_N\right\} - E\left\{\frac{1}{N} X_N^T e_N\right\} \quad (5.39)$$

Suponiendo ergodicidad, la expresión (5.39) es la misma que (5.36) y la (5.34) queda como:

$$E\left\{\frac{1}{N} X_N^T X_N^T\right\} \hat{\theta}_m = E\left\{\frac{1}{N} X_N^T Y_N\right\} \quad (5.40)$$

$$\text{donde } \hat{\theta}_m = \lim_{N \rightarrow \infty} \hat{\theta}_N.$$

Restando (5.40) y (5.39):

$$E\left\{\frac{1}{N} X_N^T X_N^T\right\} [\hat{\theta}_m - \theta] = E\left\{\frac{1}{N} X_N^T e_N\right\} \quad (5.41)$$

suponiendo $E\left\{\frac{1}{N} X_N^T X_N^T\right\}$ no singular:

$$\hat{\theta}_m - \theta = E\left\{\frac{1}{N} X_N^T X_N^T\right\}^{-1} E\left\{\frac{1}{N} X_N^T e_N\right\} \quad (5.42)$$

en (5.42) se puede verificar que si $E\left\{\frac{1}{N} X_N^T e_N\right\} = 0$ el estimador es insesgado.

Observación:

Métese que aunque los errores no sean independientes, si las entradas y los errores sí lo son, las últimas $n+1$ ecuaciones de (5.36) siguen siendo válidas.

En la sección 5.12.8 se muestran ejemplos de estimación sesgada por ruido correlacionado.

Para resolver el problema de identificación de un sistema, en presencia de ruido correlacionado, hay que recurrir a otras técnicas de identificación, algunas de las cuales son extensiones del método de mínimos cuadrados. Algunas de las técnicas utilizadas son:

- (a) mínimos cuadrados repetidos
- (b) mínimos cuadrados generalizados
- (c) método de variables instrumentales
- (d) método de máxima verosimilitud

5.4 El esquema recursivo de Plackett

Cuando se computa el estimador de mínimos cuadrados de un sistema dinámico en línea, en cada instante de medición se obtienen solo dos nuevos datos $u(t)$, $y(t)$. La expresión analítica del estimador de mínimos cuadrados permite el cómputo del estimador en el instante t , actualizando los cálculos obtenidos con los datos anteriores, con base en la nueva observación.

Veremos un esquema muy eficiente de cálculo recursivo del estimador, introducido por Plackett, [41] en 1950.

El estimador de mínimos cuadrados dados las observaciones hasta $t = N$ es:

$$\hat{\theta}_N = (X_N^T X_N)^{-1} X_N^T Y_N \quad (5.43)$$

$$\therefore \hat{\theta}_N = P_N b_N \quad (5.44)$$

con

$$\begin{cases} P_N = (X_N^T X_N)^{-1} \\ b_N = X_N^T Y_N \end{cases} \quad (5.45)$$

$$(5.46)$$

si se conocen ahora las observaciones en $t = N + 1$:

$$\hat{\theta}_{N+1} = P_{N+1}^{-1} B_{N+1} \quad (5.47)$$

para

$$P_{N+1}^{-1} = X_{N+1}^T X_{N+1} = \begin{bmatrix} X_N^T & ; & X_{N+1} \end{bmatrix} \begin{bmatrix} X_{N+1} \\ \vdots \\ X_{N+1}^T \end{bmatrix} \quad (5.48)$$

$$\therefore P_{N+1}^{-1} = X_N^T X_N + X_{N+1} X_{N+1}^T = P_N^{-1} + X_{N+1} X_{N+1}^T \quad (5.49)$$

El segundo sumando es la contribución de los nuevos datos a la inversa de la matriz de covarianzas P_{N+1}^{-1} .

El vector B_{N+1} también puede ponerse en forma recursiva:

$$B_{N+1} = \begin{bmatrix} X_N^T & ; & X_{N+1} \end{bmatrix} \begin{bmatrix} Y_N \\ \vdots \\ Y_{N+1} \end{bmatrix}$$

$$\therefore B_{N+1} = B_N + X_{N+1} Y_{N+1} \quad (5.50)$$

Busquemos ahora una forma de poner $\hat{\theta}_{N+1}$ recursivamente:

Premultiplicando (5.49) por P_{N+1} , postmultiplicando por P_N y postmultiplicando por X_{N+1}^{-1} :

$$P_N X_{N+1} = P_{N+1} X_{N+1} + P_{N+1} X_{N+1} \underbrace{X_{N+1}^{-1} P_N X_{N+1}}_P$$

notando que r es escalar:

$$P_N X_{N+1} = P_{N+1} X_{N+1} \left[1 + X_{N+1}^T P_N X_{N+1} \right] \quad (5.51)$$

$$\therefore P_{N+1} X_{N+1} = P_N X_{N+1} \left[1 + X_{N+1}^T P_N X_{N+1} \right]^{-1} \quad (5.52)$$

postmultiplicando por $X_{N+1}^T P_N$

$$P_{N+1} X_{N+1} X_{N+1}^T P_N = P_N X_{N+1} \left[1 + X_{N+1}^T P_N X_{N+1} \right]^{-1} X_{N+1}^T P_N \quad (5.53)$$

por (5.49)

$$X_{N+1} X_{N+1}^T = (P_N^{-1} - P_{N+1}^{-1})$$

entonces

$$P_{N+1} (P_N^{-1} - P_{N+1}^{-1}) P_N = P_N X_{N+1} \left[1 + X_{N+1}^T P_N X_{N+1} \right]^{-1} X_{N+1}^T P_N \quad (5.54)$$

$$\therefore P_{N+1} = P_N - P_N X_{N+1} \left[1 + X_{N+1}^T P_N X_{N+1} \right]^{-1} X_{N+1}^T P_N \quad (5.55)$$

hemos llegado a una expresión recursiva para P_{N+1} en función de P_N y los nuevos datos X_{N+1} . Luego, por (5.47):

$$\hat{\theta}_{N+1} = (P_N - P_N X_{N+1} \left[1 + X_{N+1}^T P_N X_{N+1} \right]^{-1} X_{N+1}^T P_N) (B_N + X_{N+1} Y_{N+1}) \quad (5.56)$$

por (5.44):

$$\hat{\theta}_{N+1} = \hat{\theta}_N - P_N X_{N+1} \left[1 + X_{N+1}^T P_N X_{N+1} \right]^{-1} X_{N+1}^T \hat{\theta}_N + P_N X_{N+1} Y_{N+1} \left[1 + X_{N+1}^T P_N X_{N+1} \right]^{-1} \left[1 - \frac{X_{N+1}^T P_N X_{N+1}}{1 + X_{N+1}^T P_N X_{N+1}} \right]$$

$$\therefore \hat{\theta}_{N+1} = \hat{\theta}_N - P_N X_{N+1} (X_{N+1}^T \hat{\theta}_N - Y_{N+1}) \left[1 + X_{N+1}^T P_N X_{N+1} \right]^{-1} \quad (5.57)$$

Observaciones

Se ha encontrado una forma recursiva para $\hat{\theta}_{N+1}$ con algunas ventajas:

- 1) No es necesario repetir todo el cálculo en cada recursión, tomando en cuenta todos los datos anteriores $\{u(t)\}$, $\{y(t)\}$.

Toda la información necesaria sobre los datos anteriores está contenida en P_N y $\hat{\theta}_N$. Los nuevos datos entran a través del vector v_{N+1} y la observación y_{N+1} .

- 2) No es necesario invertir ninguna matriz, en lugar de esto en cada recursión se invierte un número: $1 + x_{N+1}^T P_N x_{N+1}$

- 3) En cada recursión se obtiene la matriz de covarianzas

$P_N = (X_N^T X_N)^{-1}$ necesaria para estimar la varianza de los estimadores.

- 4) Observando (5.57) puede notarse que el nuevo estimador es el anterior, actualizado con una corrección proporcional al error de predicción: $y(N+1) - x_{N+1}^T \hat{\theta}_N$. En particular si el sistema no tiene perturbaciones ($e(t) = 0 \forall t$), una vez alcanzado el valor real de θ en $t = N$, la estimación $\hat{\theta}_{N+1}$ se mantiene en θ .

El factor de proporcionalidad de la corrección es:

$$K_N = P_N x_{N+1} [1 + x_{N+1}^T P_N x_{N+1}]^{-1} \quad (5.58)$$

El método puede ser considerado como un procedimiento multidimensional de optimización por gradiente (Ver (54)).

En resumen el algoritmo queda como sigue:

- En la recursión $N + 1$:

- i) Se calcula P_{N+1} en base a x_{N+1}^T y P_N con:

$$P_{N+1} = P_N - P_N x_{N+1} (1 + x_{N+1}^T P_N x_{N+1})^{-1} x_{N+1}^T P_N$$

- ii) Se calcula $\hat{\theta}_{N+1}$ en base a x_{N+1} , $y(N+1)$, $\hat{\theta}_N$, P_N con:

$$\hat{\theta}_{N+1} = \hat{\theta}_N - P_N x_{N+1} (x_{N+1}^T \hat{\theta}_N - y(N+1)) [1 + x_{N+1}^T P_N x_{N+1}]^{-1}$$

- Valores iniciales P_0 , $\hat{\theta}_0$ (que estudiaremos enseguida).

Inicialización del algoritmo

El algoritmo puede inicializarse de dos formas:

- i) En forma exacta, resolviendo las ecuaciones normales para un N_0 a partir del cual se sigue con las recursiones.

Hay que tener en cuenta que si $N_0 < m + 1 = \dim \theta$, entonces la matriz $X_{N_0}^T X_{N_0}$ es singular y no se puede realizar la inversión. Existe más de una solución para los parámetros que dé residuos nulos para las N_0 observaciones consideradas.

Un caso particular es cuando $N_0 = m + 1$, donde se puede ver que en general $\det (X_{N_0}^T X_{N_0}) \neq 0$ con probabilidad 1. En este caso se puede encontrar una solución para los parámetros que dé residuos nulos. Se resuelve en este caso un sistema de $m + 1$ ecuaciones con $m + 1$ incógnitas.

Si $N_0 > m + 1$ habrá una solución de mínimos cuadrados.

4) Inicialización arbitraria

Muchas veces en implantación en línea del algoritmo, se desconocen observaciones pasadas en el momento de la inicialización. Podemos asignar un par de valores arbitrarios \hat{O}_0 y P_0 a partir de los cuales procesar el algoritmo. Hay una amplia gama de valores de \hat{O}_0 y P_0 que no afectan la identificación, es decir que $\hat{O}_N \rightarrow \hat{O}_N$ y $P_N \rightarrow P_N$ rápidamente, donde \hat{O}_N y P_N son los valores estimados con condiciones iniciales arbitrarias. Estudiaremos un caso especialmente útil:

$$\begin{cases} P_0 = \frac{1}{\epsilon} I & (\epsilon \text{ pequeño}) & (5.59) \\ \hat{O}_0 = 0 & & (5.60) \end{cases}$$

Lo veremos primero en forma determinística y luego daremos una interpretación estadística. Primero observamos que

$$\begin{aligned} P_0^{-1} &= P_0^{-1} \\ P_1^{-1} &= P_0^{-1} + x_1 x_1^T \\ \vdots \\ P_N^{-1} &= P_0^{-1} + x_1 x_1^T + \dots + x_N x_N^T = P_0^{-1} + P_N^{-1} \end{aligned} \quad (5.61)$$

si $P_0 = \frac{1}{\epsilon} I$, $P_0^{-1} = \epsilon I$. P_N^{-1} y P_N^{-1} difieren solamente en ϵ sobre su diagonal, ésta es además de elementos sumas de cuadrados:

$$P_{ii}^{-1}(N) = \sum_{j=1}^N \omega_j^2(j)$$

donde ω_j son entradas o salidas del sistema, pero lo que interesa es que $P_{ii}^{-1}(N)$ es creciente con N , i.e. el peso de la diferencia ϵ es decreciente. Supongamos ahora el mismo sistema en dos situaciones, la primera es la estimación para las observaciones dadas

$$\hat{O}_N = (X_N^T X_N)^{-1} X_N^T Y_N$$

la segunda es si hubieramos observado entradas y salidas desde $t = -(m+n)$ hasta $t = 0$ tales que

$$X'_0 = \begin{bmatrix} \sqrt{\epsilon} & & & \\ & \sqrt{\epsilon} & & \\ & & \ddots & \\ & & & \sqrt{\epsilon} \end{bmatrix} \quad Y'_0 = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \quad (5.62)$$

donde X'_0 , Y'_0 indican la situación hipotética. Luego el estimador en $t = N$ es:

$$\hat{O}'_N = (X'^T_N X'_N)^{-1} X'^T_N Y'_N$$

pero

$$X'_N = \begin{bmatrix} \sqrt{\epsilon} & & & & 0 \\ & \sqrt{\epsilon} & & & \\ & & \ddots & & \\ & & & \sqrt{\epsilon} & \\ & & & & X_N \end{bmatrix} \begin{matrix} \uparrow m+n+1 \\ \\ \\ \uparrow m+n+1 \end{matrix} \quad Y'_N = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ Y_N \end{bmatrix} \begin{matrix} \uparrow m+n+1 \\ \\ \\ \uparrow m+n+1 \end{matrix} \quad (5.63)$$

de donde $X'^T_N X'_N = X'^T_N X_N$

i.e. $\hat{O}'_N = P'_N X'^T_N Y'_N$

recordando que

$$\hat{\theta}_N = P_N X_N^T Y_N$$

y que $P_N^{-1} = P_{N-1}$, entonces $\hat{\theta}_N^* \rightarrow \hat{\theta}_N$.

Para concluir debe observarse que la hipótesis de las observaciones X_N^* y Y_N^* es equivalente a $\hat{\theta}_0 = 0$, $P_0 = \frac{1}{\epsilon} I$, entonces el estimador $\hat{\theta}_N^*$ con estas condiciones iniciales converge a $\hat{\theta}_N$.

En las simulaciones hemos utilizado valores de $\epsilon: 10^{-2} \leq \epsilon \leq 10^{-4}$ con excelentes resultados. Pueden verse ejemplos en la sección 5.12.6.

Una interpretación probabilística resulta de considerar P_0 como la matriz de covarianzas de los estimadores, e inicializar el algoritmo con un estimador finito ($\hat{\theta}_0 = 0$ por ejemplo) y una $P_0 = \frac{1}{\epsilon} I$ con elementos diagonales grandes que indiquen una gran incertidumbre a priori del estimador $\hat{\theta}_0 = 0$.

Este algoritmo se implementó con el programa MICU/S, que se encuentra en el apéndice D.

Extensiones

Es fácil ver que el procedimiento de mínimos cuadrados es inmediatamente extensible a sistemas con varias entradas y una salida, agregando en el vector de observaciones x_i dichas entradas, y en θ los parámetros correspondientes.

Es inmediatamente extensible a sistemas multientradas, multisalidas con modelos de matrices polinomiales. También es extensible a sistemas con sesgo en la salida, que puede interpretarse como un sistema con dos entradas, una de las cuales permanece constante.

Hemos implementado el programa MICU/C para cubrir este caso donde se supone un sistema de la forma:

$$y(t) = -a_1 y(t-1) - \dots - a_n y(t-n) + b_0 u(t) + \dots + b_m u(t-m) + e(t) + c$$

El procedimiento de mínimos cuadrados tiene la misma forma que antes, poniendo:

$$x_i^T = (-y(i-1), \dots, -y(i-n), u(i), \dots, u(i-m), 1)$$

y

$$\theta^T = (a_1, \dots, a_n, b_0, \dots, b_m, c)$$

Formulación general

Reescribamos el algoritmo de una forma general que luego permitirá la comparación con otros algoritmos

$$\begin{cases} \hat{\theta}(N+1) = \hat{\theta}(N) + K(N) [y_{N+1} - x_{N+1}^T \hat{\theta}(N)] \\ K(N) = P(N) x_{N+1} [1 + x_{N+1}^T P(N) x_{N+1}]^{-1} \\ P(N+1) = P(N) - K(N) x_{N+1}^T P(N) \end{cases} \quad (5.64)$$

En nuestro caso $\alpha = 1$ más adelante discutiremos otros valores para α .

Número de operaciones

Una característica importante de los algoritmos recursivos en relación con su implantación es el número de operaciones por recursión. Veamos esto para el algoritmo de Plackett.

En cada recursión (no tomamos en cuenta la inicialización), el cálculo se puede efectuar de la siguiente manera:

Cálculo

- 1) Cálculo de $P_N x_{N+1}^T$ ($x_{N+1}^T P_N$ es lo mismo puesto que $(P_N x_{N+1}^T)^T = x_{N+1}^T P_N$)
- 2) Cálculo de $(x_{N+1}^T P_N x_{N+1} + 1)$
- 3) Cálculo de $x_{N+1}^T \hat{O}_N - v(N+1)$
- 4) Actualización de \hat{O}_{N+1} con (5.57)
- 5) Actualización de P_{N+1} con (5.55)
(tomando en cuenta que P_{N+1} es simétrica)

Total:

No de operaciones (aproximado)	
Sumas	Productos
p^2	p^2
p	p
p	p
p	$2p$
$\frac{p(p+1)}{2}$	$p(p+1)$
$\frac{1}{2} [3p^2 + 7p]$	$2p^2 + 5p$

Cuando $p \gg 3$ los términos de mayor orden son dominantes y el número de sumas $\approx 1.5 p^2$ y el de productos $\approx 2p^2$.

El número de productos es algo mayor que el de sumas, y su número puede ser determinante en relación al tiempo de muestreo para determinar por ejemplo en una implantación con un procesador digital, si es necesario incorporar multiplicadores por hardware, ya que por software el tiempo de ejecución de una multiplicación es varias veces superior al tiempo de ejecución de una suma.

5.5 Esquema recursivo SROOT

Veremos un esquema recursivo, numéricamente robusto, basado en la descomposición QR.

En el cálculo del estimador de mínimos cuadrados, es necesario resolver las ecuaciones normales:

$$X_N^T X_N \hat{O}_N = X_N^T Y_N$$

donde la incógnita es el vector \hat{O}_N , para ello hay muchos métodos numéricos. Cuando la matriz $X_N^T X_N$ esté malcondicionada, pueden haber problemas numéricos en el cálculo de

$$\hat{O}_N = (X_N^T X_N)^{-1} X_N^T Y_N$$

como se puede ver en el apéndice A, un método para resolver el problema $Ax = b$ que no empeora el número de condición de la matriz A , es la descomposición QR, que básicamente consiste en encontrar Q ortogonal tal que

$$A = QR$$

con R triangular superior, y resolver el problema

$$Rx = Q^T b$$

1) Solución de las ecuaciones normales en $t = N$ Sea Q_N ortogonal tal que

$$Q_N^T X_N = \begin{bmatrix} R_{11} \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \uparrow p \\ \downarrow N-p \end{matrix} \quad (5.65)$$

con R_N triangular superior. Llamemos además:

$$Q_N^T Y_N = \eta_N = \begin{bmatrix} \eta_{11} \\ \vdots \\ \eta_{21} \end{bmatrix} \begin{matrix} \uparrow p \\ \downarrow N-p \end{matrix} \quad (5.66)$$

La suma de los cuadrados de los residuos es (por (5.10)):

$$S_N(\hat{O}_N) = (Y_N - X_N \hat{O}_N)^T (Y_N - X_N \hat{O}_N) \quad (5.67)$$

como Q es ortogonal, $Q^T Q = I$, entonces:

$$S_N(\hat{O}_N) = (Y_N - X_N \hat{O}_N)^T Q_N Q_N^T (Y_N - X_N \hat{O}_N) \quad (5.68)$$

$$= (Q_N^T Y_N - Q_N^T X_N \hat{O}_N)^T (Q_N^T Y_N - Q_N^T X_N \hat{O}_N) \quad (5.69)$$

por 5.65 y 5.66:

$$S_N(\hat{O}_N) = \left[(\eta_{11} : \eta_{21}) - \hat{O}_N^T (R_N^T : 0) \right] \begin{bmatrix} \eta_{11} \\ \vdots \\ \eta_{21} \\ 0 \end{bmatrix} - \begin{bmatrix} R_N \\ 0 \end{bmatrix} \hat{O}_N$$

$$\therefore S_N(\hat{O}_N) = (\eta_{11} - R_N \hat{O}_N)^T (\eta_{11} - R_N \hat{O}_N) + \eta_{21}^T \eta_{21} \quad (5.70)$$

La solución de mínimos cuadrados, que minimiza $S_N(\hat{O}_N)$ es obviamente

$$R_N \hat{O}_N = \eta_{11} \quad (5.71)$$

o

$$\hat{O}_N = R_N^{-1} \eta_{11} \quad (5.72)$$

La solución numérica de 5.71 es sencilla puesto que R_N es triangular superior. La función de costo (suma de los cuadrados de los residuos) mínima es:

$$S_N(\hat{O}_N) = \eta_{21}^T \eta_{21}$$

2) Recursión en $t = N+1$ Se supone conocidos \hat{O}_N , R_N , $S_N(\hat{O}_N)$, y los nuevos datos x_{N+1} , $y(N+1)$, busquemos una recursión para calcular \hat{O}_{N+1} , R_{N+1} , $S_{N+1}(\hat{O}_{N+1})$

$$S_{N+1}(0) = S_N(0) + (y(N+1) - x_{N+1}^T 0)^2$$

de (5.70):

$$S_{N+1}(0) = (\eta_{11} - R_N 0)^T (\eta_{11} - R_N 0) + \eta_{21}^T \eta_{21} + (y(N+1) - x_{N+1}^T 0)^2$$

$$S_{N+1}(0) = \begin{bmatrix} \eta_{11} \\ \vdots \\ \eta_{21} \\ y_{N+1} \end{bmatrix} - \begin{bmatrix} R_N \\ x_{N+1}^T \end{bmatrix} 0 \quad \begin{bmatrix} \eta_{11} \\ \vdots \\ \eta_{21} \\ y_{N+1} \end{bmatrix} - \begin{bmatrix} R_N \\ x_{N+1}^T \end{bmatrix} 0 \quad + S_N(\hat{O}_N) \quad (5.73)$$

es fácil encontrar una transformación ψ_{N+1} ortogonal tal que

$$\psi_{N+1}^T \begin{bmatrix} R_N \\ \vdots \\ x_{N+1}^T \\ 0 \end{bmatrix} = \begin{bmatrix} R_{N+1} \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \uparrow p \\ \downarrow 1 \end{matrix}$$

donde R_{N+1} es triangular superior. En efecto, sea $\psi_N^T = \Lambda_p \dots \Lambda_1$

donde

$$\Lambda_i = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ 0 & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}, \quad i=1, \dots, p, \quad \text{con} \begin{cases} c_i = (R_N)_{ii} / \lambda_i \\ s_i = (x_{N+1})_i / \lambda_i \\ \lambda_i = [(R_N)_{ii}^2 + (x_{N+1})_i^2]^{1/2} \end{cases}$$

... (5.74)

ψ_{i+1}^T construida de esta forma cumple con (5.74). Se puede ver que

las matrices Λ_i son matrices de rotaciones. Definamos ahora:

$$\psi_{N+1}^T \begin{bmatrix} \eta_N \\ \vdots \\ y_{N+1} \end{bmatrix} = \begin{bmatrix} \eta_{N+1} \\ \vdots \\ v_{N+1} \end{bmatrix} \begin{matrix} p \\ \vdots \\ 1 \end{matrix} \quad (5.75)$$

La ecuación (5.73) se puede poner:

$$S_{N+1}(0) = \begin{bmatrix} \eta_N \\ \vdots \\ y_{N+1} \end{bmatrix} - \begin{bmatrix} R_N \\ \vdots \\ x_{N+1}^T \end{bmatrix} 0 \quad \psi_{N+1}^T \psi_{N+1}^T \begin{bmatrix} \eta_N \\ \vdots \\ y_{N+1} \end{bmatrix} - \begin{bmatrix} R_N \\ \vdots \\ x_{N+1}^T \end{bmatrix} 0 + S_N(\hat{0}_N)$$

para $\psi_{N+1}^T \psi_{N+1} = I$. Entonces:

$$S_{N+1}(0) = \begin{bmatrix} \eta_{N+1} \\ \vdots \\ v_{N+1} \end{bmatrix} - \begin{bmatrix} R_{N+1} \\ \vdots \\ 0 \end{bmatrix} 0 + S_N(\hat{0}_N)$$

$$\therefore S_{N+1}(0) = (\eta_{N+1} - R_{N+1} 0)^T (\eta_{N+1} - R_{N+1} 0) + v_{N+1}^2 + S_N(\hat{0}_N)$$

el estimador de mínimos cuadrados $\hat{0}_{N+1}$ minimiza $S_{N+1}(0)$ y es obvio

d) Se actualiza $\hat{0}_N$ con

$$\hat{0}_{N+1} = R_{N+1}^{-1} \eta_{N+1}$$

e) $S_{N+1}(\hat{0}_{N+1})$ de:

$$S_{N+1}(\hat{0}_{N+1}) = S_N(\hat{0}_N) + v_{N+1}^2$$

2) Se pasa a la recursión siguiente.

Falta discutir los valores iniciales.

Inicialización

El algoritmo puede inicializarse en forma exacta a partir de un η_0 o en forma aproximada.

La inicialización exacta es la que se discute en la parte 1) de este párrafo, solo falta encontrar la matriz Q_N^T ortogonal, tal que

$$Q_N^T X_N = \begin{bmatrix} \eta_N \\ \vdots \\ 0 \end{bmatrix} \quad (N \geq p, \text{ número de parámetros})$$

con R_N triangular superior. Este problema está resuelto en la sección A.3 del apéndice A. De acuerdo a los resultados que allá se obtienen, la matriz Q_N^T se construye como

$$Q_N^T = Q'_1 \dots Q'_p$$

Las matrices Q'_i , $i=0, \dots, p-1$ van triangularizando sucesivamente la matriz X_N , columna por columna. La matriz $Q'_1 \dots Q'_p X_N$ tiene la forma:

que es el que anula el primer sumando (los otros dos son constantes).

$$y \quad \left. \begin{aligned} \dots R_{N+1} \hat{O}_{N+1} &= n_{N+1} \\ S_{N+1}(\hat{O}_{N+1}) &= S_N(\hat{O}_N) + v_{N+1}^2 \end{aligned} \right\} \quad (5.76)$$

$$\hat{O}_{N+1} = R_{N+1}^{-1} n_{N+1} \quad (5.77)$$

si R_{N+1} es no singular.

Otra vez la solución de (5.76) es sencilla puesto que R_{N+1} es triangular superior. Podemos resumir el algoritmo como sigue:

Algoritmo: en $t = N+1$

1) Dadas \hat{O}_N , R_N , n_N , $S_N(\hat{O}_N)$ y los nuevos datos x_{N+1}^T , $y_{(N+1)}$ se calculan:

a) $\psi_{N+1}^T = A_1 \dots A_N$ con (5.74)

b) Se actualiza R_N :

$$\begin{bmatrix} R_{N+1} \\ 0 \end{bmatrix} = \psi_{N+1}^T \begin{bmatrix} R_N \\ x_{N+1}^T \end{bmatrix}$$

c) Se actualizan n_N y v_N de

$$\begin{bmatrix} n_{N+1} \\ v_{N+1} \end{bmatrix} = \psi_{N+1}^T \begin{bmatrix} n_N \\ v_N \end{bmatrix}$$

$$Q^1_{i-1} \dots Q^1_0 X_N = \begin{bmatrix} \lambda_1 & x_{\dots} & x_{\dots} & \dots & x_{\dots} & \dots & x_{\dots} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \lambda_i & x_{\dots} & \dots & x_{\dots} & \dots & x_{\dots} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & x_i & x_{\dots} & x_{\dots} & \dots & x_{\dots} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & x_{N-i} & x_{\dots} & x_{\dots} \end{bmatrix}$$

y la matriz Q^1_i se construye a partir de ésta como:

$$Q^1_i = \begin{bmatrix} I_i & 0 \\ 0 & H_i \end{bmatrix} \quad \text{con} \quad \begin{cases} I_i = \text{matriz } (i \times i) \text{ identidad} \\ H_i = \text{matriz de Householder} = [h_{ij}] \\ i, j = 1, \dots, N-i \end{cases}$$

los elementos de la matriz H_i de Householder son, de acuerdo a (A.12):

$$h_{k\ell} = \begin{cases} = \frac{x_k x_\ell}{\lambda} & \text{si } k=1, \ell=1, 2, \dots, N-i \\ = 1 - \frac{x_k^2}{\lambda^2 - x_k^2} & \text{si } k=\ell, k=2, \dots, N-i \\ = -\frac{x_k x_\ell}{\lambda^2 - x_k^2} & k < \ell, k=2, \dots, N-i \end{cases} \quad (5.78)$$

con $\lambda^2 = x_1^2 + x_2^2 + \dots + x_{N-i}^2$, el signo de λ es $\text{sign}(\lambda) = -\text{sign}(x_1)$

el resto de los elementos se construyen por simetría de H_i .

Obsérvese que este resultado permite que no sólo el algoritmo sino su propia inicialización se puede hacer en forma recursiva.

Otra forma de inicializar el algoritmo es una elección arbitraria de \hat{O}_0 , R_0 , S_0 (\hat{O}_0), η_0 .

Una elección equivalente a la que hicimos en el anterior algoritmo recursivo es:

$$\begin{aligned} R_0 &= \delta I_p \text{ (pequeño)} \\ \hat{O}_0 &= 0 \\ S_0(\hat{O}_0) &= 0 \\ \eta_0 &= 0 \end{aligned} \quad (5.79)$$

En efecto, veamos la relación entre ambos algoritmos. Recordemos que habíamos llamado

$$\begin{aligned} \text{como } X_N &= O_N \begin{bmatrix} R_N \\ 0 \end{bmatrix} \text{ (Q ortogonal)} \\ P_N &= (X_N^T X_N)^{-1} \\ P_N &= \begin{bmatrix} R_N^T & 0 \end{bmatrix} Q^T Q \begin{bmatrix} R_N \\ 0 \end{bmatrix}^{-1} = [R_N^T R_N]^{-1} \\ \therefore P_N &= R_N^{-1} R_N^{-T} \end{aligned} \quad (5.80)$$

El algoritmo recibe el nombre de square-root debido a que en lugar de proyectar la matriz P_N proyectamos la matriz R_N . Esta matriz es una "raíz cuadrada" de P_N .

De este hecho proviene además la robustez numérica del algoritmo. No importa qué malcondicionado esté el problema de mínimos cuadrados, siempre se asegura que P_N es positiva definida porque:

$$P_N = R_N^{-1} R_N^{-T}$$

La elección $P_0 = \frac{1}{c} I$ es equivalente a $R_0 = \sqrt{c} I$. (Llamando $c = \delta^2$ obtenemos la expresión (5.79)).

Además por (5.71)

$$R_N \hat{O}_N = \eta_N$$

como $\hat{O}_N = 0$, $\eta_N = 0$. Finalmente $S_0(\hat{O}_0)$ se asume cero (se ajustan p observaciones con p parámetros).

El algoritmo ha sido implantado con el programa SROOT/S para sistemas sin sesgo en la salida, y SROOT/C para sistemas con sesgo en la salida. Ambos programas fueron escritos en el lenguaje PASCAL y utilizados en la computadora B-6800 del Centro de Servicios de Computo de la UNAM (con compilador PASCAL de SAN DIEGO). Pruebas con datos simulados se presentan al final del capítulo. Los programas están en el apéndice D.

Número de operaciones

Veamos el número de operaciones elementales necesarias en una recursión del algoritmo SROOT. Consideraremos sumas y productos (o divisiones) como dos categorías distintas.

No se toma en cuenta la inicialización del algoritmo. En cada recursión se necesita:

1) Cálculo de las matrices

$$A_i, i=1, \dots, p$$

2) Cálculo del producto

$$A_p \dots A_1 \begin{bmatrix} x_{N+1}^T \\ \vdots \\ x_1^T \end{bmatrix}$$

3) Cálculo de $A_p \dots A_1$

$$\begin{bmatrix} 0_{N+1} \\ \vdots \\ 1 \end{bmatrix}$$

4) Resolución del sistema

$$\text{triangular } R_{N+1} \hat{O}_{N+1} = \begin{bmatrix} 0_{N+1} \\ \vdots \\ 1 \end{bmatrix}$$

Total

	No de operaciones aproximado:	
	Sumas	Productos
1)	p	5p
2)	p(p+1)-p	2(p(p+1))-2p
3)	2p	4p
4)	$\frac{p(p+1)}{2} - p$	$\frac{p(p+1)}{2}$
Total	$\frac{1}{2} [3p + 5p]$	$\frac{1}{2} [3p + 11p]$

Se ha contabilizado la elevación al cuadrado y la toma de raíz cuadrada en el cálculo de las matrices A_i como si fueran productos.

Puede verse que el número de operaciones de este algoritmo es mayor que para el algoritmo de Plackett, obteniéndose a cambio mayor robustez numérica.

Como ejemplo, para algunos valores de p el No.de operaciones necesarias para llevar a cabo una recursión con ambos algoritmos es:

p	Algoritmo de Plackett		Algoritmo SROOT	
	Sumas	Productos	Sumas	Productos
2	13	18	11	29
3	24	33	21	51
4	38	52	34	78
5	55	75	50	110
6	75	102	69	147

5.6 Cómputo de la matriz de covarianzas

Hemos visto que la matriz de covarianzas del estimador \hat{O}_N de mínimos cuadrados del modelo (5.2) es

$$\text{cov } \hat{O}_N = E \{ (\hat{O}_N - O) (\hat{O}_N - O)^T \} = \sigma^2 E \{ X_N^T X_N \}^{-1}$$

cundo se cumple que los errores de dicho modelo tienen $\text{cov } \{e(t)\} = \sigma^2 I_N$. Usando los resultados de los teoremas 5.1 y 5.2 estimaremos la matriz de covarianzas con

$$\hat{\text{cov}} \hat{O}_N = \hat{\sigma}^2 (X_N^T X_N)^{-1} \quad (5.81)$$

donde $\hat{\sigma}^2$ es un estimador de σ^2 ; de acuerdo al teorema 5.4 podría ser por ejemplo

$$\hat{\sigma}_N^2 = \frac{1}{N} S_N (\hat{O}_N) \quad (5.82)$$

$$S'_N = \frac{1}{N-p} S_N(\hat{O}_N) \quad (5.83)$$

que se demostró son débilmente consistentes. Discutiremos cómo computar el estimador \hat{O}_N en los dos algoritmos recursivos propuestos.

a) Covarianzas en el algoritmo de Plackett

Para estimar σ^2 utilizaremos S'_N , que posee mejores propiedades que $\hat{\sigma}_N^2$ respecto al sesgo (ver capítulo 2).

En este algoritmo la matriz $(X_N^T X_N)^{-1}$ se computa recursivamente para el cálculo de \hat{O}_N , la hemos llamado P_N . Lo único que queda por calcular es la función de costo $S_N(\hat{O}_N)$, buscaremos un esquema recursivo de cálculo.

$$S_N(\hat{O}_N) = [Y_N - X_N \hat{O}_N]^T [Y_N - X_N \hat{O}_N] = Y_N^T Y_N - \hat{O}_N^T X_N^T X_N Y_N + Y_N^T X_N \hat{O}_N - \hat{O}_N^T X_N^T X_N \hat{O}_N = Y_N^T Y_N - Y_N^T X_N (X_N^T X_N)^{-1} X_N^T Y_N \quad \dots \quad (5.84)$$

luego

$$S_{N+1}(\hat{O}_{N+1}) = Y_{N+1}^T Y_{N+1} - Y_{N+1}^T X_{N+1} (X_{N+1}^T X_{N+1})^{-1} X_{N+1}^T Y_{N+1}$$

haciendo una partición de Y_{N+1} , X_{N+1} , Y_{N+1} = $\begin{bmatrix} Y_N \\ \vdots \\ Y_{N+1} \end{bmatrix}$, X_{N+1} = $\begin{bmatrix} X_N \\ \vdots \\ X_{N+1} \end{bmatrix}$ y utilizando la actualización de $(X_N^T X_N)^{-1}$ ya vista:

$$S_{N+1}(\hat{O}_{N+1}) = (Y_N^T Y_N) + y^2(N+1) - [Y_N^T X_N + y(N+1) X_{N+1}^T] [(X_N^T X_N)^{-1} - \dots - \frac{(X_N^T X_N)^{-1} X_{N+1} X_{N+1}^T (X_N^T X_N)^{-1}}{1 + \lambda}] [X_N^T Y_N + y(N+1) X_{N+1}^T] \quad \dots \quad (5.85)$$

donde hemos llamado λ al escalar $\lambda = X_{N+1}^T (X_N^T X_N)^{-1} X_{N+1}$

Recordando que $\hat{O}_N = (X_N^T X_N)^{-1} X_N^T Y_N$ entonces:

$$S_{N+1}(\hat{O}_{N+1}) = (Y_N^T Y_N) - \hat{O}_N^T X_N^T X_N \hat{O}_N + \frac{y(N+1)^2}{1 + \lambda} - \frac{2y(N+1)}{1 + \lambda} Y_N^T X_N \hat{O}_N + \frac{\hat{O}_N^T X_N^T X_{N+1} X_{N+1}^T \hat{O}_N}{1 + \lambda} \quad \dots \quad (5.86)$$

observando que los dos primeros sumandos son $S_N(\hat{O}_N)$, (ver (5.84), y agrupando los últimos:

$$S_{N+1}(\hat{O}_{N+1}) = S_N(\hat{O}_N) + \frac{c'(N+1)^2}{1 + \lambda} \quad (5.87)$$

donde

$$c'(N+1) = y(N+1) - X_{N+1}^T \hat{O}_N \quad (5.88)$$

es el error de predicción en el tiempo $N+1$ con el estimador \hat{O}_N

$$\lambda = X_{N+1}^T (X_N^T X_N)^{-1} X_{N+1}$$

hemos obtenido una fórmula recursiva para la función de costo $S_N(\hat{O}_N)$. Como las cantidades c' , $1+\lambda$ ya han sido computadas como parte del algoritmo para la actualización de \hat{O}_N y $(X_N^T X_N)^{-1}$, entonces en el cálculo de $S_{N+1}(\hat{O}_{N+1})$ sólo se agrega un producto, una división y una suma.

b) Covarianzas en el algoritmo SQUARE-ROOT

En este algoritmo la función de costo se computa recursivamente según:

$$S_{N+1}(\hat{\theta}_{N+1}) = S_N(\hat{\theta}_N) + v_{N+1}^2$$

de acuerdo a (5.76), lo que queda por calcular es $P_N = (X_N^T X_N)^{-1}$. Por (5.80)

$$P_N = R_N^{-1} (R_N^{-1})^T$$

donde la matriz R_N es triangular superior y se computa recursivamente. Siendo R_N triangular superior, su inversa también lo es. En el apéndice A, sección A.5 hay un algoritmo recursivo para llevar a cabo la inversión. Llamamos $S = R^{-1}$

$$\therefore P_N = SS^T = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ & S_{22} & \dots & S_{2p} \\ & & \ddots & \vdots \\ 0 & & & S_{pp} \end{bmatrix} \begin{bmatrix} S_{11} & & & \\ S_{12} & S_{22} & 0 & \\ \vdots & \ddots & \ddots & \\ S_{1p} & \dots & \dots & S_{pp} \end{bmatrix} \quad (5.80)$$

$p = \text{no. de parámetros} = \text{dim}(\theta)$.

Entonces

$$P_N(i,j) = \sum_{k=1}^p S_{ik} S_{kj}^* = \sum_{k=1}^p S_{ik} S_{jk} = \sum_{k=\max(i,j)}^p S_{ik} S_{jk} \quad (5.90)$$

(pues $S_{ik} = 0$ si $i > k$)

como sólo nos interesan los elementos de la diagonal para estimar las varianzas de los estimadores de los parámetros:

$$P_N(i,i) = \sum_{k=i}^p (S_{ik})^2$$

En resumen, dada $R_N(i,j)$ $i, j = 1, \dots, p$ calculamos $S_N(i,j)$ $i, j = 1, \dots, p$ conocida $S_N(i,j)$ $i, j = 1, \dots, p$ calculamos $P_N(i,i) = \sum_{k=i}^p (S_{ik})^2$ $i=1, \dots, p$.

La estimación de la matriz de covarianzas fue incorporada a los programas MICO/S, MICO/C que realizan el algoritmo de Plackett y SEEST/C, SROOT/C que realizan el algoritmo SQUARE-ROOT. Los resultados de las pruebas con datos simulados se discuten al final del capítulo, y se comparan en la sección 5.12.5 con los estimadores que se verán en seguida.

c) Otros estimadores de σ^2

Cuando se nos presentó el problema de estimar σ^2 desconocíamos la expresión para actualizar $\hat{\sigma}_N^2$ por lo que derivamos otras formas para estimar la cantidad σ^2 .

De acuerdo a los resultados del capítulo 2 y la extensión demostrada a través del teorema 5.4, asintóticamente ($N \rightarrow \infty$) se cumple que

$$E \left(\frac{1}{N-p} \sum_{i=1}^N v_N^2(i) \right) = \sigma^2 \quad (5.91)$$

donde $v_N(i)$ son los residuos calculados con el estimador $\hat{\theta}_N$, y la suma de sus cuadrados es la función de costo minimizada en la estima -

ción:

$$e_N(i) = y(i) - x_i^T \hat{O}_N \quad (5.92)$$

$$S_N(\hat{O}_N) = \sum_{i=1}^N e_N(i)^2 \quad (5.93)$$

(5.92) se debe a la definición de residuos (ver (5.10) y (5.93) es la definición de la función de costo (ver (5.11)). Por (5.91) se cumple asintóticamente que:

$$E\left\{\sum_{i=1}^N e_N(i)^2\right\} = \sigma^2 (N-p) \quad (5.94)$$

Suponiendo que $e_N(i)$ $i = 1 \dots N$ estén igualmente distribuidos:

$$E\{e_N^2(i)\} = \sigma^2 \left(\frac{N-p}{N}\right) \quad i = 1 \dots N$$

en particular para $i = N$

$$E\{e_N^2(N)\} = \sigma^2 \left(\frac{N-p}{N}\right) \quad (5.95)$$

El elemento $e_N(N)$ es fácil de calcular: es el último residuo de la estimación actual, por (5.92):

$$e_N(N) = y(N) - x_N^T \hat{O}_N$$

luego

$$E\left\{\sum_{j=1}^N e_j^2(j)\right\} = \sum_{j=p+1}^N E\{e_j^2(j)\} = \sigma^2 \sum_{j=p+1}^N \frac{1-p}{j} \quad (5.96)$$

la suma comienza desde $p+1$ pues hasta p los residuos son nulos (el ajuste del modelo es exacto).

$$\therefore E\left[\frac{\sum_{j=p+1}^N e_j^2(j)}{\sum_{j=p+1}^N \frac{1-p}{j}}\right] = \sigma^2 (N-p) \quad (5.97)$$

Coherentemente con el último resultado proponemos el estimador:

$$S_N^{11} = \frac{\sum_{j=p+1}^N e_j^2(j)}{\sum_{j=p+1}^N \frac{1-p}{j}} \quad (5.98)$$

El estimador S_N^{11} es actualizable recursivamente. Si el estimador S_N^{11} se utiliza con el algoritmo de Plackett inicializado arbitrariamente, puede suponerse que las observaciones comenzaron en $t = -p$, y el estimador queda:

$$S_N^{11} = \frac{\sum_{j=1}^N e_j^2(j)}{\sum_{j=1}^N \frac{1}{j+p}} \quad (5.99)$$

Este estimador fue implementado en MICU/C/versión 2, los resultados de pruebas comparativas con los otros estimadores se ven en el final del capítulo con datos de sistemas simulados.

De la misma forma que antes puede encontrarse otro estimador de σ^2 :

$$S_N^{111} = \frac{1}{N-p} \sum_{j=p+1}^N \left[\frac{e_j^2(j)}{\left(\frac{1-p}{j}\right)} \right] \quad (5.100)$$

como antes con inicialización arbitraria, el estimador queda:

$$\hat{x}_N = \frac{1}{N} \sum_{j=1}^N \begin{bmatrix} 1 \\ \frac{c_j(j)}{1+j^2} \end{bmatrix} \quad (5.101)$$

5.7 Mínimos cuadrados y filtro de Kalman

Consideremos un sistema dinámico lineal en tiempo discreto con representación en variables de estado de la forma:

$$x(t+1) = \phi x(t) + r(t) \quad (5.102)$$

$$y(t) = C x(t) + v(t)$$

donde $x(t) \in \mathbb{R}^p$ es el vector de estados, $\{r(t), t=1,2,\dots\}$, $\{v(t), t=1,2,\dots\}$ son secuencias de variables aleatorias independientes, igualmente distribuidas con media cero y matrices de covarianzas:

$$E\{r(t) r(t)^T\} = R_1 \quad (5.103)$$

$$E\{v(t) v(t)^T\} = R_2$$

El vector de estados $x(0)$ es una variable aleatoria con media m y matriz de covarianzas R_0 . Entonces se puede establecer el siguiente:

TEOREMA 5.5 (Kalman)

El mejor estimador de $x(t)$ en el sentido de mínimos cuadrados es el que satisface las ecuaciones recursivas:

$$\hat{x}(t) = \hat{x}(t-1) + K(t) (y(t) - C\phi \hat{x}(t-1)) \quad (5.104)$$

$$\hat{x}(0) = m$$

donde

$$K(t) = S(t) C^T [C S(t) C^T + R_2]^{-1} \quad (5.105)$$

$$S(t) = \phi P(t-1) \phi^T + R_1 \quad (5.106)$$

$$P(t) = S(t) - K(t) C S(t) \quad (5.107)$$

$$S(0) = R_0 \quad (5.108)$$

Puede interpretarse la matriz $S(t)$ como un estimador a priori de la matriz de covarianzas de $\hat{x}(t)$ dados $y(1) \dots y(t-1)$, y la matriz $P(t)$ como la estimación a posteriori de la matriz de covarianzas de $\hat{x}(t)$ dadas las observaciones $v(1) \dots v(t)$.

Consideremos ahora la identificación por mínimos cuadrados de un sistema dinámico discreto, lineal, invariante en el tiempo como el (5.2).

Las ecuaciones recursivas del estimador son (5.64) con $n = 1$.

Supongamos ahora un sistema como (5.102) donde el vector de estados es el vector de parámetros del sistema que estamos identificando, es decir:

$$\left\{ \begin{array}{l} x(t) = (a_1, \dots, a_n, b_0, b_1, \dots, b_m)^T \\ \Phi = I_p \\ r(t) \text{ una v.a. de media cero y matriz de covarianzas } R_1 = 0 \\ C = C(t) = (-y(t-1), \dots, -v(t-n), u(t), \dots, u(t-m)) \\ v(t) = e(t) \text{ una sucesión de v.a. igualmente distribuidas, indepen-} \\ \text{dientes de media cero y covarianza } \sigma^2, \dots, R_2 = \sigma^2 \text{ Es de} \\ \text{cir que es un proceso estacionario caracterizable como rui-} \\ \text{do blanco.} \end{array} \right. \quad (5.109)$$

Entonces el sistema (5.102) queda como:

$$\left\{ \begin{array}{l} x(t+1) = x(t) \\ y(t) = C(t)x(t) + e(t) \end{array} \right. \quad (5.110)$$

El estimador $\hat{x}(t)$ del vector de estados obtenidos con el filtro de Kalman se reduce a:

$$\left\{ \begin{array}{l} \hat{x}(t) = \hat{x}(t-1) + K(t) [y(t) - C(t)\hat{x}(t-1)] \\ K(t) = P(t-1)C(t)^T [\sigma^2 + C(t)P(t-1)C(t)^T]^{-1} \\ P(t) = P(t-1) - K(t)C(t)P(t-1) \end{array} \right. \quad (5.111)$$

Reescribamos las ecuaciones (5.64) del estimador de mínimos cuadrados con dos ligeras variantes: lo haremos para la recursión N en lugar de la $N+1$, y en segundo lugar en vez de escribir las ecuaciones en función de la matriz $P_N = (X_N^T X_N)^{-1}$ lo haremos con $\hat{P}_N = \sigma^2 P_N$ ya que es el estimador de la matriz de covarianzas de \hat{O}_N es decir que tiene la misma interpretación que $P(t)$ en $t = N$ del filtro de Kalman:

$$\left\{ \begin{array}{l} \hat{O}_N = \hat{O}_{N-1} + K(N-1) [y(N) - x_N^T \hat{O}_{N-1}] \\ K(N-1) = P_{N-1} x_N \sigma^2 + x_N^T P_{N-1} x_N^{-1} \\ P_N^{\hat{}} = P_{N-1}^{\hat{}} - K(N-1) x_N^T P_{N-1} \end{array} \right. \quad (5.112)$$

Ambas recursiones son idénticas. En cuanto a las condiciones iniciales, si no hay información sobre el estado inicial $x(0)$, puede suponerse que es una variable aleatoria con media cero y matriz de covarianzas βI con $\beta \gg 1$. Luego:

$$\left\{ \begin{array}{l} \hat{x}(0) = 0 \\ P(0) = \beta I \text{ (si las condiciones iniciales son} \\ \text{nulas)} \end{array} \right. \quad (5.113)$$

Compárese con la inicialización arbitraria discutida para el algoritmo de mínimos cuadrados.

Una discusión más detallada sobre las relaciones entre filtro de Kalman y mínimos cuadrados puede leerse en las referencias (3), (4) y (54).

5.8 Extensiones para sistemas variables con el tiempo

La forma recursiva del estimador de mínimos cuadrados sugiere que se pueden hacer ligeras modificaciones para identificar sistemas cuyos parámetros no permanecen constantes a lo largo del proceso de mediciones. Veremos dos formas de atacar el problema.

5.8.1 Olvido exponencial

En su exposición sobre el método de mínimos cuadrados, Gauss (23), propone pesar los cuadrados de los residuos con una medida de la exactitud de las observaciones correspondientes. Lo que se propone con esta técnica es dar más peso a las últimas observaciones que a las anteriores, pasando en forma exponencial los datos pasados. Consideremos la función de costo

$$S_{N+1}(0) = c_N^2 + \alpha c_{N-1}^2 + \dots + \alpha^{N-1} c_1^2 \quad (5.114)$$

que puede ponerse como

$$\tilde{S}_{N+1}(0) = \alpha \tilde{S}_N(0) + (y_{N+1} - x_{N+1}^T \hat{0})^2 \quad (5.115)$$

donde $0 < \alpha < 1$. Nótese que para $\alpha = 1$ se tiene la misma función de costo que en el procedimiento de mínimos cuadrados ya estudiados. El estimador $\hat{0}_N$ que minimiza la función de costo (5.114) se dice que es el de mínimos cuadrados con peso exponencial, y satisface la recursión

$$\begin{aligned} \hat{0}_{N+1} &= \hat{0}_N + K_{N+1} (y_{N+1} - x_{N+1}^T \hat{0}_N) \\ K_{N+1} &= P_N x_{N+1} (a + x_{N+1}^T P_N x_{N+1})^{-1} \\ P_{N+1} &= \frac{1}{\alpha} [P_N - P_N x_{N+1} x_{N+1}^T P_N (\alpha x_{N+1}^T P_N x_{N+1})^{-1} P_N] \end{aligned} \quad (5.116)$$

Una demostración por inducción puede verse en (26). Otra forma de obtener (5.116) es seguir la demostración del algoritmo de Plackett considerando que se minimiza una función de costo de la forma

$$\tilde{S}_N(0) = (Y_N - X_N 0)^T W_N (Y_N - X_N 0) \quad (5.117)$$

donde W_N es diagonal de la forma:

$$W_N = \begin{bmatrix} \alpha^{N-1} & & & & 0 \\ & \alpha^{N-2} & & & \\ & & \ddots & & \\ 0 & & & \alpha^0 & \\ & & & & \ddots \\ & & & & & \alpha^0 \end{bmatrix} \quad \text{además } W_N = I_N^T \Gamma_N$$

donde

$$\Gamma_N = \begin{bmatrix} \sqrt{\alpha^{N-1}} & & & & 0 \\ & \ddots & & & \\ & & \ddots & & \\ 0 & & & \sqrt{\alpha^0} & \\ & & & & \ddots \\ & & & & & \sqrt{\alpha^0} \end{bmatrix}$$

$$\text{Si } Y_N = X_N 0 + \epsilon_N + \Gamma_N Y_N = I_N X_N 0 + \Gamma_N \epsilon_N \quad (5.118)$$

el estimador de mínimos cuadrados con peso exponencial se puede obtener de acuerdo a (2.98) como:

$$\hat{0} = (X_N^T W_N X_N)^{-1} X_N^T W_N Y_N \quad (5.119)$$

que es equivalente a obtener el estimador de mínimos cuadrados para (5.118).

Llamando

$$\begin{cases} X_N^{\hat{A}} = F_N^{\hat{A}} X_N & ; & P_N^{\hat{A}} = (X_N^{\hat{A}} X_N^{\hat{A}})^{-1} \\ Y_N^{\hat{A}} = F_N^{\hat{A}} Y_N & ; & B_N^{\hat{A}} = X_N^{\hat{A}T} Y_N^{\hat{A}} \end{cases} \quad (5.120)$$

Notando que

$$\begin{cases} P_{N+1}^{\hat{A}-1} = \alpha P_N^{\hat{A}-1} + x_{N+1} x_{N+1}^T \\ B_{N+1}^{\hat{A}} = \alpha B_N^{\hat{A}} + x_{N+1} y(N+1) \\ \hat{O}_{N+1}^{\hat{A}} = P_{N+1}^{\hat{A}} B_{N+1}^{\hat{A}} \end{cases}$$

Entonces se puede seguir todo el razonamiento de la demostración del algoritmo de PLACKETT para llegar finalmente a las ecuaciones (5.116).

Nótese la diferencia entre P_N y $P_N^{\hat{A}}$.

Para la estimación de la matriz de covarianzas puede usarse el resultado (2.101), que cuando $E\{e e^T\} = \sigma^2 I$ queda:

$$\text{cov} \begin{pmatrix} \hat{O}_N \\ \hat{O}_N \end{pmatrix} = \sigma^2 \begin{pmatrix} X_N^T & W_N X_N \end{pmatrix}^{-1} X_N^T W_N W_N X_N \begin{pmatrix} X_N^T & W_N X_N \end{pmatrix}^{-1} \quad (5.121)$$

debe tenerse en cuenta que este resultado fue derivado bajo hipótesis que no se aseguran en identificación de sistemas dinámicos.

El nombre de "olvido exponencial" para el algoritmo resulta claro de la definición de la función de costo. El peso que tiene el residuo correspondiente a un tiempo t decrece exponencialmente para N crecientes. La relación entre los pesos de los residuos correspondientes a tiempos t y $t - k$ es constante respecto N , y vale: α^k .

La selección de α depende pues, de qué tan rápido se espere que varíen los parámetros, es entonces importante cierta información *a priori*, sobre dicha variación. La razón es que no se puede disminuir α sin costo alguno. Una disminución de α significa una velocidad de asentamiento de los parámetros más grande, pues rápidamente la información anterior deja de pesar. Sin embargo, esto se hace a costa de una mayor incertidumbre en la determinación de los parámetros: la varianza de la estimación aumenta, y el algoritmo se vuelve más sensible a la perturbación $u(t)$. Esto se debe a que al hacer "olvidadizo" el algoritmo, el efecto es como si se tomara en cuenta un menor número de observaciones.

En el caso estudiado en el Capítulo 2 esto se traducía en que el estimador de mínimos cuadrados pesados no era el de varianza mínima (ver sección 2.5.4).

Se presenta una gráfica de α^k para distintos valores de α , $k = 0 \dots 150$ para comparar la rapidez de decrecimiento del peso de los residuos.

Este algoritmo ha sido implantado a través del programa *MINI/S/OLVEXP* escrito en PASCAL para la B-6800 del CSC de la UNAM. Ejemplos con datos simulados se discuten al final del Capítulo, en la sección 5.12.7.

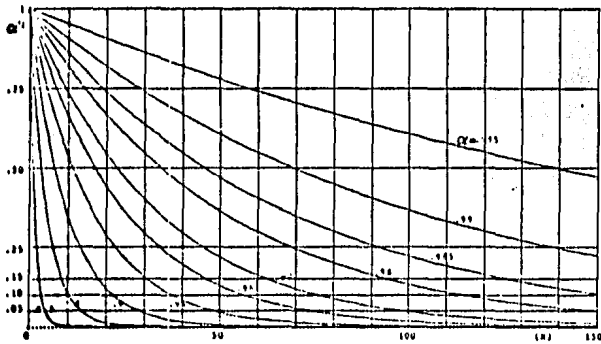


Fig 5.1 Curvas de la familia a^N , $0 < N < 150$ para $\alpha = 0.5, 0.8, 0.9, 0.94, 0.96, 0.97, 0.98, 0.985, 0.99, 0.995$.

Factor de olvido variable

Ocurre a veces que los primeros datos no son confiables y se quiere prescindir rápidamente de ellos pero no de los que siguen, se puede utilizar entonces un factor de olvido variable, creciente en el tiempo. Una forma de hacerlo puede ser por ejemplo un factor de peso con crecimiento exponencial hacia 1:

$$a_{k+1} = a_k + (1-a_k) \Delta a = (1-\Delta a) a_k + \Delta a \quad (5.122)$$

(ver por ejemplo (21)).

Si el factor de olvido α es constante, existe una cota mínima para los elementos de la matriz P_N^* , $N \rightarrow \infty$. Un factor de olvido exponencialmente creciente disminuye esta cota.

Puede demostrarse que para α_k variable el algoritmo minimiza la función de costo:

$$S^* = \sum_{k=1}^N \sum_{i=k}^N \lambda(i) \epsilon_k^2 \quad (5.123)$$

El algoritmo puede reescribirse como

$$\begin{cases} \hat{O}(N+1) = \hat{O}(N) + K(N+1) \left(y(N+1) - x_{N+1}^T \hat{O}_N \right) \\ K_{N+1} = P_N^* x_{N+1} \left(\alpha(N+1) + x_{N+1}^T P_N^* x_{N+1} \right)^{-1} \\ P_{N+1}^* = \frac{1}{\alpha_{N+1}} \left[P_N^* - \frac{P_N^* x_{N+1} x_{N+1}^T P_N^*}{\alpha_{N+1} + x_{N+1}^T P_N^* x_{N+1}} \right] \end{cases} \quad (5.124)$$

la recursión en α es (5.122), con valor inicial α_0 y constante Δa ó la recursión

$$a_{k+1} = a_0 a_N + (1 - a_0) \quad (5.125)$$

si se considera $\alpha_0 = 1 - \Delta a$.

En este caso se pueden calcular los factores de peso de (5.123) de 5.125:

$$a_N = 1 - a_0^N (1 - a(0))$$

y asintóticamente para N grande:

$$\prod_{i=k}^N a(i) = a_0^k \left[\frac{1-\alpha(0)}{1-a_0} (1-a_0^{N+1-k}) \right]$$

en el límite para $N \rightarrow \infty$:

$$\prod_{i=k}^{\infty} a(i) = a_0^{-k} \frac{1-\alpha(0)}{1-a_0} \quad (5.125')$$

(La aproximación es válida p. ej. para $N \geq 100$ y $a_0 \leq .97$)

Se presentan familias de curvas $\prod_{i=k}^{\infty} a(i)$ parametrizadas en a_0 y $\alpha(0)$ que representan el peso de los residuos ϵ_k en la función de costo cuando $N \rightarrow \infty$.

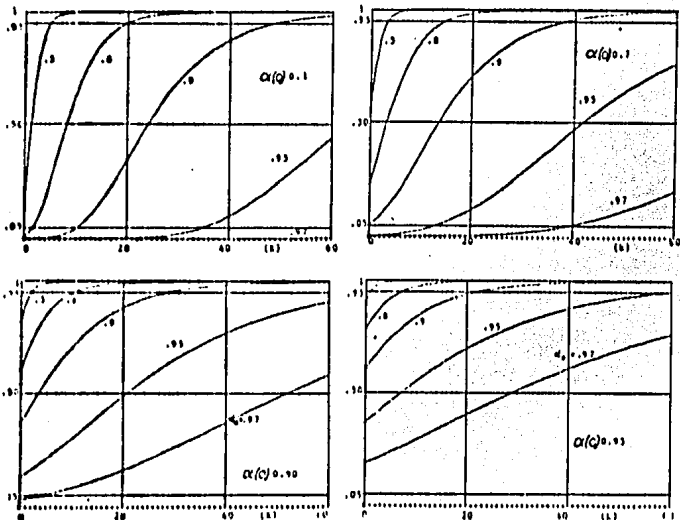


Fig 5.2 Curvas de la forma (5.125') para $\alpha(0) = 0.1, 0.7, 0.9, 0.95$,
 $a_0 = 0.5, 0.8, 0.9, 0.95, 0.97$

Quando se esperan variaciones bruscas de los parámetros, puede utilizarse el algoritmo (5.124), de forma que al detectarse una variación de los parámetros se reinicialice la recursión (5.125) del factor de olvido a , con la consiguiente reducción del peso de los datos anteriores al cambio.

El problema de la detección de dicho cambio puede ser hecho como prueba estadística de la hipótesis nula: los nuevos datos son generados por el mismo modelo vs los nuevos datos no son generados por el mismo modelo.

Basados en los resultados del teorema 2.19 se propone como cantidad de prueba la estadística:

$$U = S_N(\hat{0}_N) / S_{N+1}(\hat{0}_{N+1})$$

que debe aproximar una distribución $U(\frac{N-1}{2}, 1/2)$ para errores gaussianos si $N \rightarrow \infty$

5.8.2 Marcha aleatoria y filtro de Kalman

Un enfoque bastante más elegante es sugerido por P. Young (54). Supongamos otra vez el modelo (5.2), descrito vectorialmente por (5.6) donde el vector de parámetros 0 no permanece constante en el tiempo. Se supone que $e(t)$ es ruido blanco con $\text{var}(e(t)) = \sigma^2$.

Supongamos que 0_N satisface una ecuación matricial en diferencias de la forma:

$$0_N = \Phi 0_{N+1} + \Gamma q_{N-1} \quad (5.126)$$

donde $\Phi(N, N-1)$ es una matriz de transición $p \times p$, $\Gamma(N, N-1)$ es una matriz $(p \times r)$ de entrada, y q_{N-1} es un vector de r variables aleatorias que sucesivamente son independientes, con media cero y matriz de covarianzas Q

$$E[q_N] = 0, \quad E[q_N q_N^T] = Q \delta_{NN} \quad (5.127)$$

Suponiendo conocida la ley de variación (estocástica) (5.126) del parámetro θ_N , Young arriba a un algoritmo de estimación de θ_N de la forma predicción - corrección incorporando la información a priori contenida en (5.126).

llamando: P_N^A la matriz de covarianzas de $\hat{\theta}_N$, $\hat{\theta}_N$ el estimador de θ_N con la información en $t = N-1$ y la ley (5.126) a priori de las nuevas observaciones en $t = N$, y $P_{N/N-1}^A$ la matriz de covarianzas del estimador a priori $\hat{\theta}_{N/N-1}$, entonces:

predicción:

$$\begin{aligned} \hat{\theta}_{N/N-1} &= \Phi \hat{\theta}_{N-1} \\ P_{N/N-1}^A &= \Phi P_{N-1}^A \Phi^T + \Gamma Q \Gamma^T \end{aligned} \quad (5.128)$$

corrección:

$$\begin{aligned} \hat{\theta}_N &= \hat{\theta}_{N/N-1} - P_{N/N-1}^A x_N^T (\sigma^2 x_N^T P_{N/N-1}^A x_N + 1)^{-1} [x_N^T \hat{\theta}_{N/N-1} - y(N)] \\ P_N^A &= P_{N/N-1}^A - P_{N/N-1}^A x_N^T (\sigma^2 x_N^T P_{N/N-1}^A x_N + 1)^{-1} x_N P_{N/N-1}^A \end{aligned}$$

Obsérvese que mínimos cuadrados es un caso particular de las ecuaciones (5.128), en efecto: comparemos (5.128) con (5.112) cuando $\Phi = I_p$, $Q = 0$. Un caso muy importante es cuando (5.126) se transforma en una marcha aleatoria multidimensional:

$$\theta_N = \theta_{N-1} + q_{N-1} \quad (5.129)$$

En este caso el algoritmo se reduce a:

$$\begin{aligned} \hat{\theta}_N &= \hat{\theta}_{N-1} - (P_{N-1}^A + Q) x_N (\sigma^2 + x_N^T (P_{N-1}^A + Q) x_N)^{-1} [x_N^T \hat{\theta}_{N-1} - y(N)] \\ P_N^A &= (P_{N-1}^A + Q) - (P_{N-1}^A + Q) x_N (\sigma^2 + x_N^T (P_{N-1}^A + Q) x_N)^{-1} x_N^T (P_{N-1}^A + Q) \end{aligned} \quad (5.130)$$

Este algoritmo es bastante más versátil que el de olvido exponencial, pues la matriz Q se puede arreglar de forma que algunos parámetros puedan presentar variaciones mayores, y otras pequeñas, incluso nulas. La matriz Q puede ser matriz diagonal y sus elementos son esperanzas de velocidad de variación de los parámetros correspondientes.

P. Young, (54), señala que en la práctica el algoritmo general (5.128) es bueno para variaciones lentas de los parámetros. Cuando se espera que los parámetros varíen al mismo tiempo se propone como ley de variación un modelo de Gauss-Markov de primer orden

$$\theta_N = \beta \theta_{N-1} + q_{N-1} \quad (5.131)$$

donde ahora β es una matriz $p \times p$ diagonal, con elementos β_i tales que $-1 < \beta_i < 1$.

Otra forma de llegar a los resultados (5.128) y (5.130) es plantear las ecuaciones recursivas del filtro de Kalman para un sistema dinámico consistente con las leyes (5.126) ó (5.129) respectivamente. A modo de ejemplo lo haremos en el último caso.

Sea un sistema dinámico, en variables de estado:

$$\begin{cases} x(t+1) = x(t) + r(t) \\ y(t) = C x(t) + v(t) \end{cases}$$

$x(t) \in \mathbb{R}^p$ es el vector de estados que representa los parámetros del sistema a identificar. $v(t) = e(t)$ es ruido blanco. Definimos $C = C(t)$ como en (5.109), $r(t) \in \mathbb{R}^p$ es una sucesión de vectores cuyos elementos son variables aleatorias tales que

$$E[r(t)] = 0 \quad E[r(t)r(j)^T] = Q \delta_{tj}$$

Entonces las ecuaciones del filtro de Kalman quedan:

$$\begin{cases} \hat{x}(t) = \hat{x}(t-1) + k(t) [y(t) - C(t) \hat{x}(t-1)] \\ k(t) = [P(t-1) + Q] C^T(t) [C(t) \{P(t-1) + Q\} C^T(t) + \sigma^2]^{-1} \\ P(t) = \{P(t-1) + Q\} - k(t) C(t) [P(t-1) + Q] \end{cases} \quad (5.132)$$

Simplemente por observación se concluye en la identidad de las ecuaciones (5.130) y (5.132), tomando se cuenta la definición del vector fila $C(t)$.

5.9 Mínimos cuadrados repetidos

Supongamos que se quiere identificar un sistema dinámico lineal que responde a una ley:

$$A(q^{-1}) y(t) = B(q^{-1}) u(t) + n(t) \quad (5.133)$$

donde los polinomios $A(q^{-1})$ y $B(q^{-1})$ están definidos en (5.3) y (5.4), Si $n(t)$ no es una señal de ruido blanco, si la sucesión $n(t)$ está correlacionada en el tiempo, hemos visto en la sección 5.3 que el método de mínimos cuadrados da lugar a una estimación sesgada de los parámetros, aún cuando se conoce la estructura del modelo (en nuestro caso equivale a conocer n y m : los grados de A y B respectivamente).

Está claro que si repetimos el procedimiento para valores de m y n crecientes, la función de costo irá disminuyendo, puesto que hay más parámetros que ajustar. Planteemos dos problemas:

- ¿Cuándo abandonar la búsqueda de un mejor modelo de mayor orden?

- ¿Podemos a partir de dicha estimación conocer A y B ?

Supondremos a partir de ahora que $n(t)$ es una señal de espectro racional, gaussiana, que puede ser aproximada por la autorregresión:

$$n(t) = \frac{1}{11(q^{-1})} e(t) \quad (5.134)$$

$$11(q^{-1}) = 1 + h_1 q^{-1} + \dots + h_p q^{-p} \quad (5.135)$$

riendo $e(t)$ una señal de ruido blanco gaussiano de varianza $E\{e(t)^2\} = \sigma^2$. De (5.134) y (5.133) se desprende que en forma aproximada:

$$A(q^{-1}) H(q^{-1}) y(t) = b(q^{-1}) H(q^{-1}) u(t) + e(t) \quad (5.136)$$

Comparando (5.136) y (5.2) se observa que esta es la forma necesaria para la aplicación del procedimiento de mínimos cuadrados.

Si se aplica dicho procedimiento al sistema (5.136) se puede estimar sin sesgo los coeficientes de AH y bH . Sin embargo aún cuando se conocen n y m difícilmente se conoce l ; el procedimiento que se propone es:

repetir mínimos cuadrados para grados de $A' = AH$, $b' = bH$ crecientes, hasta que no haya una reducción significativa de la función de costo (suma de cuadrados de los residuos). Para detectar esto se propone el uso de la prueba F introducida en el capítulo 2 (teorema 2.18).

Bajo la hipótesis nula de que hemos llegado al modelo adecuado, incrementamos de n_1 a n_2 el no. de parámetros, entonces la cantidad:

$$F = \frac{\frac{E_1}{k_1} - \frac{E_2}{k_2}}{\frac{E_2}{k_2}} \cdot \frac{n_2 - n_1}{n_2 - n_1} \sim F(n_2 - n_1, n - n_2) \quad (5.137)$$

p. ej: hemos visto que si $n_2 - n_1 = 1$, $n - n_2 \geq 60$ $F_{0.95}(1, 60) = 4$ entonces alcanza $t \geq 4$ para rechazar la hipótesis nula con 5% de riesgo.

Hemos implantado a través de GROOT/REPE/C y GROOT/HEPE/F el esquema de mínimos cuadrados repetidos para sistemas con constante aditiva y sin constante aditiva respectivamente, se incluye en ellos la prueba sobre el orden del sistema.

Un caso particular es la determinación de los órdenes n_1, m del modelo de mínimos cuadrados, es decir, en el caso de residuos no correlacionados. En efecto: si $H(q^{-1}) = 1$ entonces la determinación de los grados de $A'(q^{-1})$ y $B'(q^{-1})$ equivale a encontrar n_1, m respectivamente. Para terminar de determinar la estructura del modelo se verá enseguida la búsqueda del retraso del sistema.

Debe notarse que hemos respondido a la primer pregunta que nos hacían, pero no a la segunda. Esta queda planteada en términos de: determinar A y B , estimados $A' = AH$ y $B' = bH$. A veces se puede determinar A y B buscando ceros comunes en A' y B' , no siempre resulta preciso el método: hay que tomar en cuenta que los coeficientes son aproximados y en la práctica suele ocurrir un fenómeno de "vanillitas" en \hat{A}' y \hat{B}' (estimados) donde hay una zona común de probable ocurrencia de ceros, a veces bastante extendida.

Este problema se resuelve con los procedimientos de mínimos cuadrados generalizados y máxima verosimilitud, que se verán más adelante.

5.10 Determinación del retardo

Otro problema frecuente en identificación es la determinación del no. de instantes de muestreo de retardo en la respuesta del sistema a identificar.

El procedimiento de mínimos cuadrados en forma repetida es útil en este sentido. Podemos reiterarlo para distintos tiempos de retraso, es decir, con la restricción de que los primeros coeficientes de $B(q^{-1})$ sean nulos.

Luego podemos reiterar la prueba F de la sección anterior para modelos donde se agrupen coeficientes de menor grado en el polinomio $B(q^{-1})$ (y no de mayor grado como antes).

Una sugerencia muy común en la literatura es acompañar la identificación con algún método de estimación de la sucesión de peso (respuesta en impulso) del sistema en cuestión. Normalmente la observación de dicha respuesta, aunque ésta sea aproximada, da buena idea de la estructura del modelo.

Una discusión más extensa acerca del uso de mínimos cuadrados repetidos para la determinación del retraso y la determinación de m y n se da al final del capítulo, sobre datos simulados en 5.12.11.

5.11 Aplicación a la estimación de la respuesta a impulso

Sea el sistema (5.2) con $e(t)$ ruido blanco de varianza σ^2 . La salida $y(t)$ se puede poner en la forma de suma de convolución, de acuerdo a (3.4):

$$y(t) = \sum_{i=0}^M u(t-i) h(i) \quad (5.138)$$

para señales de entrada causales

$$y(t) = \sum_{i=0}^t u(t-i) h(i) \quad (5.139)$$

Se quiere determinar la sucesión de peso $h(i)$, $i=0,1,2,\dots$. Una forma de hacerlo es identificar el sistema, para obtener los polinomios $\hat{A}(q^{-1})$, $\hat{B}(q^{-1})$, estimados de $A(q^{-1})$ y $B(q^{-1})$ respectivamente.

Si se realiza la expansión de $\hat{A}(q^{-1})/\hat{B}(q^{-1})$ en potencias crecientes de q^{-1} , la sucesión de coeficientes de q^{-i} , $i=0,1,2,\dots$ es la sucesión de peso $\hat{h}(i)$, $i=0,1,2,\dots$ del modelo estimado. En general resulta una buena estimación de la sucesión de peso del sistema. Veamos 2 formas más de estimar la sucesión de peso:

5.11.1 Entradas no aleatorias

Aproximemos el sistema como un filtro FIR (respuesta a impulso finita), es decir, que existe algún K tal que $h(i) = 0 \forall i > K$, entonces

$$Y_N = \sum_{i=0}^K U_N h(i) + u_N \quad (5.140)$$

donde

$$Y_H = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(H) \end{bmatrix}; U_H = \begin{bmatrix} u(1) & u(0) & \dots & u(1-k) \\ u(2) & u(1) & \dots & u(2-k) \\ \vdots & \vdots & \ddots & \vdots \\ u(H) & u(H-1) & \dots & u(H-k) \end{bmatrix}; h = \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(k) \end{bmatrix}; a_H = \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(H) \end{bmatrix}$$

La ecuación (5.140) tiene la misma forma que (5.6), y el vector $h(y)$ por lo tanto la sucesión de peso, puede ser estimado por mínimos cuadrados de la misma manera que el vector θ :

$$\hat{h}_H = (U_H^T U_H)^{-1} U_H^T Y_H \quad (5.141)$$

para ello basta con aplicar cualquiera de los métodos ya descritos.

Una desventaja de este método es que si se quiere aproximar la respuesta a impulso en un número importante de puntos, el problema de mínimos cuadrados tiene la misma dimensión.

5.11.2 Entradas aleatorias

Si es posible computar en forma aproximada las funciones de correlación de entrada-salida y autocorrelación de la entrada, será posible estimar la sucesión de peso. Suponemos como antes que $h(k)$ tal que $h(i) = 0 \forall i > k$. Luego, por la ecuación de Wiener-Kolmogorov (ecuación 3.26):

$$R_{uy}(\tau) = \sum_{i=0}^K h(i) R_{uu}(\tau-i)$$

Si se computan las estimaciones de las funciones de correlación: $\hat{R}_{uy}(\tau)$, $\hat{R}_{uu}(\tau)$, $\tau = 0, 1, 2, \dots, N$ para $N \gg K$, entonces se tiene un sistema de ecuaciones sobredeterminado. Supondremos que en cada ecuación se comete un error $e^i(\tau)$, $\tau = 0, 1, \dots, N$, entonces:

$$R_{uy} = R_{uu} h + e^i \quad (5.142)$$

donde R_{uy} , R_{uu} , e^i se definen como:

$$R_{uy} = \begin{bmatrix} \hat{R}_{uy}(0) \\ \hat{R}_{uy}(1) \\ \vdots \\ \hat{R}_{uy}(N) \end{bmatrix}; R_{uu} = \begin{bmatrix} \hat{R}_{uu}(0) & \hat{R}_{uu}(-1) & \dots & \hat{R}_{uu}(-k) \\ \hat{R}_{uu}(1) & \hat{R}_{uu}(0) & \dots & \hat{R}_{uu}(1-k) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{uu}(N) & \hat{R}_{uu}(N-1) & \dots & \hat{R}_{uu}(N-k) \end{bmatrix}; h = \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(k) \end{bmatrix}; e^i = \begin{bmatrix} e^i(0) \\ e^i(1) \\ \vdots \\ e^i(N) \end{bmatrix}$$

Nótese que los elementos $\hat{R}_{uu}(\tau)$ en R_{uu} con argumentos negativos no es necesario computarlos puesto que $\hat{R}_{uu}(-\tau) = \hat{R}_{uu}(\tau)$.

Al modelo (5.142) se le puede aplicar el método de mínimos cuadrados obteniendo la estimación:

$$\hat{h} = (R_{uu}^T R_{uu})^{-1} R_{uu}^T R_{uy} \quad (5.143)$$

La desventaja de este procedimiento, como el caso anterior, es la dimensión del problema de mínimos cuadrados que hay que resolver. A ella se le agrega otra: antes hay que computar las correlaciones.

Los ejemplos sobre sistemas simulados pueden verse en la sección 5.12.10.

donde

$$Y_H = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}; U_H = \begin{bmatrix} u(1) & u(0) & \dots & u(1-k) \\ u(2) & u(1) & \dots & u(2-k) \\ \vdots & \vdots & \ddots & \vdots \\ u(N) & u(N-1) & \dots & u(N-k) \end{bmatrix}; h = \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(k) \end{bmatrix}; u_H = \begin{bmatrix} u(1) \\ u(2) \\ \vdots \\ u(N) \end{bmatrix}$$

La ecuación (5.140) tiene la misma forma que (5.6), y el vector h y por lo tanto la sucesión de peso, puede ser estimado por mínimos cuadrados de la misma manera que el vector θ :

$$\hat{h}_H = (U_H^T U_H)^{-1} U_H^T Y_H \quad (5.141)$$

para ello basta con aplicar cualquiera de los métodos ya descritos.

Una desventaja de este método es que si se quiere aproximar la respuesta a impulso en un número importante de puntos, el problema de mínimos cuadrados tiene la misma dimensión.

5.11.2 Entradas aleatorias

Si es posible computar en forma aproximada las funciones de correlación de entrada-salida y autocorrelación de la entrada, será posible estimar la sucesión de peso. Suponemos como antes que $h(k)$ tal que $h(i) = 0 \forall i > k$. Luego, por la ecuación de Wiener-Kolmogorov (ecuación 3.26):

$$R_{uy}(\tau) = \sum_{i=0}^k h(i) R_{uu}(\tau-i)$$

Si se computan las estimaciones de las funciones de correlación: $\hat{R}_{uy}(\tau)$, $\hat{R}_{uu}(\tau)$, $\tau = 0, 1, 2, \dots, N$ para $N \gg k$, entonces se tiene un sistema de ecuaciones sobredeterminado. Supondremos que en cada ecuación se comete un error $e^i(\tau)$, $\tau = 0, 1, \dots, N$, entonces:

$$R_{uy} = R_{uu} h + e^i \quad (5.142)$$

donde R_{uy} , R_{uu} , e^i se definen como:

$$R_{uy} = \begin{bmatrix} \hat{R}_{uy}(0) \\ \hat{R}_{uy}(1) \\ \vdots \\ \hat{R}_{uy}(N) \end{bmatrix}; R_{uu} = \begin{bmatrix} \hat{R}_{uu}(0) & \hat{R}_{uu}(-1) & \dots & \hat{R}_{uu}(-k) \\ \hat{R}_{uu}(1) & \hat{R}_{uu}(0) & \dots & \hat{R}_{uu}(1-k) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{uu}(N) & \hat{R}_{uu}(N-1) & \dots & \hat{R}_{uu}(N-k) \end{bmatrix}; h = \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(k) \end{bmatrix}; e^i = \begin{bmatrix} e^i(0) \\ e^i(1) \\ \vdots \\ e^i(N) \end{bmatrix}$$

Nótese que los elementos $\hat{R}_{uu}(\tau)$ en R_{uu} con argumentos negativos no es necesario computarlos puesto que $\hat{R}_{uu}(-\tau) = \hat{R}_{uu}(\tau)$.

Al modelo (5.142) se le puede aplicar el método de mínimos cuadrados obteniendo la estimación:

$$\hat{h} = (R_{uu}^T R_{uu})^{-1} R_{uu}^T R_{uy} \quad (5.143)$$

La desventaja de este procedimiento, como el caso anterior, es la dimensión del problema de mínimos cuadrados que hay que resolver. A ella se le agrega otra: antes hay que computar las correlaciones.

Los ejemplos sobre sistemas simulados pueden verse en la sección 5.12.10.

5.12 Pruebas con Datos Simulados

En esta sección se describen algunas de las pruebas realizadas para comprobar que se cumplen las propiedades asintóticas del método de los mínimos cuadrados.

Todo el volumen del presente trabajo no alcanzaría para describir todas y cada una de las pruebas con datos simulados que se realizaron durante su desarrollo. Por otro lado estas pruebas son redundantes en el sentido de que se observan resultados similares; hemos preferido, siguiendo los consejos de los doctores J. L. Farah y A. Alonso Concheiro presentar algunos ejemplos claros y representativos.

5.12.1 Las simulaciones

Se eligieron ejemplos de sistemas cuya identificación condujera a resultados sencillos, fácilmente analizables y visualizables. Para que estos resultados pudieran ser comparados, se optó por utilizar un mismo conjunto de señales en las simulaciones.

Para verificar tendencias de convergencia, se hicieron simulaciones con las mismas características y se promediaron los resultados obtenidos. Los procesos simulados son de la forma (5.2);

$$A(q^{-1})y(t) = B(q^{-1})u(t) + e'(t), \text{ simulados}$$

con el programa SIMULA, escrito en PASCAL para la computadora B-6800 del C.S.C. de la UNAM; dicho programa puede encontrarse en el Apéndice D.

El polinomio $B(q^{-1}) = b_0 q^{-1} + \dots + b_m q^{-(m+1)}$ no contiene términos de transmisión directa; la señal $e'(t)$ fue compuesta como:

$$e'(t) = \sigma e(t) \quad (5.144)$$

con $e(t)$ ruido blanco de media cero y varianza unitaria. De esta forma la perturbación $e'(t)$ es ruido blanco de media cero y varianza σ^2 , siendo σ^2 otro parámetro factible de ser estimado.

Se eligieron tres sistemas discretos lineales de primero, segundo y tercer orden respectivamente; estos fueron excitados con entradas PRBS127 y se perturbaron con las señales RUID01, RUID02, RUID03, RUID04, RUID05, RUID06, y RUID07, para distintos valores de la relación señal a ruido (S/N).

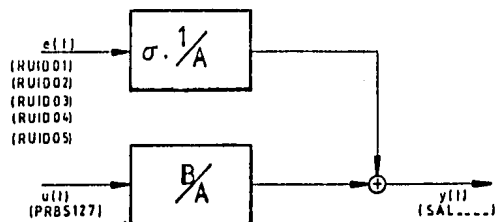


Fig 5.3 Sistema simulado

Los sistemas mencionados son:

- Sistema de primer orden

$$(1-0.8 q^{-1}) y(t) = (q^{-1}) u(t) + \sigma e(t) \quad (5.145)$$

para

$$\left\{ \begin{array}{ll} S/N = \infty & \sigma = 0 \\ S/N = 10 & \sigma = 1/\sqrt{10} = 0.316228 \\ S/N = 4 & \sigma = 1/2 \\ S/N = 1 & \sigma = 1 \\ S/N = 0.1 & \sigma = \sqrt{10} = 3.16228 \end{array} \right.$$

- Sistema de segundo orden

$$(1-1.5 q^{-1}+0.7 q^{-2}) y(t) = (q^{-1}+0.5 q^{-2}) u(t) + \sigma e(t) \quad (5.146)$$

para

$$\left\{ \begin{array}{ll} S/N = \infty & \sigma = 0 \\ S/N = 10 & \sigma = 0.46177 \\ S/N = 4 & \sigma = 0.73013 \\ S/N = 1 & \sigma = 1.46026 \\ S/N = 0.1 & \sigma = 4.61774 \end{array} \right.$$

- Sistema de tercer orden

$$(1-1.9q^{-1}+1.3q^{-2}-0.3q^{-3}) y(t) = (q^{-1}-0.5q^{-2}+0.3q^{-3}) u(t) + \sigma e(t) \quad (5.147)$$

para

$$\left\{ \begin{array}{ll} S/N = \infty & \sigma = 0 \\ S/N = 10 & \sigma = 0.240503 \\ S/N = 4 & \sigma = 0.38027 \\ S/N = 1 & \sigma = 0.76054 \\ S/N = 0.1 & \sigma = 2.40503 \end{array} \right.$$

El cálculo de S/N se hizo con el algoritmo que se describe en el párrafo 4.3.2.

Para las pruebas donde se promediaron los resultados se utilizó como entrada la señal PRG6127, y distintas señales de ruido.

En las siguientes tablas se presentan las simulaciones realizadas bajo estas condiciones y distintos valores de S/N:

Simulaciones con el sistema de primer orden:

S/N \ Perturbación	∞	10	4	1	0.1
RUIDO1	SAL101	SAL102	SAL103	SAL104	SAL105
RUIDO2	-	SAL202	SAL203	SAL204	SAL205
RUIDO3	-	SAL302	SAL303	SAL304	SAL305
RUIDO4	-	SAL402	SAL403	SAL404	SAL405
RUIDO5	-	SAL502	SAL503	SAL504	SAL505

Estas señales fueron generadas a partir de condiciones iniciales nulas. La señal SAL505 con 1127 elementos, el resto con 645 elementos. Luego se eliminaron los primeros 127 elementos de cada señal como una forma de eliminar el efecto de condiciones iniciales particulares y desconocimiento de $e(t)$ ($t < 0$).

- Simulaciones con el sistema de segundo orden:

S/N Perturbación	∞	10	4	1	0,1
RUIDO1	SALA101	SALA102	SALA103	SALA104	SALA105
RUIDO2	-	-	-	SALA204	-
RUIDO3	-	-	-	SALA304	-
RUIDO4	-	-	-	SALA404	-
RUIDO5	-	-	-	SALA504	SALA505

Como antes estas señales fueron generadas a partir de condiciones iniciales nulas, con 1127 elementos SALA505, y 645 el resto. Fueron suprimidos los primeros 127 elementos de cada señal.

- Simulaciones con el sistema de tercer orden

Perturbación: RUIDO5

S/N	∞	10	1	0,1
Señal:	SALC501	SALC502	SALC504	SALC505

Como antes, las señales se generaron a partir de condiciones iniciales nulas, con 1127 elementos cada una. Se suprimieron los primeros 127 elementos.

Otras señales utilizadas en las pruebas fueron generadas especialmente para alguna prueba particular, por ejemplo con perturbaciones ENT1 o RUIDO7, solo que las señales antes mencionadas fueron usadas ampliamente, en casi todas las pruebas.

De todos modos, al describir las pruebas se mencionará la fuente de perturbación, y la relación señal a ruido.

Para visualizar el efecto del ruido en las señales de salida, se muestran en las figuras (5.4), (5.5) y (5.6) las salidas de los sistemas de primer, segundo y tercer orden para distintos valores de la relación señal a ruido S/N.

La entrada en los tres casos es la señal PRB: 127 y la fuente de perturbación es una simulación de normal con media cero, y varianza de acuerdo al valor de S/N.

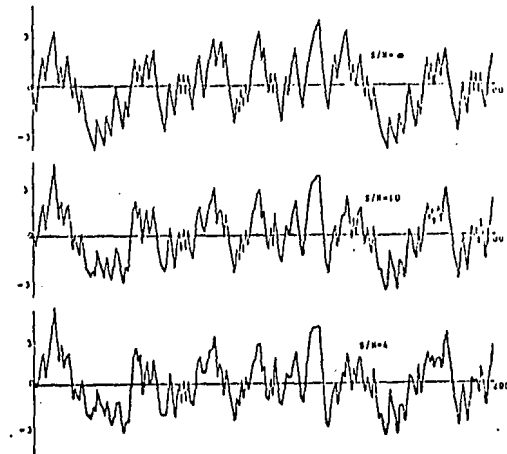


Fig 5.4 Salida del sistema de primer orden para S/N = ∞ , 10, 1, 0,1

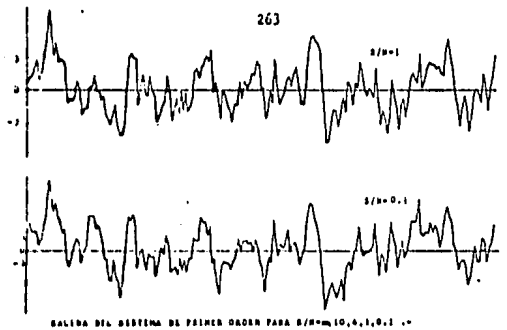


Fig 5.4 Continuación

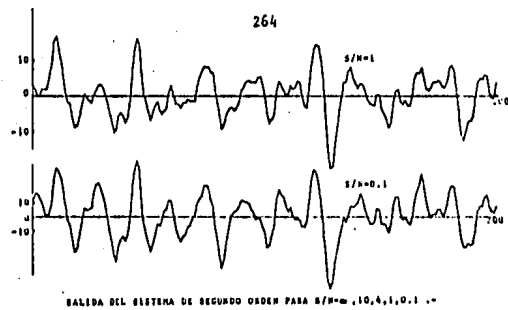


Fig 5.5 Continuación

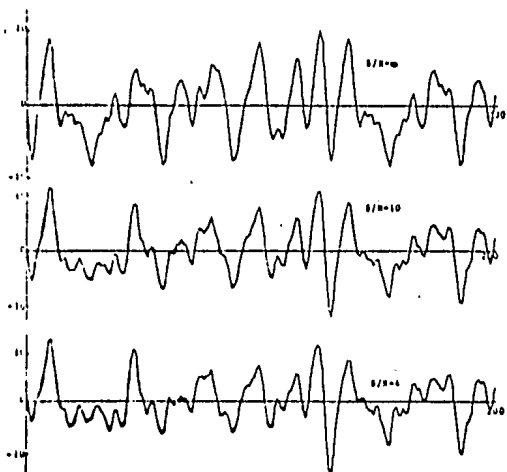


Fig 5.5 Salida del sistema de segundo orden para $S/N = \infty, 10, 1, 0.1$.

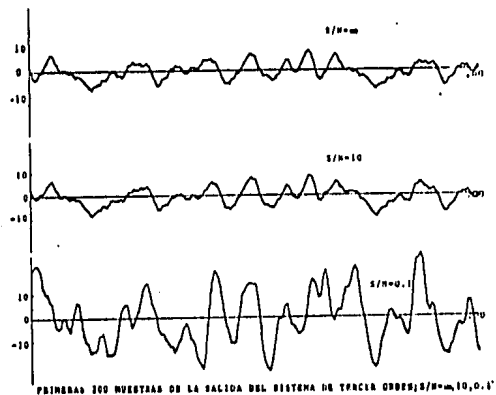


Fig 5.6 Primeras 200 muestras de la salida del sistema de tercer orden; $S/N = \infty, 10, 0.1$.

La teoría provee la magnitud de la varianza del estimador de mínimos cuadrados como una medida de bondad del estimador. Distingamos tres cantidades:

- La desviación standard prevista por la teoría
- La desviación standard estimada en la identificación
- La desviación standard encontrada como desviación cuadrática media a lo largo de muchas identificaciones

Comparamos estas tres cantidades en distintas partes de las siguientes secciones.

En la próxima sección se entrelazan las tres en un ejemplo donde se estudia la ubicación de los estimadores de los parámetros respecto al supuesto teórico de concentración, para el sistema de primer orden.



La relación 2 está implícita en el párrafo que corresponde a intervalo de estimación.

La relación 3 es el núcleo del párrafo donde se estudian las propiedades de convergencia promediando los resultados de varias simulaciones.

La relación 1 se estudia en el párrafo denominado: estimación de las varianzas de los parámetros.

Conviene tener presente siempre estas tres cantidades para evitar confusiones y distinguir lo que se estudia en cada párrafo.

5.12.2 Precisión e Intervalo de estimación

En las simulaciones de los tres sistemas con la entrada RUID127 y las perturbaciones RUID01, ..., RUID07 se cumplen las hipótesis de los teoremas 5.1, 5.2, y 5.3 tomando en cuenta que los sistemas no están realimentados (si lo estuvieran habría que echar mano al teorema 5.3).

De acuerdo al teorema 5.1 el estimador $\hat{\theta}_N$ de mínimos cuadrados cumple:

$$\begin{aligned} \hat{\theta}_N &\xrightarrow{\text{m.c.}} \theta \\ \hat{\theta}_N &\xrightarrow{\text{ley}} \mathcal{N}^n(0, \sigma^2 R^{-1}) \end{aligned}$$

donde $R = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i x_i^T = E[x_i x_i^T]$ (pues u, e son procesos estocásticos ergódicos en nuestras simulaciones).

$$R = \begin{bmatrix} E\{y^2} & 0 \\ 0 & E\{u^2} \end{bmatrix} \quad (5.148)$$

para el sistema de 1er orden

$$J_N^{-1} = \frac{\partial \hat{\theta}_N^{-1}}{\partial \theta} = \begin{bmatrix} \frac{\sigma^2}{NE\{y^2}} & 0 \\ 0 & \frac{\sigma^2}{NE\{u^2}} \end{bmatrix} \quad (5.149)$$

Si las perturbaciones $e(t)$ son además gaussianas J_N^{-1} es la matriz de

información de Fisher. Para el sistema de primer orden propuesto:

$$\sigma^T = (-0.8, 1) = (a_1, b_0) \quad (5.150)$$

como usaremos entradas con $E u^2 = 1$, con el algoritmo del capítulo 4

se calculó la varianza de la parte determinística del sistema

$$E y^2 = 2.7778$$

La varianza de la salida del sistema es:

$$E y^2 = 2.7778 (1 + \sigma^2)$$

entonces la matriz de covarianzas (asintótica) J_N^{-1} queda:

$$J_N^{-1} = \begin{bmatrix} \frac{\sigma^2}{(1+\sigma^2)} & \frac{1}{2.7778 N} & 0 \\ 0 & 0 & \frac{\sigma^2}{N} \end{bmatrix} \quad (5.151)$$

Los elementos diagonales son las varianzas de los parámetros: $\sigma_{a_1}^2$ y $\sigma_{b_0}^2$ respectivamente. La Tabla 5.1 presenta los valores de estas varianzas para distintos números de muestras y relaciones señal a ruido (para el sistema propuesto puede verse que $\sigma^2 = 1/(S/N)$)

Tabla 5.1. Varianzas teóricas de los parámetros del sistema de primer orden para $N = 50, 100, 150, 300, 500$ y $S/N = 10, 4, 1, 0.1$

S/N \ N	10	4	1	0.1
50 $\sigma_{a_1}^2$	0.0006545	0.0014400	0.0036000	0.0065455
50 $\sigma_{b_0}^2$	0.0020000	0.0050000	0.0200000	0.2000000
100 $\sigma_{a_1}^2$	0.0003273	0.0007200	0.0018000	0.0032727
100 $\sigma_{b_0}^2$	0.0010000	0.0025000	0.0100000	0.1000000
150 $\sigma_{a_1}^2$	0.0002182	0.0004800	0.0012000	0.0021818
150 $\sigma_{b_0}^2$	0.0006667	0.0016667	0.0066667	0.0666667
300 $\sigma_{a_1}^2$	0.0001091	0.0002400	0.0006000	0.0010909
300 $\sigma_{b_0}^2$	0.0003333	0.0008333	0.0033333	0.0333333
500 $\sigma_{a_1}^2$	0.0000655	0.0001440	0.0003600	0.0006545
500 $\sigma_{b_0}^2$	0.0002000	0.0005000	0.0020000	0.0200000

Nótese que para grandes valores de σ^2 el efecto del ruido es muy perjudicial para la estimación de b_0 , pero no para la de a_1 donde su varianza $\sigma_{a_1}^2 \rightarrow \frac{1}{2.7778N}$ para $\sigma^2 \rightarrow \infty$.

Para una relación señal a ruido $S/N = 1$, se estimaron los parámetros del sistema de primer orden sobre cinco realizaciones con las perturbaciones RUID01, RUID02, RUID03, RUID04, RUID05.

Se efectuaron estimaciones para 100, 200 y 500 intervalos de observación. Los resultados obtenidos se muestran graficados en las figuras (5.7), (5.8) y (5.9) respectivamente, junto con los elipsoides teóricos de concentración:

$$(\hat{\theta}_N^T - \theta^T) J_N^{-1} (\hat{\theta}_N - \theta) = 4$$

Los resultados observados son ampliamente satisfactorios, tanto por el hecho de que todos los estimados caen dentro del elipsoide correspondiente, como que para valores crecientes de N esa concentración parece ser mayor, en elipsoides obviamente más reducidos.

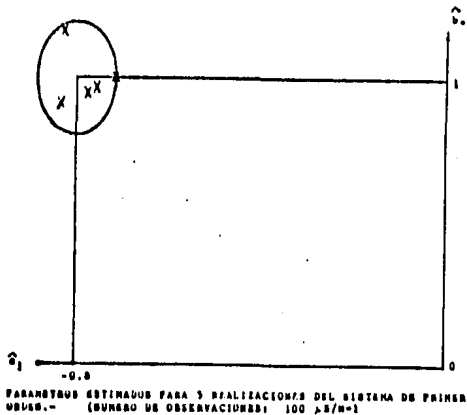


Fig 5.7.

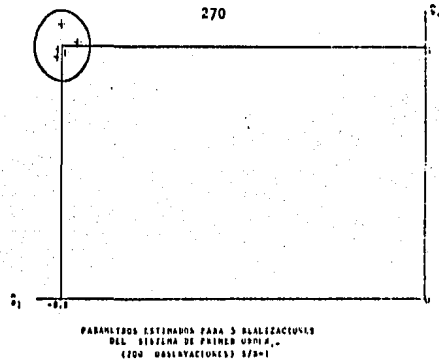


Fig 5.8

Plantearnos ahora el problema inverso, es decir, conocido el valor estimado ¿cómo se ubica el parámetro real respecto al primero?. Dicho de otra forma: dado el valor estimado del parámetro, y el elipsoide de concentración estimado ¿cómo se ubica el parámetro real respecto de dicho elipsoide?.

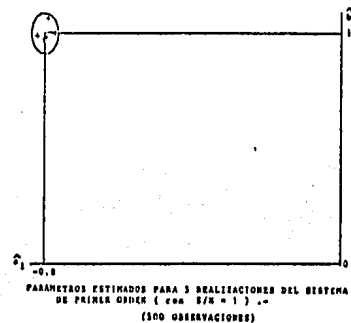


Fig 5.9

En la figura (5.10) se dibujan para una relación señal a ruido $S/N = 1$ y 100 observaciones los elipsoides de concentración estimados en cinco realizaciones. Estos se estimaron con el estimador $\hat{\sigma}^2 (X^T X)^{-1}$, de la matriz de covarianzas y se centró cada elipsoide en el parámetro estimado correspondiente.

Junto a los cinco elipsoides estimados, fue dibujado en línea más gruesa el elipsoide de concentración teórico.

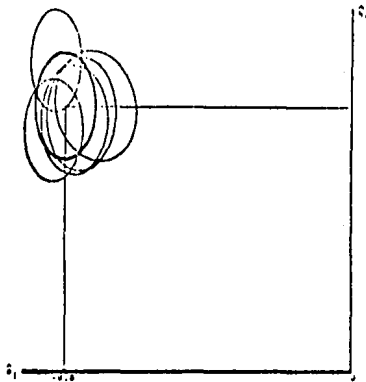


Fig 5.10 Elipsoides de concentración estimados para 5 realizaciones del sistema de primer orden. (100 observaciones, $S/N=1$)

A pesar de que $N = 100$ no es un número muy grande de observaciones, si tenemos en cuenta que la relación señal a ruido es 1, otra vez los resultados son muy satisfactorios. Podemos observar varias cosas:

- Todos los elipsoides contienen en su interior al valor real del parámetro estimado: $[-0.8, 1]^T$

- Eso sucede aún para los elipsoides correspondientes a los estimados más alejados (ver fig 5.7)
- Los elipsoides estimados (y por lo tanto la matriz de covarianzas) son muy similares al elipsoide (y por lo tanto a la matriz de covarianzas teórica) teórico.

Esto sucede en la orientación, lo que implica que la matriz de covarianza tiende a ser diagonal, y en sus dimensiones. Se desprende inmediatamente que los autovalores que son las varianzas de los parámetros están estimando bien las varianzas teóricas.

- Por último como son de dimensiones semejantes, también la varianza generalizada de Wilks es semejante.

No dibujamos los experimentos análogos para 200 y 500 observaciones, pues los resultados son semejantes y los elipsoides son naturalmente más reducidos y concentrados.

Las figuras de esta sección se obtuvieron con el programa ELLIPSOIDE que dibuja los elipsoides de concentración pasando al cálculo de la forma polar, este programa se encuentra en el Apéndice D.

Intervalo de estimación

Como los estimadores tienden a la distribución normal multivariada ya mencionada, podemos basarnos en las mismas consideraciones que el teorema 2.16 para construir intervalos de estimación del parámetro real con un nivel de confianza determinado.

En efecto, como $\hat{O}_n \xrightarrow{\text{dist}} N(O, \frac{\sigma^2 K^{-1}}{n})$, pudiéndose estimar $\frac{K^{-1}}{n}$ con $(X^T X)^{-1}$, podemos estimar cov \hat{O}_i (correspondiente al i-ésimo parámetro - pero con el elemento i-ésimo de la diagonal de $\sigma^2(X^T X)^{-1}$): llamémosle $\sigma^2 a_{ii}$.

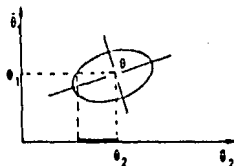
Obtenemos entonces un intervalo de estimación:

$$\hat{O}_i \pm K \sigma \sqrt{a_{ii}}$$

donde K es un factor a seleccionar del mismo modo que en los comentarios al teorema 2.15. Esto supone conocida σ^2 , pero si no se conoce σ^2 , y se lo estima con el estimador S_y^2 de (5.27) se puede construir un intervalo para θ con la distribución t de Student como en el teorema 2.16.

Los valores de K para distintos niveles de confianza están en las tablas 2.II y 2.III para las distribuciones normal y t respectivamente. Cabe agregar que estos intervalos son algo sobrados en el sentido que un nivel de confianza vale para estimaciones independientes, lo que no es en general nuestro caso.

Podrían obtenerse intervalos más finos diagonalizando la matriz de covarianzas (pero estos intervalos serían para combinaciones lineales de los parámetros)



Como ejemplo presentaremos resultados con datos simulados con entrada PRBS 127.

Sistema de primer orden

Para este sistema (descrito por (5.145)) es $p = 2$ (n2 de parámetros). Si se estima con 122 observaciones, la tabla 2.III señala para $n-p = 120$ y un nivel de confianza de 90% un intervalo:

$$\hat{O}_i \pm 1.658 v_i \quad (5.152)$$

donde v_i^2 es la varianza estimada de \hat{O}_i estimado σ^2 con S_y^2 de (5.27).

Se identificó el sistema con 122 observaciones, de realizaciones con relación señal a ruido $S/N = 4$ y $S/H = 1$. El método empleado fue de mínimos cuadrados y en la tabla 5.II, se muestran los valores estimados de los parámetros \hat{O}_i junto con la estimación v_i de su desviación Standard.

S/N	$\hat{a}_1(-0.8)$		$\hat{a}_0(1.0)$		$\hat{\sigma}$	(real)	PUNTO
n	-0.77308	0.02854	0.99113	0.06746	0.52206	0.5	PUNTO1
n	-0.81313	0.02376	1.06663	0.04668	0.49126	0.5	PUNTO2
n	-0.78132	0.02966	0.97642	0.05064	0.55654	0.5	PUNTO3
1	-0.76985	0.04570	0.95406	0.10124	0.11298	1.0	PUNTO3
1	-0.77535	0.04665	0.95192	0.09093	0.09033	1.0	PUNTO4
1	-0.81181	0.03848	0.88042	0.08705	0.05702	1.0	PUNTO5

Tabla 5. II Estimación de los parámetros del sistema de primer orden con 122 muestras, y sus desviaciones. Relación señal a ruido $S/N = 4$, $S/H = 1$.

Puede verificarse que en todos los casos el valor real del parámetro θ_i encuentra dentro del intervalo (5.152).

Sistema de segundo orden

Este sistema (con ecuación en diferencias (5.146)) tiene cuatro parámetros, $p = 4$. Luego, si se realiza la identificación para $N = 124$ observaciones vale el mismo intervalo (5.152).

Se efectuó la identificación por mínimos cuadrados para tres realizaciones de este sistema con $S/N = 1$. Los resultados se muestran en la tabla 5.III.

Tabla 5.III Estimación de los parámetros del sistema de segundo orden con 124 observaciones y una relación señal a ruido $S/N = 1$.

Parámetro valor real	a_1 -1.5	desv. 0.7	a_2 0.7	desv. 0.5	b_0 1.0	desv. 1.0	b_1 0.5	desv. 0.5	J	Ruido
	-1.50000	0.05276	0.77933	0.05361	1.03737	0.13928	0.47134	0.14097	1.50677	RU1D01
	-1.05000	0.05150	0.65004	0.05181	1.17026	0.13146	0.51272	0.13376	1.48066	RU1002
	-1.40500	0.05142	0.64201	0.05091	.95236	0.13300	0.51821	0.14237	1.46223	RU1D03

También en este caso el valor real de cada parámetro cae dentro del intervalo estimado, de 90% de confianza.

Sistema de tercer orden

Para este sistema (de ecuación (5.147)) es $p = 6$. Se efectuaron identificaciones sobre realizaciones del sistema perturbado con ruido y relaciones señal a ruido $S/N = 10, 1, 0.1$. Se consideran 500 observaciones.

Como $N - p = 494 \gg 120$ se ha aproximado la distribución t con una normal, utilizándose para elegir el intervalo de estimación el mismo renglón de la tabla 2.III. El intervalo es:

$$\hat{\theta}_i \pm 1.645 v_i \quad (5.153)$$

para 90% de confianza. Los resultados se listan en la tabla 5. IV.

Parámetro	$S/N=10$	$S/N=1$	$S/N=0.1$	valor real
a_1	-1.50000	-1.50000	-1.50000	-1.5
a_2	0.77933	0.77933	0.77933	0.7
b_0	1.03737	1.03737	1.03737	1.0
b_1	0.47134	0.47134	0.47134	0.5
a_3	0.05276	0.05276	0.05276	0.1
a_4	0.05361	0.05361	0.05361	0.1
a_5	0.05150	0.05150	0.05150	0.1
a_6	0.05181	0.05181	0.05181	0.1
b_2	0.13928	0.13928	0.13928	0.1
b_3	0.14097	0.14097	0.14097	0.1
b_4	0.13376	0.13376	0.13376	0.1
b_5	0.14237	0.14237	0.14237	0.1
v	1.50677	1.50677	1.50677	1.50677

La forma de la tabla 5.IV se usará repetidamente pues es el formato de impresión de resultados de los programas HICU/S, HICU/C, SROOT/S, SROOT/C, SROOT/REPL/S, SROOT/REPL/C. Al lado del nombre del parámetro aparece el estimado, y a continuación su desviación Standard estimada.

En la línea que se indica V (SU para los programas SPOOT/REPE/S y SPOOT/REPE/C) se imprime el valor de la función de costo (suma de los cuadrados de los residuos), y a continuación la estimación de la desviación Standard de las perturbaciones.

Como puede verse en la tabla 5.1V que también en este caso, todas las veces el valor real de los parámetros cae dentro del intervalo de 90% de confianza (5.153).

5.12.3 Propiedades asintóticas y efecto del ruido

En este párrafo se describen las pruebas y sus resultados para comprobar las propiedades asintóticas del estimador de mínimos cuadrados, y la influencia del nivel de ruido sobre éste.

Primero veremos algunos ejemplos de identificación de sistemas a medida que llegan nuevos datos de entrada y salida. Las figuras (5.11) a (5.14) muestran la evolución del estimador de mínimos cuadrados para N creciente, conforme se incorporan nuevos datos a la estimación. Estas figuras muestran la identificación para distintas realizaciones de los sistemas de primer, segundo y tercer orden con distintas relaciones señal a ruido.

En el eje de abscisas se ha colocado el número de observaciones y en ordenadas los valores de la estimación de los parámetros para cada número de observaciones.

- Sistema de primer orden

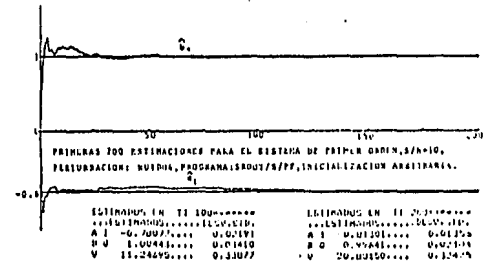


Fig 5.11

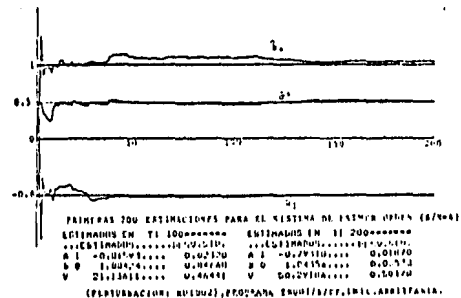


Fig 5.12

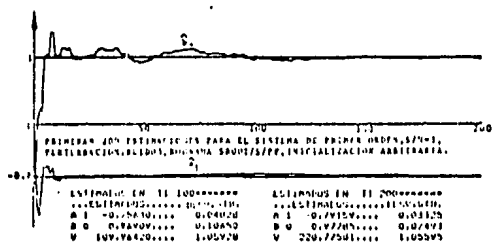


Fig 5.13

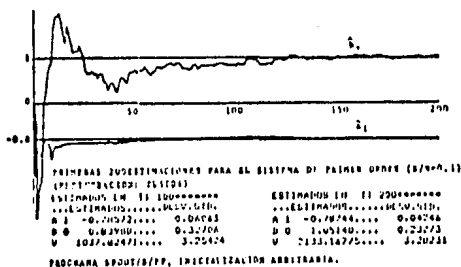


Fig 5.14

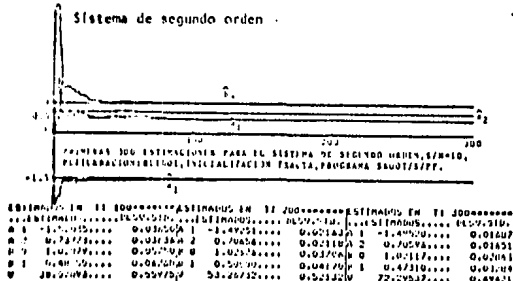


Fig 5.15

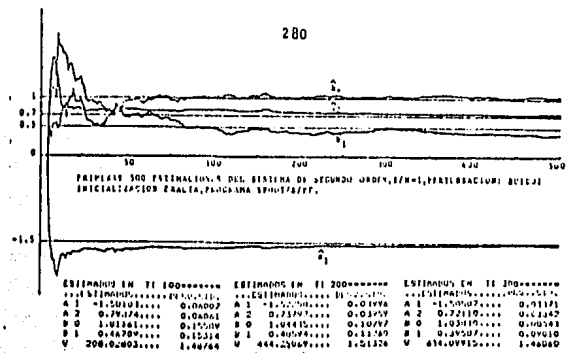


Fig 5.16

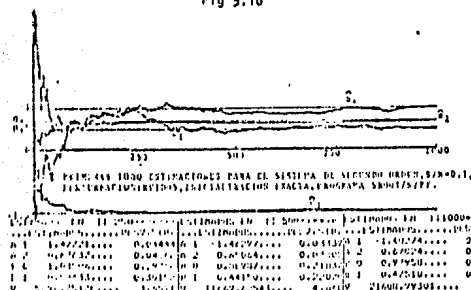


Fig 5.17

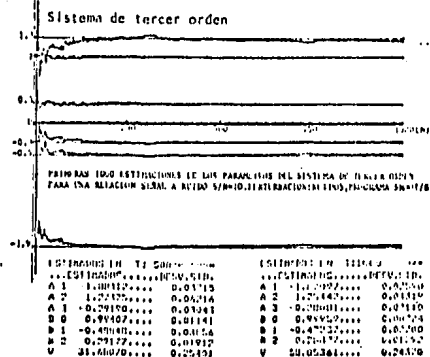


Fig 5.18

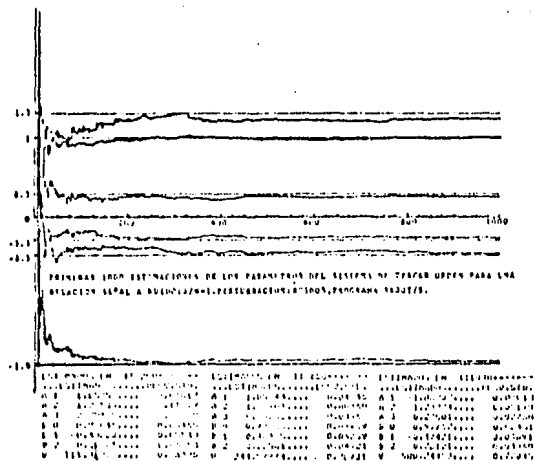


Fig 5.19

En la figura (5.12) fue agregada la evolución del estimador de la varian-
za del ruido (σ^2).

De la observación de las figuras (5.11) a (5.19) pueden comprobarse al-
gunos efectos esperados:

- Se muestra en todos los casos la tendencia del estimador de mínimos cuadrados de converger al valor real de los parámetros
- Quanto menor es el nivel de ruido (mayores valores de γ/σ^2), el valor estimado es más próximo al valor real y tiene menos oscilaciones.

Recordemos que se espera que el estimador tenga una varianza inversamen-
te proporcional al número de observaciones, y que dicha varianza sea -
creciente con el nivel de la señal a ruido.

Con esta última afirmación hay que ser cuidadosos, no se espera que la
varianza del estimador sea proporcional a σ^2 , y las figuras corroboran
esta cuestión.

En efecto: la matriz J_{11}^{-1} de covarianzas esperadas (5.151) del sistema
de primer orden prevé una varianza de la estimación del parámetro b_1
proporcional a σ^2 , en cambio para a_1 esta varianza es proporcional a
 $\sigma^2/(10^2)$. Esto hace que disminuyamos γ/σ^2 de 10 a 0.1 la varianza de
 b_1 cruza 100 veces y la de a_1 10 veces.

Las figuras (5.11) a (5.14) confirman esta previsión: al disminuir la
relación señal a ruido la estimación de b_1 desmejora notablemente en
comparación con la estimación de a_1 , cuya precisión no mejora mucho a
simple vista. Más adelante confirmaremos con más simulaciones que la

varianza de a_1 sigue la ley prevista por la teoría.

Una situación similar sucede con la estimación de los parámetros del sistema de segundo orden. Poco más adelante, en las ecuaciones (5.161), la teoría indica una varianza de \hat{b}_0 , \hat{b}_1 de 7 a 8 veces mayor que la de \hat{a}_1 , \hat{a}_2 para una relación señal a ruido $S/N = 1$, en la figura (5.16) puede apreciarse que en general la estimación de a_1 y a_2 es algo mejor que la de b_0 y b_1 .

Todo esto nos lleva a concluir que dada la estructura del sistema, y para entradas y perturbaciones concordantes con los teoremas de convergencia, la observación de la matriz $\sigma^2 E(\kappa_i \kappa_i^T)$ es vital para entender donde podemos y donde no podemos esperar precisión. Más aún, parecería ser (no lo hemos estudiado en profundidad), que el estudio de la matriz de covarianzas podría arrojar algo de luz en un tópico bastante delicado: determinación de la estructura del modelo y validación del modelo.

Algunas técnicas referidas a esta matriz ya han sido descritas, p. ej: el estudio de su determinante (ver p. ej: (52)), pero parecería no haber sido bastante explotada la comparación entre la matriz de covarianzas prevista, la estimada, y la evolución de la estimación.

Volviendo al análisis de las figuras anteriores, reiteramos que en general son concordantes (y mucho) con las tendencias previstas. En el resto del párrafo se describen pruebas para verificar dichas tendencias.

Dependencia del número de observaciones

Para estudiar la dependencia del estimador de mínimos cuadrados respecto del número de observaciones y el nivel de ruido, consideraremos tres factores de mérito:

1. $\text{Var}(\hat{\theta}_1)$
2. $\delta^2 = E \frac{||\hat{\theta} - \theta||^2}{||\theta||^2}$
3. $V_2(\hat{\theta}) = E \left\{ \left[\frac{\hat{u}(q^{-1})}{A(q^{-1})} - \frac{B(q^{-1})}{A(q^{-1})} e(t) \right]^2 \right\}$

1. El primer factor: $\text{var}(\hat{\theta}_1)$ ya ha sido discutido como factor de mérito de un estimador. Hemos visto además que en nuestras simulaciones se espera que sea proporcional a $1/N$, y que crezca junto con σ^2 (según una función racional de σ^2).

Para el sistema de primer orden, y con $E(u^2) = 1$, las expresiones (5.149) y (5.151) muestran que:

$$\begin{cases} \text{var}(\hat{a}_1) = \frac{\sigma^2}{1 + \sigma^2} \frac{1}{2.7778 N} \\ \text{var}(\hat{b}_0) = \frac{\sigma^2}{N} \end{cases} \quad (5.156)$$

para el sistema de segundo orden podemos derivar la expresión de la matriz J_N^{-1} de covarianzas (asintótica) del estimador:

Para un sistema estable (como el propuesto), con $u(t)$ persistentemente excitante, y $e(t)$ ruido blanco de media cero y varianza 1:

$$\hat{O}_1^{-1} \frac{1}{N} \sum_{t=1}^N \mathcal{N}(0, J_N^{-1}) \text{ para } N \rightarrow \infty$$

donde $J^{-1} = \frac{\sigma^2 R^{-1}}{N}$, y R es:

$$R = \begin{bmatrix} E y^2(t) & E y(t) y(t-1) & 0 & -E u(t) y(t+1) \\ E y(t) y(t-1) & E y^2(t) & 0 & 0 \\ 0 & 0 & E u^2(t) & 0 \\ -E u(t) y(t+1) & 0 & 0 & E u^2(t) \end{bmatrix} \quad (5.157)$$

es decir

$$R = \begin{bmatrix} R_{yy}(0) & R_{yy}(1) & 0 & -R_{uy}(1) \\ R_{yy}(1) & R_{yy}(0) & 0 & 0 \\ 0 & 0 & R_{uu}(0) & 0 \\ -R_{uy}(1) & 0 & 0 & R_{uu}(0) \end{bmatrix} \quad (5.158)$$

en nuestro caso $E u^2 = 1$.

Con el algoritmo del capítulo 4 se ha calculado la varianza de la parte determinística y_1 del sistema, y de la parte de perturbaciones y_2 :

$$E y_1^2 = 10.86021 \cdot E u^2 = 10.86021$$

$$E y_2^2 = 8.85417 \cdot \sigma^2$$

$$\therefore E y^2 = R_{yy}(0) = E y_1^2 + E y_2^2 = 18.86021(1 + \frac{1}{S/N}) = 18.86021 + 8.85417 \sigma^2$$

para calcular $R_{yy}(1)$ y $R_{uy}(1)$ consideremos la ecuación del sistema:

$$y(t) = -a_1 y(t-1) - a_2 y(t-2) + b_0 u(t-1) + b_1 u(t-2) + \omega(t)$$

multiplicando ambos lados por $u(t-1)$ y tomando esperanzas:

$$\begin{aligned} R_{uy}(1) &= b_0 R_{uu}(0) \\ \therefore R_{uy}(1) &= b_0 \end{aligned} \quad (5.159)$$

multiplicando por $y(t-1)$ y tomando esperanzas:

$$R_{yy}(1) = -a_1 R_{yy}(0) - a_2 R_{yy}(1) + b_1 R_{uy}(1)$$

$$\therefore R_{yy}(1) = \frac{b_1 b_1 - a_1 R_{yy}(0)}{1 + a_1} \quad (5.160)$$

sustituyendo en (5.158) los valores de los parámetros del sistema, se calcula J_N^{-1} para $S/N = 1$ ($\sigma = 1.46026$)

$$J_N^{-1} = \frac{1}{N} \begin{bmatrix} 0.31171 & -0.2774 & 0 & 0.31171 \\ -0.2774 & 0.30346 & 0 & -0.2774 \\ 0 & 0 & 2.13236 & 0 \\ 0.31171 & -0.2774 & 0 & 2.44407 \end{bmatrix} \quad (5.161)$$

La matriz hallada es la matriz teórica de covarianzas del estimador de mínimos cuadrados para el sistema de segundo orden propuesto, con $S/N = 1$. Los elementos diagonales son las varianzas de la estimación de cada uno de los parámetros.

- El segundo factor: $\sqrt{\sigma^2}$ es un factor global para el conjunto de los parámetros. Estudia la desviación standard relativa del parámetro vectorial \hat{O}_N .

ha sido ampliamente utilizado en la literatura, ver por ej. ref. (21). Es uno de los factores de mérito propuesto por Iserman ver por ejemplo: (31)

Se espera que este factor sea proporcional a $1/\sqrt{N}$.

Para el sistema de primer orden, de acuerdo a las ecuaciones (5.156) y los valores de los parámetros

$$E \left[\sqrt{\delta^2} \right] = \sqrt{\frac{1.30 \sigma^2 + (\sigma^2)^2}{1.05 \cdot 3 (1.70 \sigma^2)}} \quad (5.162)$$

Para el sistema de segundo orden, de acuerdo a la matriz \hat{H}^{-1} calculada en (5.161), para una relación señal a ruido $S/N = 1$:

$$E \left[\sqrt{\delta^2} \right] = \sqrt{\frac{0.111b}{3.99 N}} \quad (5.163)$$

3. El tercer factor de mérito: $V_2(\hat{\theta})$ es una medida de qué tan bien aproxima el modelo estimado la parte determinística del sistema.

Usamos el factor $V_2(\hat{\theta})$ con u : ruido blanco de varianza unitaria y media cero. En tal caso:

$$\begin{aligned} V_1(\hat{\theta}) &= E \left[\sum_{i=0}^{\infty} \left[\hat{h}(i) - h(i) \right] u(t) \right]^2 \\ V_2(\hat{\theta}) &= E \left[\sum_{i=0}^{\infty} \left[\hat{h}(i) - h(i) \right]^2 u(t)^2 \right] \\ V_3(\hat{\theta}) &= \left[\sum_{i=0}^{\infty} \left[\hat{h}(i) - h(i) \right]^2 \right] E \left[u(t)^2 \right] \\ V_4(\hat{\theta}) &= \sum_{i=0}^{\infty} \left[\hat{h}(i) - h(i) \right]^2 \end{aligned} \quad (5.164)$$

Donde $h(i)$ y $\hat{h}(i)$ son respectivamente la respuesta a pulso del sistema simulado y la respuesta a pulso del sistema estimado.

El método utilizado para verificar que se cumplen las leyes previstas por la teoría es el siguiente:

Se utilizarán las simulaciones descritas al principio del capítulo, con entrada $u = \text{M35127}$, y perturbaciones gaussianas RUID01, RUID02, RUID03, RUID04, RUID05, para identificar los sistemas de primer y segundo orden. Con los parámetros estimados se estimarán las cantidades $\text{var}(\hat{\theta}_i)$, $\sqrt{\delta^2}$, $V_2(\hat{\theta})$ promediando los resultados obtenidos con cinco realizaciones distintas (manteniendo un mismo valor de N y σ^2). El procedimiento se repite para distintos valores de N y σ^2 .

La forma de estimar $\text{var}(\hat{\theta}_i)$, $\sqrt{\delta^2}$, $V_2(\hat{\theta})$ es:

$$\begin{cases} \text{var}(\hat{\theta}_i) = \frac{1}{5} \sum_{j=1}^5 (\hat{\theta}_i^j - \theta_i)^2 \\ \sqrt{\delta^2} = \sqrt{\frac{1}{5} \left| \hat{\theta} \right| \left[\sum_{j=1}^5 \left[\hat{\theta}_i^j - \theta_i \right]^2 \right]} \\ V_2(\hat{\theta}) = \frac{1}{5} \sum_{j=1}^5 V_2(\hat{\theta}^j) \end{cases} \quad (5.165)$$

donde p es el número de parámetros, el superíndice j : 1,2,3,4,5 indica de qué realización se obtuvo el estimado (realización perturbada con la señal RUIDOj). A continuación se describen los resultados obtenidos.

Sistema de primer orden

Para relaciones señal a ruido $S/N = 10, 4, 1, 0.1$ se efectuaron las identificaciones del sistema de primer orden. Para cada valor de S/N se realizó la identificación sobre 5 señales obtenidas como resultado de simular el sistema con las perturbaciones RUIDO1, RUIDO2, RUIDO3, RUIDO4 y RUIDO5.

Para valores fijos de S/N se obtuvieron las cantidades $\hat{\sigma}_{d1}^2, \hat{\sigma}_{d2}^2, \sqrt{\hat{\sigma}^2}$ como promedio de las cinco realizaciones de acuerdo a las ecuaciones (5.165); estas cantidades se computaron para valores de $N = 50, 100, 150, 300$ y 500 .

Los resultados obtenidos se grafican en las figuras que siguen. En el eje de abscisas el n^2 de observaciones N , y en ordenadas la cantidad computada. Los cinco puntos así obtenidos (correspondientes a $N = 50, 100, 150, 300, 500$) se unieron con línea gruesa. Con línea delgada se dibujó la curva que prevé la teoría (ver tabla 5.1).

La escala es logarítmica, por lo que las curvas teóricas (que en escala lineal son hipérbolas) son rectas.

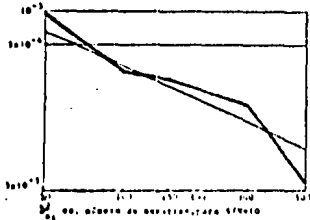


Fig 5.20

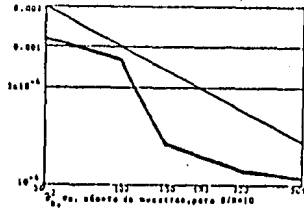


Fig 5.21

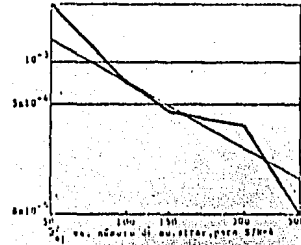


Fig 5.22

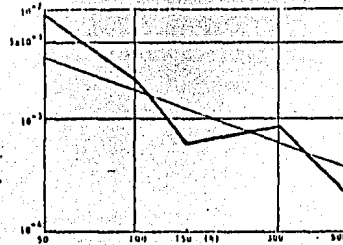


Fig 5.23

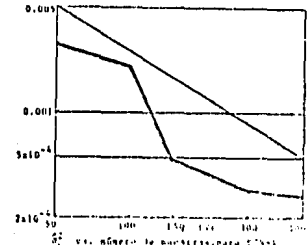


Fig 5.24

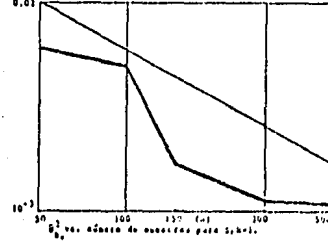


Fig 5.25

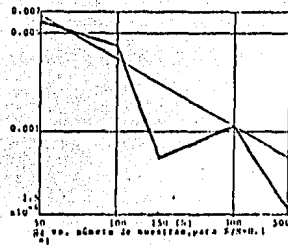


Fig 5.26

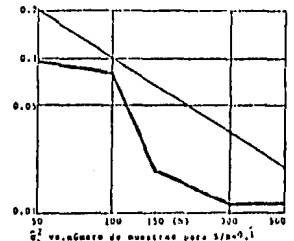


Fig 5.27

En las figuras (5.20) a (5.27) se graficaron las cantidades $\hat{\sigma}_{d1}^2$ y $\hat{\sigma}_{d0}^2$ contra el número de muestras, para $S/N = 10, 4, 1, 0.1$.

En las gráficas correspondientes a $\hat{\sigma}_{d1}^2$ se observa una enorme coincidencia entre los resultados teóricos y los obtenidos en las simulaciones.

En las gráficas correspondientes a $\hat{\sigma}_{d0}^2$ se observan resultados prácticos algo mejores que los previstos por la teoría: aunque coinciden en orden de magnitud y tendencia, todos los promedios computados se encuentran algo por debajo del valor teórico. De hecho podría ajustarse mejor a los valores calculados, una varianza igual a la mitad de la teórica.

En todos los casos, la tendencia proporcional a $1/N$ se corrobora. Las siguientes figuras son las gráficas de $\sqrt{\hat{\sigma}^2}$ correspondientes a las mismas simulaciones del sistema del primer orden.

Así como los valores teóricos de las figuras anteriores corresponden a las ecuaciones (5.156), los de las figuras que siguen corresponden a la ecuación (5.162) derivada de las mismas ecuaciones.

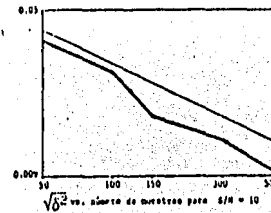


Fig 5.28

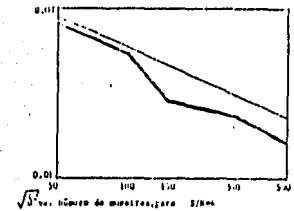


Fig 5.29

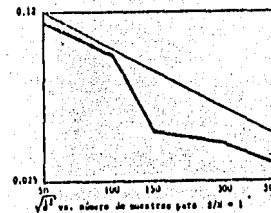


Fig 5.30

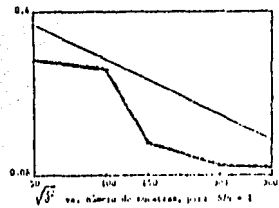


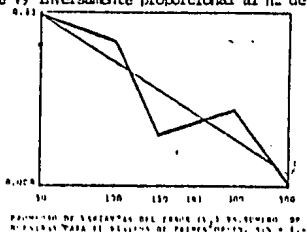
Fig 5.31

También para el factor $\sqrt{\hat{\sigma}^2}$, se encuentra una enorme coincidencia entre el valor teórico y la cantidad computada. Los resultados prácticos son algo mejores que los teóricos en correspondencia con el hecho que $\hat{\sigma}_{d0}^2$ es algo más pequeño que el valor teórico. Su comportamiento asimila no que $\sqrt{\hat{\sigma}^2}$ es proporcional al factor $1/\sqrt{N}$.

Para una relación señal a ruido $S/N = 1$ se estimó el factor $V_2(\hat{\sigma}_N)$ con $N = 50, 100, 150, 200$ y 250 , de acuerdo a (5.165). En la figura adjunta se grafican en línea gruesa los valores computados.

Se encontró una tendencia decreciente con el no. de observaciones, que confirma desde otro punto de vista la tendencia del sistema estimado a aproximar mejor al sistema que genera los datos, conforme $N \rightarrow \infty$.

Se ajustó a los resultados (en escala logarítmica) una recta por mínimos cuadrados. Esta tiene una pendiente de -1.180 , lo que hace pensar en una tendencia de V , inversamente proporcional al no. de observaciones.



Sistema de segundo orden

El mismo procedimiento que para el sistema de primer orden, se llevó a cabo para el de segundo orden, con un solo valor de la relación señal a ruido: $S/N = 1$.

Los resultados se presentan en las gráficas que siguen. En la figura (5.32) se presentan conjuntamente las cantidades $\hat{\sigma}_{a_1}^2$, $\hat{\sigma}_{a_2}^2$, $\hat{\sigma}_{b_0}^2$, $\hat{\sigma}_{b_1}^2$ para comprobar el hecho previsto de la observación de la matriz de covarianzas asintótica (5.161), de que las varianzas de \hat{a}_1 y \hat{a}_2 son similares entre sí, así como las de \hat{b}_0 y \hat{b}_1 entre sí. Además las varianzas de \hat{b}_0 y \hat{b}_1 son entre 7 y 8 veces mayores que las de \hat{a}_1 y \hat{a}_2 .

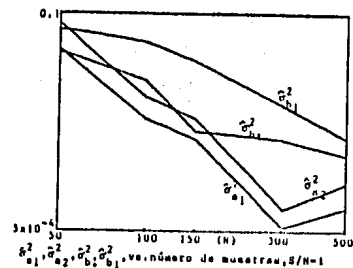


Fig 5.32

En las siguientes figuras se grafican $\hat{\sigma}_{a_1}^2$, $\hat{\sigma}_{a_2}^2$, $\hat{\sigma}_{b_0}^2$, $\hat{\sigma}_{b_1}^2$, contra el número de muestras,

En líneas delgadas se dibujan las tendencias previstas.

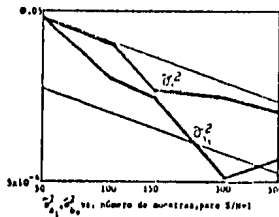


Fig 5.33

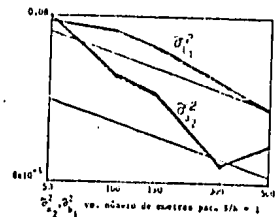


Fig 5.34

Otra vez los valores prolijos y los reales se aproximan mucho. No hay que olvidarse que en todas estas comprobaciones sólo hemos tomado en cuenta promedios sobre cinco realizaciones.

Las curvas teóricas son válidas para $N \rightarrow \infty$ pero aún así se muestra que son muy buenas aproximaciones para valores finitos de N relativa-

mento pequeño ($N < 500$).

En las varianzas de las estimaciones de a_1 y a_2 , los valores encontrados para 50 muestras no son muy buenos, pero esta situación mejora notablemente para $N \geq 100$.

La figura 5.35 muestra la gráfica de los valores computados de $\sqrt{\hat{\sigma}_1^2}$.

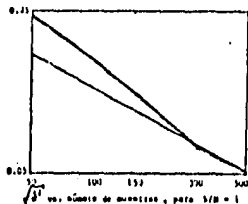


Fig 5.35

La proximidad entre los valores teóricos y computados para $N = 50, 100$ y 150 es buena, pero para $N = 300, 500$ es casi exacta.

Dependencia de la varianza de las perturbaciones

Figurémonos ahora el número de muestras N y variaremos el nivel de la señal de ruido, es decir, la relación señal a ruido S/N ; trabajaremos solamente con señales del sistema de primer orden.

En cuanto al comportamiento de $\sigma_{a_1}^2$ y $\sigma_{a_2}^2$, sus valores para $N \rightarrow \infty$ están dados por las ecuaciones (5.156). Estas indican que $\sigma_{a_1}^2$ es proporcional a $\frac{\sigma^2}{1+S^2}$ mientras que $\sigma_{b_0}^2$ lo es a σ^2 .

En las figuras (5.36) y (5.37) se grafican los valores computados de $\hat{\sigma}_{a_1}^2$ y $\hat{\sigma}_{b_0}^2$ respectivamente.

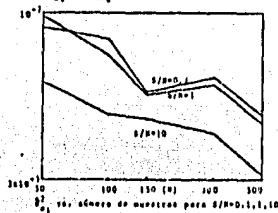


Fig 5.36

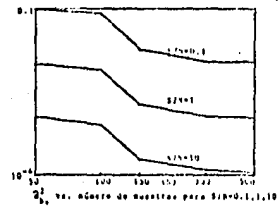


Fig 5.37

Recordando que para nuestro sistema de primer orden $S/N = 1/\sigma^2$ las figuras son coherentes con una ley del tipo (5.156). Es decir, la varianza de \hat{b}_0 parece ser proporcional a \ln del ruido: σ^2 , en cambio la de \hat{a}_1 crece cada vez menos para σ^2 creciente.

Para el estimador \hat{b}_0 esto se verifica para todos los largos de muestras graficadas (50, 100, 150, 300 y 500), en cambio la tendencia general de $\hat{\sigma}_{a_1}^2$ no se verifica en el menor de los largos de muestras considerables (50).

En las siguientes figuras se comparan los valores computados de $\hat{\sigma}_{a_1}^2$ y $\hat{\sigma}_{b_0}^2$, con los valores calculados (teóricos) de $\sigma_{a_1}^2$ y $\sigma_{b_0}^2$; esto se hace para algunos valores particulares de N . En trazo delgado los valores teóricos.

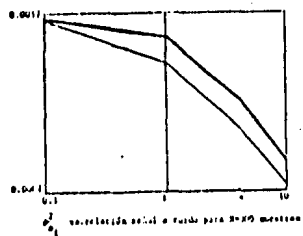


Fig 5,38

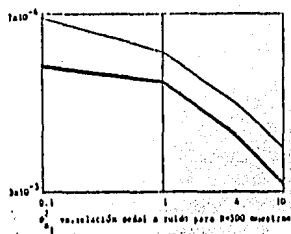


Fig 5,39

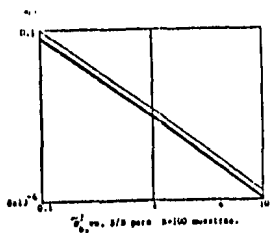


Fig 5,40

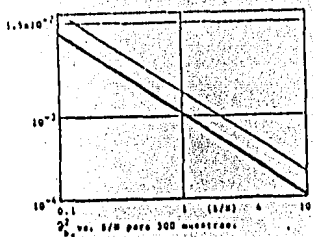


Fig 5,41

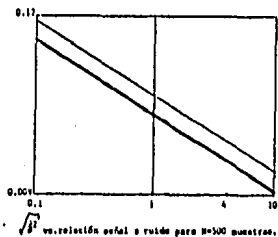


Fig 5,42

Como cuando estudiamos la dependencia respecto del número de muestras, el valor de $\hat{\sigma}_{b_1}^2$ ajusta bastante bien lo previsto; en la figura (5.39) correspondiente a $N = 300$ el alejamiento del estimador respecto del valor real, es algo menor que el previsto, y en la figura (5.39) correspondiente a $N = 500$, es algo mayor.

El estimador \hat{b}_0 tiene un alejamiento estimado según $\hat{\sigma}_{b_0}^2$ que sigue una ley casi perfectamente proporcional a $(S/N)^{-1}$ y muy próxima a la prevista.

El estimador \hat{b}_0 es algo más próximo a b_0 que lo que se espera de acuerdo a (5.156).

Para concluir, en la figura (5.42) se grafica $\sqrt{\hat{\sigma}_{b_1}^2}$ para $N = 500$; esta cantidad sigue casi perfectamente el tipo de ley prevista, con un valor ligeramente por debajo del esperado, ya que tanto $\hat{\sigma}_{b_1}^2$ y $\hat{\sigma}_{b_0}^2$ son un poco inferiores a su ley asintótica en $N = 500$ (ver figs (5.39) y (5.41)).

5.12.4 Estimación de la varianza de los estimadores

Hasta ahora hemos comparado el estimador con el valor real del parámetro, pero en general de poco sirve la estimación si no se acompaña de una medida de su precisión. De la diagonal de la matriz de covarianzas:

$$\text{diag} \left[\frac{\sigma^2}{N} \Gamma(x_i, x_i^T) \right]$$

se puede conocer la varianza de los estimadores. Pero cuando se está identificando no se conoce esta matriz puesto que para conocerla se necesitaría haber resuelto el problema de identificación: conocer el modelo del sistema.

Lo que se puede hacer es estimar también la varianza de los estimadores. Hemos visto que $\hat{\sigma}^2 (X^T X)^{-1}$ (donde $\hat{\sigma}^2$ es el estimador de la varianza del ruido) es un estimador de la matriz de covarianzas de $\hat{\theta}_N$ (ver capítulo II y teoremas 5.1, 5.2 y 5.3).

En este párrafo compararemos la estimación de la varianza de los estimadores (diagonal de $\hat{\sigma}^2 (X^T X)^{-1}$) con la varianza asintótica de los estimadores (diagonal de $\frac{\sigma^2}{N} \Gamma(x_i, x_i^T)$).

Sistema de primer orden

Las ecuaciones (5.156) determinan la varianza asintótica de los estimadores. Se estimó dicha varianza de acuerdo a (5.81) y (5.83) con:

$$\text{cov} \hat{\theta}_N = \hat{\sigma}^2 (X^T X)^{-1}$$

con el estimador $\hat{\sigma}^2$ de σ^2 : $S_N^1 = \frac{1}{N-p} S_N(\hat{\theta}_N)$

donde $S_N(\hat{\theta}_N)$ es la función de costo (suma de cuadrados de los residuos), p es el no. de parámetros, y N es el no. de observaciones.

Para una relación señal a ruido $S/N = 10$, se consideraron las cinco realizaciones que resultaron de excitar el sistema con entrada $u(t) = \text{PRBS127}$, y perturbaciones RUIDO1, RUIDO2, RUIDO3, RUIDO4, RUIDO5 respectivamente. Para $N = 50, 100, 150, 300$ y 500 observaciones se estimaron las varianzas de \hat{a}_1 y \hat{b}_0 .

En cada valor de N considerado se tomó la estadística: rango ($\hat{\theta}_i$) = (máx. desv. std. ($\hat{\theta}_i$), mín. desv. std. ($\hat{\theta}_i$)) es decir el intervalo entre la máxima y la mínima desviación std. estimada dentro de las cinco estimaciones.

En la figura (5.43) se dibujan en líneas gruesas las desv. std. de \hat{a}_1 y \hat{b}_0 , predichas por las ecuaciones (5.156) y en líneas delgadas se unen los extremos de los intervalos donde caen las estimaciones de dichas desviaciones std. La zona limitada por estas líneas está sombreada.

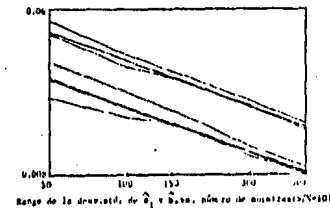


Fig 5.43

Dijo notarse la enorme coincidencia entre los valores observados y los predichos por la teoría, teniéndose en cuenta que no se está graficando ninguna clase de promedio, sino todo el rango de valores donde caen las estimaciones de las desviaciones std, que puede considerarse como una estadística del "poor caso".

La coincidencia es doblemente buena: primero porque todo el rango de valores es muy próximo al teórico, y segundo porque se va estrechando alrededor de la ley asintótica para $N \rightarrow \infty$.

Se pueden hacer más comparaciones para el sistema de primer orden con las estimaciones ya tabuladas en páginas anteriores. Para ello, la Tabla 5.V presenta desv. std. teóricas para distintos valores de N y S/N .

Tabla 5.V Algunos valores teóricos de la desv. std. de \hat{a}_1 y \hat{b}_2 para el sistema de primer orden.

N	Parámetro	S/N = 4	S/N = 1
100	\hat{a}_1	0.02683	0.04743
	\hat{b}_2	0.05	0.1
122	\hat{a}_1	0.02429	0.03861
	\hat{b}_2	0.04527	0.09054
200	\hat{a}_1	0.01897	0.03
	\hat{b}_2	0.03536	0.07071

Compárese las filas de $N = 122$ con los resultados de la tabla 5.II.
Compárese de la columna $S/N = 4$, con $N = 100$ y 200 , los valores tabu-

lados y las estimaciones al pie de la figura (5.12).

Compárese los valores tabulados en la columna $S/N = 1$, con $N = 100$ y 200 , con los resultados de las estimaciones al pie de la figura (5.13). En todos los casos las estimaciones son enormemente próximas al valor esperado dado por la tabla 5.V.

Sistema de segundo orden

Se empleó el mismo método que para el sistema de primer orden. Las figuras (5.44) a (5.47) muestran el rango de las estimaciones de las desviaciones std de \hat{a}_1 , \hat{a}_2 , \hat{b}_2 y \hat{b}_1 , respectivamente para una relación señal a ruido $S/N = 1$, y valores de $N = 50, 100, 150, 300$ y 500 observaciones. Los valores teóricos son calculados tomando raíz cuadrada de la diagonal de (5.161). Otra vez la coincidencia entre los valores previstos y estimados es enorme.

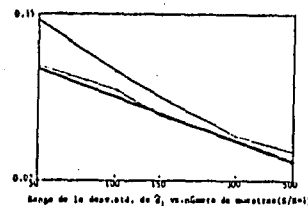


Fig 5.44

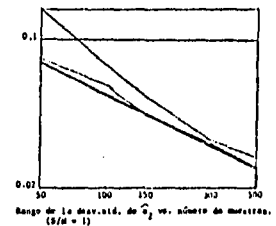


Fig 5.45

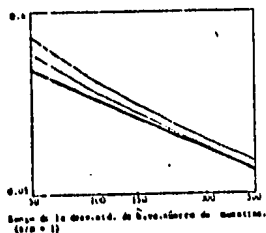


Fig 5.46

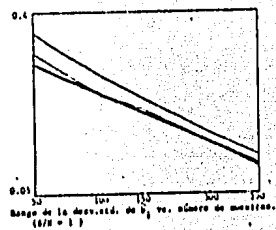


Fig 5.47

También en este caso puede observarse la tendencia de las estimaciones de las desv. std., de estrechar su rango de variación alrededor de su ley asintótica para valores de N crecientes.

Puede observarse una tendencia a la sobreestimación de las desv. std. para los valores más pequeños de N (50, 100 y 150), especialmente en las desviaciones de los estimadores \hat{a}_1 y \hat{a}_2 .

Aún esto no representa una dificultad, puesto que ya hemos visto que la desviación contrada de \hat{a}_1 y \hat{a}_2 es mayor que la prevista en valores pequeños de N (ver figuras 5.33 y 5.34) de modo que este exceso en la estimación de la desv. std. de \hat{a}_1 y \hat{a}_2 están más cerca de la situación real de las desv. std. de \hat{a}_1 y \hat{a}_2 , que la ley asintótica.

También se puede verificar en ejemplos ya vistos, el comportamiento de la desv. std. estimada.

La tabla 5.VI presenta los valores teóricos de la desv. std. de \hat{a}_1 , \hat{a}_2 , \hat{b}_0 y \hat{b}_1 del sistema de segundo orden, para $S/N = 1$ y distintos valores de N .

Tabla 5.VI. Algunas desviaciones std. de la estimación de los parámetros del sistema de 2º orden.

Parámetro	$N = 100$	$N = 124$	$N = 200$	$N = 300$
a_1	0.05583	0.05014	0.03948	0.03223
a_2	0.05509	0.04947	0.03895	0.0318
b_0	0.14603	0.13113	0.10326	0.08431
b_1	0.15634	0.14039	0.11055	0.09026

Comúnense los valores de la columna correspondiente a $N = 124$ con los resultados tabulados en la tabla 5.III.

Comúnense los valores de las columnas correspondientes a $N = 100$, 200 y 300 con los resultados correspondientes de la figura (5.16).

Otra vez se comprueba una gran semejanza entre los valores estimados de la desviación Standard de los estimadores y la ley asintótica de dichas desviaciones.

5.12.5 Estimación de la varianza del ruido

Aunque la estimación de la varianza σ^2 del ruido del modelo (5.2) ha estado implícita en la estimación de la varianza del estimador de mínimos cuadrados, vale la pena estudiarla por separado y comparar algunos estimadores.

El estimador que hemos usado normalmente es (5.27):

$$S_N^2 = \frac{1}{N-p} (Y_N - X_N \hat{\theta}_N)^T (Y_N - X_N \hat{\theta}_N) \quad (5.166)$$

que, de acuerdo al teorema 5.4 es débilmente consistente.

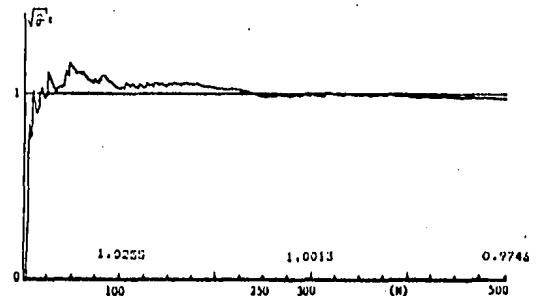
Consideremos la desviación std. del ruido como otro parámetro más a estimar; la matriz de información de Fisher correspondiente a perturbaciones gaussianas independientes se amplía en una fila y una columna, ambas con ceros excepto el elemento diagonal que es:

$$\begin{bmatrix} \sigma^2 \\ \frac{\sigma^2}{2N} \end{bmatrix}^{-1} \quad (5.167)$$

Entonces la cantidad $\sigma^2/2N$ es la cota de Cramer-Rao de la varianza de la estimación de la desviación Standard del ruido. (ver p. ej (2)). Esta cota se alcanza asintóticamente en el caso que estamos considerando: perturbaciones independientes y gaussianas.

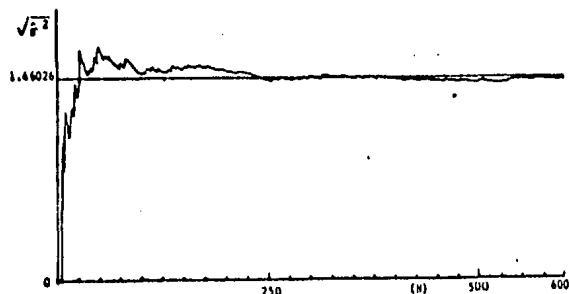
En las siguientes figuras se presenta la evolución de la estimación de la desv. std. para tres casos, uno del sistema de primer orden, uno del sistema de segundo orden, y uno del sistema de tercer orden.

En las tres simulaciones se utilizó una relación $S/N = 1$. El estimador de σ es $\sqrt{\hat{\sigma}^2}$ donde el estimador $\hat{\sigma}^2$ de σ^2 es S_N^2 dado por (5.166).



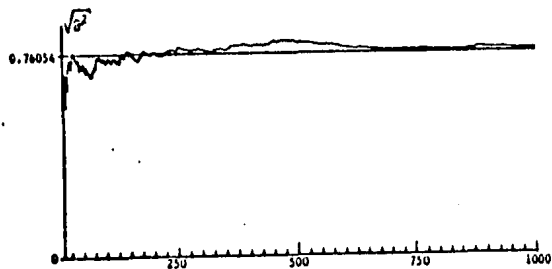
ESTIMACION DE LA DESV. STD. DEL RUIDO. SISTEMA DE ORDEN 1, $\sigma^2=1$, $S/N=1$.
PERTURBACION: RUIDO1.

Fig 5,18



ESTIMACION DE LA DESV. STD. DEL RUIDO. SISTEMA DE ORDEN 2, $\sigma^2=2.13235$, S/N=1.-
PERTURBACION : RUIDO1

Fig 5.49



ESTIMACION DE LA DESV. STD. DEL RUIDO. SISTEMA DE ORDEN 3, $\sigma^2=0.37841$, S/N=1.-
PERTURBACION : RUIDO3

Fig 5.50

En los tres casos la estimación es enormemente próxima a σ , y es compatible con la provisión de consistencia del teorema 5.4. La perturbación utilizada en los tres casos es gaussiana y la entrada al sistema es PRBS127.

Para verificar la tendencia inversamente proporcional a N de la varianza del estimador de σ , se realizaron simulaciones sobre el sistema de primer y segundo orden.

En ambos casos se excitó el sistema con entrada PRBS127 y se perturbaron los sistemas con RUIDO1, RUIDO2, RUIDO3, RUIDO4, RUIDO5, obteniéndose así cinco realizaciones del sistema de primero y segundo orden. En el de primer orden se utilizó S/N = 10, y en el segundo orden S/N = 1.

Se computó la desviación cuadrática media de la estimación de σ para las cinco realizaciones en ambos casos, esto es:

$$DCH = \left[\frac{1}{5} \sum_{j=1}^5 (\sqrt{\hat{\sigma}_j^2} - \sigma)^2 \right]^{1/2} \quad (5.168)$$

donde el índice j señala el número de realización. Esto fue hecho para distintos valores de N (50, 100, 150, 300 y 500).

Los resultados se presentan en las figuras (5.51) y (5.52), en línea gruesa, contra la ley asintótica (en línea delgada) dada por la inversa de la expresión (5.167).

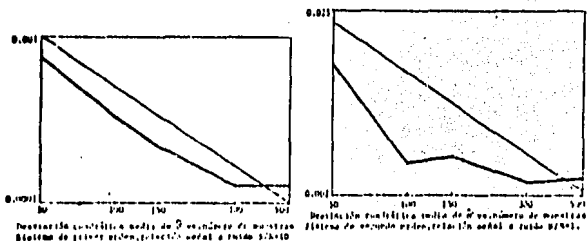


Fig 5.51

En ambos casos los resultados manifiestan una tendencia similar a la prevista y para los valores más altos de N computados (300 y 500) los datos muestran mayor proximidad a la ley teórica.

Sin lugar a dudas la coincidencia es muchísimo mejor en las simulaciones sobre el sistema de primer orden. En ambos casos para $N \leq 300$ se ha encontrado una estimación de σ pronomialmente mejor que lo que se esperaba.

Ceros estimadores de la varianza han sido discutidos en el párrafo

5.6.c) éstos son:

$$\begin{cases} S2 = \left(\frac{1}{N} \sum_{j=1}^N \epsilon_j^2(j) / \sum_{j=1}^N \frac{1}{j+p} \right)^{1/2} \\ S3 = \left(\frac{1}{N} \sum_{j=1}^N \frac{\epsilon_j^2(j)}{(j/j+p)} \right)^{1/2} \end{cases} \quad (5.169)$$

En donde $\epsilon_j(j)$ es el j -ésimo residuo correspondiente a la estimación en $t=j$ (es decir el último residuo correspondiente a dicha estimación).

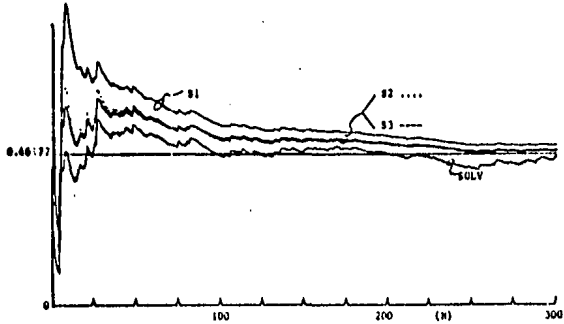
Otra forma de generar un estimador es con un factor de olvido α :

$$SOLV = \frac{(2-\alpha)}{1-\alpha} \sum_{j=1}^N \alpha^{N-j} \epsilon_j^2(j) \quad (5.170)$$

En la figura 5.53 se muestra la evolución de estos tres estimadores además de $S1 = \frac{1}{N-p} \sum_{i=1}^N \epsilon_N^2(i)$, (que es el que hemos utilizado hasta ahora), para una realización del sistema de Segundo orden con inicialización arbitraria.

En este ejemplo $S1$ es el que se mantiene más alejado (aunque los 4 estimadores convergen adecuadamente), pues depende muy fuertemente de la inicialización del algoritmo.

Los estimadores $S2$ y $S3$ en este ejemplo se mantienen muy próximos entre sí, y $SOLV$ rápidamente se acerca al valor real de σ , a costa de mantener un nivel de oscilación permanente alrededor de él debido al efecto del factor de olvido $\alpha = 0.98$.



COMPORTAMIENTO DE TRES ESTIMADORES DE LA DESV. STD. DEL RUIDO EN UNA REALIZACIÓN DEL SISTEMA DE SEGUNDO ORDEN. 3/4=10, PERTURBACIONES BLANCAS.

Fig 5.53

5.12.6 Inicialización arbitraria vs. exacta

Ya se ha discutido (ver 5.4 y 5.5) el problema de la inicialización del algoritmo de mínimos cuadrados y se ha mostrado que para $N \rightarrow \infty$ es equivalente el utilizar los primeros datos como condiciones iniciales para las siguientes recursiones, a considerar una matriz inicial $(X_0^T X_0)^{-1} = P_0^{-1} = KI$ con $K \gg 1$, y $\hat{X}_0 = 0$.

Por supuesto que la velocidad con que se aproximan entre sí los estimadores obtenidos con ambas inicializaciones no es siempre la misma, de hecho depende de muchos factores. En particular la estructura de la matriz $(X^T X)$ que depende de la estructura del sistema es muy importante.

Por ejemplo en un mismo sistema, para distintos niveles de ruido, a mayor varianza del ruido, más pronto se aproximan los estimadores. Esto se debe a que la varianza del ruido incide mucho en una parte de la diagonal de $(X^T X)$.

De la misma forma, el mismo sistema con distintas amplitudes de entrada es estimado en forma más parecida a mayor amplitud de la señal de entrada.

En la práctica los resultados obtenidos con ambas inicializaciones han resultado bastante próximos, y han confirmado las provisiones. Para avalar esta afirmación presentamos algunos ejemplos.

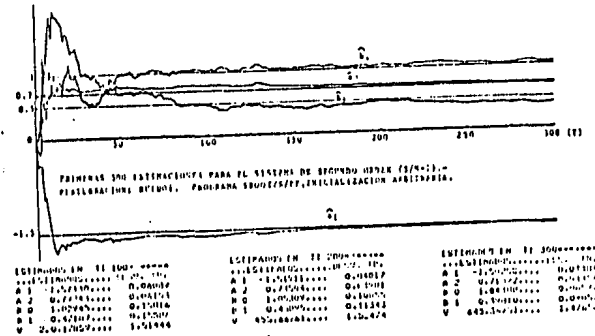


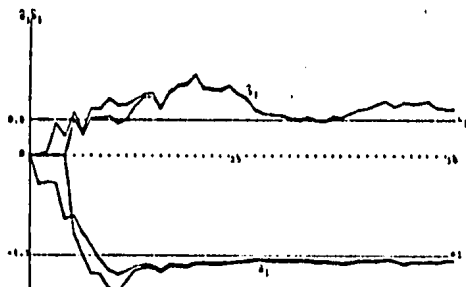
Fig 5.54

En la figura (5.54) se muestra la identificación del sistema de segundo orden para una relación $S/N = 1$, perturbado con RUIDO1. La identificación fue inicializada arbitrariamente.

La misma identificación, realizada con inicialización exacta se muestra en la figura (5.16).

Excepto en los primeros puntos, donde todavía no existe el estimado con inicialización exacta, resulta difícil distinguir diferencias a simple vista entre ambos estimadores.

Para visualizar esas diferencias se han ampliado los primeros 50 puntos de estimación de los parámetros a_1 y b_1 . A partir de $N = 16$ esta diferencia es muy pequeña y a partir de $N = 40$ la precisión de la estimación no permite distinguir diferencias.



PRIMEROS 50 ESTIMACIONES DE LOS PARÁMETROS a_1 Y b_1 DEL SISTEMA DE ORDEN 2 POR INICIALIZACIÓN EXACTA Y ARBITRARIA. RELACION SEÑAL A RUIDO $S/N=1$. (ESTRUCTURA: RUIDO1)

Para comparar numéricamente ambas estimaciones se tabulan los resultados para $N = 10, 50, 100, 200, 300, 400$ y 500 en la tabla 5.VII.

TABLA 5.VII

COMPARACIÓN DE RESULTADOS DE ESTIMACIÓN POR MÍNIMOS CUADRADOS CON INICIALIZACIÓN ARBITRARIA Y EXACTA PARA EL SISTEMA DE ORDEN 2

Número de estimaciones: 10000 cuadros, Programa: SPOUT1077.
Relación señal a ruido: $S/N=1$, Número de observaciones: 6000.
INICIALIZACIÓN ARBITRARIA. INICIALIZACIÓN EXACTA.

ESTIMACIONES EN TI 10000000		ESTIMACIONES EN TI 10000000	
A1	-1.2791000000000000	A1	-1.2791000000000000
A2	0.7181000000000000	A2	0.7181000000000000
B0	1.0000000000000000	B0	1.0000000000000000
B1	0.0000000000000000	B1	0.0000000000000000
V	10.1000000000000000	V	0.0000000000000000
ESTIMACIONES EN TI 50000000		ESTIMACIONES EN TI 50000000	
A1	-1.2500000000000000	A1	-1.2500000000000000
A2	0.7000000000000000	A2	0.7000000000000000
B0	0.9999999999999999	B0	0.9999999999999999
B1	0.0000000000000000	B1	0.0000000000000000
V	10.0000000000000000	V	10.0000000000000000
ESTIMACIONES EN TI 100000000		ESTIMACIONES EN TI 100000000	
A1	-1.2500000000000000	A1	-1.2500000000000000
A2	0.7000000000000000	A2	0.7000000000000000
B0	1.0000000000000000	B0	1.0000000000000000
B1	0.0000000000000000	B1	0.0000000000000000
V	10.0000000000000000	V	10.0000000000000000
ESTIMACIONES EN TI 200000000		ESTIMACIONES EN TI 200000000	
A1	-1.2500000000000000	A1	-1.2500000000000000
A2	0.7000000000000000	A2	0.7000000000000000
B0	1.0000000000000000	B0	1.0000000000000000
B1	0.0000000000000000	B1	0.0000000000000000
V	10.0000000000000000	V	10.0000000000000000
ESTIMACIONES EN TI 300000000		ESTIMACIONES EN TI 300000000	
A1	-1.2500000000000000	A1	-1.2500000000000000
A2	0.7000000000000000	A2	0.7000000000000000
B0	1.0000000000000000	B0	1.0000000000000000
B1	0.0000000000000000	B1	0.0000000000000000
V	10.0000000000000000	V	10.0000000000000000
ESTIMACIONES EN TI 400000000		ESTIMACIONES EN TI 400000000	
A1	-1.2500000000000000	A1	-1.2500000000000000
A2	0.7000000000000000	A2	0.7000000000000000
B0	1.0000000000000000	B0	1.0000000000000000
B1	0.0000000000000000	B1	0.0000000000000000
V	10.0000000000000000	V	10.0000000000000000
ESTIMACIONES EN TI 500000000		ESTIMACIONES EN TI 500000000	
A1	-1.2500000000000000	A1	-1.2500000000000000
A2	0.7000000000000000	A2	0.7000000000000000
B0	1.0000000000000000	B0	1.0000000000000000
B1	0.0000000000000000	B1	0.0000000000000000
V	10.0000000000000000	V	10.0000000000000000

Quédese que ya para $N = 50$ la diferencia más alta corresponde al parámetro b_0 (es de: 0.0395), siendo el único parámetro cuya diferencia de valores estimados es mayor que 1.5%.

Para N creciente esta diferencia disminuye, para $N = 200$ las diferencias máximas son del orden de 0.1% y para $N = 500$ las diferencias máximas son del orden de 0.4%.

Aunque no se sigue una ley exacta, parecería existir la tendencia a disminuir según $1/N$ las diferencias entre ambos estimadores. También para las estimaciones de las varianzas de los parámetros la proximidad es bastante alta a partir de $N = 50$. Esto sin embargo, no es siempre así, veremos en seguida un ejemplo.

En la tabla 5.VIII se presentan una serie de valores de los parámetros estimados con las dos inicializaciones, para el sistema de segundo orden y con una relación señal a ruido $S/N = 10$.

Los resultados se tabulan para $N = 10, 20, 50, 100, 200$ y 300. A partir de: $N = 50$ las diferencias son muy pequeñas; en $N = 50$ la diferencia más grande corresponde también a b_0 , y es de 0.06423, siendo el único parámetro con una diferencia de valores estimados mayor que el 1%.

Otra vez, para N creciente esta diferencia disminuye. El peor de los casos sigue siendo la diferencia correspondiente a b_0 , que decrece según $1/N$ bastante aproximadamente.

Desde hay una diferencia apreciablemente grande es en la estimación de la varianza de los estimadores. Esta proviene de una imperfecta estimación de σ^2 y no de una mala estimación de $[(X^T X)^{-1}]$.

COMPARACION DE RESULTADOS DE ESTIMACION POR MÍNIMOS CUADRADOS CON INICIALIZACIONES PARA EL SISTEMA DE ORDEN 2.

Tabla de varianzas estimadas de los parámetros, $\sigma_{\hat{\theta}_i}^2$, en donde $i=1, 2, 3$; $\sigma_{\hat{\theta}_i}^2$ es la variancia de $\hat{\theta}_i$, $\sigma_{\hat{\theta}_i}^2$ es la variancia de $\hat{\theta}_i$, $\sigma_{\hat{\theta}_i}^2$ es la variancia de $\hat{\theta}_i$.

INICIALIZACION ADELANTADA.			INICIALIZACION TARDÍA.		
ESTIMADOS EN TI 1000000000			ESTIMADOS EN TI 1000000000		
A 1	1.520000000	0.00000	A 1	1.520000000	0.00000
A 2	0.000000000	0.00000	A 2	0.000000000	0.00000
B 0	0.000000000	0.00000	B 0	0.000000000	0.00000
B 1	0.000000000	0.00000	B 1	0.000000000	0.00000
V	0.000000000	0.00000	V	0.000000000	0.00000
ESTIMADOS EN TI 2000000000			ESTIMADOS EN TI 2000000000		
A 1	1.520000000	0.00000	A 1	1.520000000	0.00000
A 2	0.000000000	0.00000	A 2	0.000000000	0.00000
B 0	0.000000000	0.00000	B 0	0.000000000	0.00000
B 1	0.000000000	0.00000	B 1	0.000000000	0.00000
V	0.000000000	0.00000	V	0.000000000	0.00000
ESTIMADOS EN TI 5000000000			ESTIMADOS EN TI 5000000000		
A 1	1.520000000	0.00000	A 1	1.520000000	0.00000
A 2	0.000000000	0.00000	A 2	0.000000000	0.00000
B 0	0.000000000	0.00000	B 0	0.000000000	0.00000
B 1	0.000000000	0.00000	B 1	0.000000000	0.00000
V	0.000000000	0.00000	V	0.000000000	0.00000
ESTIMADOS EN TI 10000000000			ESTIMADOS EN TI 10000000000		
A 1	1.520000000	0.00000	A 1	1.520000000	0.00000
A 2	0.000000000	0.00000	A 2	0.000000000	0.00000
B 0	0.000000000	0.00000	B 0	0.000000000	0.00000
B 1	0.000000000	0.00000	B 1	0.000000000	0.00000
V	0.000000000	0.00000	V	0.000000000	0.00000
ESTIMADOS EN TI 20000000000			ESTIMADOS EN TI 20000000000		
A 1	1.520000000	0.00000	A 1	1.520000000	0.00000
A 2	0.000000000	0.00000	A 2	0.000000000	0.00000
B 0	0.000000000	0.00000	B 0	0.000000000	0.00000
B 1	0.000000000	0.00000	B 1	0.000000000	0.00000
V	0.000000000	0.00000	V	0.000000000	0.00000
ESTIMADOS EN TI 50000000000			ESTIMADOS EN TI 50000000000		
A 1	1.520000000	0.00000	A 1	1.520000000	0.00000
A 2	0.000000000	0.00000	A 2	0.000000000	0.00000
B 0	0.000000000	0.00000	B 0	0.000000000	0.00000
B 1	0.000000000	0.00000	B 1	0.000000000	0.00000
V	0.000000000	0.00000	V	0.000000000	0.00000

Como ejemplo puede verificarse en la tabla que corrigiendo con las estimaciones de σ^2 , la diagonal de $(X^T X)^{-1}$ coincide en sus tres primeras cifras decimales desde $N = 10$.

Pero la diferencia no es alarmante si percibimos que la sobreestimación de σ^2 proviene de altos errores de predicción en los primeros datos. Alcanza entonces con considerar otro estimador de σ^2 , que no tome tan en cuenta los primeros datos.

Usualmente esto no es necesario puesto que se trabaja con alguna clase de olvido (p. ej: exponencial), y naturalmente los primeros datos son rápidamente olvidados y la aproximación mejora notablemente, tanto en la estimación de los parámetros como en sus varianzas.

5.12.7 Extensiones de mínimos cuadrados: sistemas variables en el tiempo y sistemas con término constante

a. Sistemas variables en el tiempo

Hemos discutido en la sección 5.8 algunas formas de abordar el problema de la identificación de sistemas cuyos parámetros varían en el tiempo. En este párrafo se presentan algunos ejemplos del uso del algoritmo con olvido exponencial visto en el párrafo 5.8.1. Las ecuaciones recursivas son (5.116).

El valor de α (factor de olvido) debe ser tal que $0 < \alpha \leq 1$. Si $\alpha = 1$ el algoritmo es el de mínimos cuadrados ordinario pues debemos recordar que estamos minimizando en $\hat{\theta}$ la función de costo:

$$\hat{S}_N^2(\theta) = e_N^2 + \alpha e_{N-1}^2 + \dots + \alpha^{N-1} e_1^2$$

donde e_i es el residuo correspondiente a la salida y_i .

Evidentemente la elección de α es crucial y su elección es en todos los casos un compromiso entre dos factores: el primero es que menores valores de α permiten un seguimiento más ruidoso de la variación de los parámetros, el segundo es que menores valores de α implican es-

timados más variables, menos estables.

Debemos pues elegir valores de α que den resultados aceptables en ambos sentidos, en la figura 5.1 se dibujaron las gráficas de los pesos relativos de los residuos para distintos valores de α .

Presentaremos primero dos ejemplos. El primero es de un sistema de orden 1 y el segundo es de un sistema de orden 2. El primer caso se trata de un sistema que cumple la ley (5.171) y en $t = 500$ cambia su ley a (5.172)

$$\begin{cases} y(t) = 0.8 y(t-1) + u(t) + \alpha e(t) & (5.171) \\ y(t) = 0.6 y(t-1) + 1.3 u(t) + \alpha e(t) & (5.172) \end{cases}$$

donde: $\sigma = 0.316228$, $u(t)$ es ruido blanco de distribución normal (0, 1). hasta $t = 500$ la relación señal a ruido es $S/N = 10$, desde $t = 500$ la relación $S/N = 16.9$.

Como en casos anteriores, la entrada utilizada fue PRNG127, la fuente de ruido es RUIDOS. Se simularon 1127 datos y se suprimieron los primeros 127. La identificación fue llevada a cabo con el programa MICU/S/OLVEXP y el resultado de la evolución del estimador se presenta en la figura (5.56). El valor utilizado para α fue de 0.98

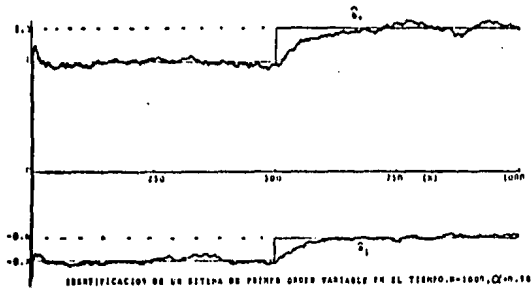


Fig 5.56

Puede observarse que luego de unos 120 instantes de muestreo los estimadores de los parámetros quedan bastante próximos al valor real que se graficó también. Compárese la figura (5.56) con la figura (5.11) que corresponde a la identificación de un sistema de primer orden invariante en el tiempo, con el algoritmo de mínimos cuadrados ordinario. Am-
 bos casos tienen relaciones señal a ruido similares, pero mientras que en un caso las varianzas de los estimadores son proporcionales a $1/N$, en el otro caso no, (veremos en seguida que son proporcionales a $\frac{1+\alpha^N}{(1-\alpha^N)(1+\alpha)}$ que para $N \rightarrow \infty$ es $1/(1-\alpha)$). En el caso que estamos tratando esto indica que para $N \rightarrow \infty$ se esperan varianzas de los estimadores equivalentes a las que se obtendrían en el primer caso con $N = 99 = 1 + 0.99/(1-0.98)$.

Un indicador de la velocidad con que el estimador puede seguir las variaciones de los parámetros es saber cuántos de los últimos pesos de los residuos suman un porcentaje determinado, (por ejemplo el 90%) del total de la suma de los pesos α^{N-1} de todos los residuos e_j . En

la figura (5.57) se grafica la curva correspondiente al 90%.

Por ejemplo, para $\alpha = 0.98$ los últimos 114 pesos contienen más del 90% de la suma total; esta coincide con la velocidad de seguimiento encontrada en el ejemplo anterior.

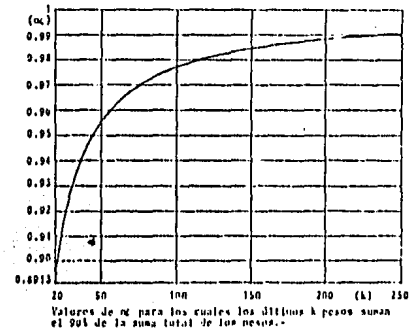


Fig 5.57

Valores de N para otros porcentajes p distintos del 90% pueden ser obtenidos de la figura (5.1) como sigue:

Para $N \rightarrow \infty$ la suma de todos los pesos desde $t = 1$ a $t = N$ tiende a $1/(1-\alpha)$; la suma de los últimos K pesos es:

$$\sum_{j=N-k+1}^{j=N} \alpha^{N-j} = \frac{1 - \alpha^k}{1 - \alpha} \quad (5.173)$$

si se quiere que esto represente un porcentaje $p \times 100$ del total $(1/1-\alpha)$ entonces:

$$\frac{1-\alpha^k}{1-\alpha} = p \frac{1}{1-\alpha} \quad (5.174)$$

$$\therefore \alpha^k = 1 - p \quad (5.175)$$

de esta última ecuación se desprende que para calcular cuantos pesos (k) contienen el porcentaje $p \times 100$ del total para un valor de α determinado, basta ver en la figura (5.1) el valor de N para el cual α^N corta la línea $1-p$.

El segundo ejemplo es un sistema de orden 2 cuyos parámetros a_1 y b_2 varían bruscamente en $t = 500$. Las ecuaciones (5.176) y (5.177) representan el sistema antes y después de $t = 500$:

$$\left\{ \begin{array}{l} y(t) = 1.5y(t-1) - 0.7y(t-2) + u(t-1) + 0.5u(t-2) + \alpha e(t) \quad (5.176) \\ y(t) = y(t-1) - 0.7y(t-2) + 1.3u(t-1) + 0.5u(t-2) + \alpha e(t) \quad (5.177) \end{array} \right.$$

siendo:

$$\left\{ \begin{array}{l} \alpha = 0.461774, e(t) \text{ ruido blanco con distribución normal } (0,1) \text{ y} \\ u(t) \text{ es ruido blanco con media cero y varianza unitaria.} \end{array} \right.$$

Con el valor elegido de α la relación señal a ruido S/N es 10 hasta $t = 500$, y 12.68 en adelante.

El proceso se simuló con $u(t) = \text{PRNG127}$, y perturbación $e(t)$; PNTD05 . Se simularon 1127 datos de los cuales se descartaron los primeros 127 y se identificó el sistema sobre los 1000 datos restantes con el programa HICU/S/OLVEXP, con $\alpha = 0.985$.

El resultado de la identificación se grafica en la figura (5.58), contra los valores reales de los parámetros y el valor de k de la figura (5.57) para $\alpha = 0.985$.

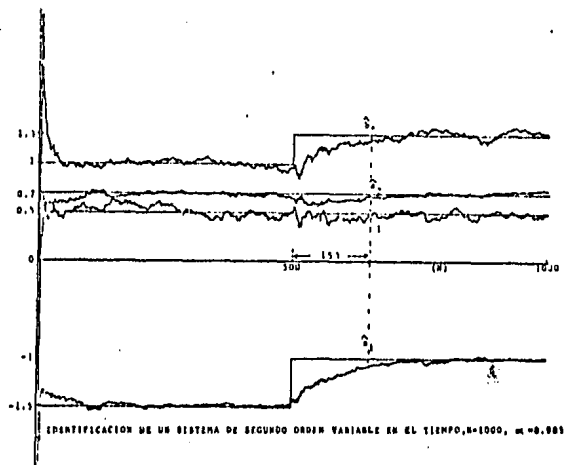


Fig 5.58

En general el resultado es bueno, los estimadores son cercanos a los valores reales. Sin embargo el olvido exponencial introduce una mayor varianza en la estimación. Compárese por ejemplo con la figura (5.15) donde se presenta la identificación del sistema para $t < 500$ con la misma relación señal a ruido.

De la figura (5.57) se prevé del orden de 153 instantes de muestreo para alcanzar una proximidad relativa al nuevo valor de los parámetros. Estas 153 instantes se dibujan en la gráfica, y a simple vista se comprueban en el resultado de la figura (5.58) como un indicador razonable.

Ejemplificaremos ahora con una simulación el efecto de variar el factor de olvido α . El sistema considerado es de primer orden, con una relación señal a ruido constante igual a 10. El sistema varía bruscamente el parámetro a_1 en $t = 500$; la ecuación en diferencias del sistema es (5.178) hasta $t = 500$ y (5.179) a partir de $t = 501$.

$$\begin{cases} y(t) [1 - 0.9 q^{-1}] = q^{-1} u(t) + \sigma e(t) & (5.178) \\ y(t) [1 - 0.3 q^{-1}] = q^{-1} u(t) + \sigma e(t) & (5.179) \end{cases}$$

con $\sigma = 0.316228$, $e(t)$ ruido blanco gaussiano y $u(t)$ ruido blanco de media cero y varianza = 1. El sistema fue simulado con entrada $u(t)$ PRD:127 y perturbación $e(t)$ RUIR05, como en casos anteriores, se simularon 1127 datos de salida, de los que se eliminaron los primeros 127. Con los 1000 datos restantes se efectuó la identifica-

ción del sistema, cuatro veces, con distintos valores de α : 0.99, 0.985, 0.98, 0.97. El programa utilizado fue MTCUS/OLVIDE.P.

Los resultados de las simulaciones se grafican en las figuras (5.59) a (5.62) contra el valor real de los parámetros.

Se han dibujado además los puntos para los cuales los pesos de los residuos posteriores al cambio del valor del parámetro suman más del 90% del total.

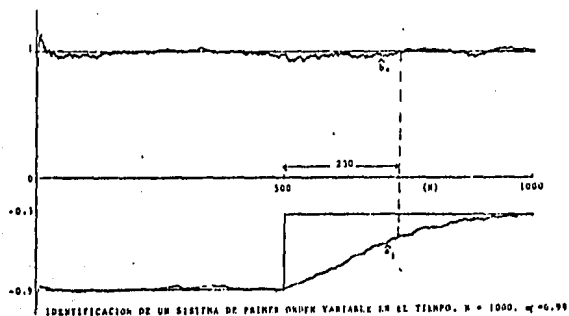


Fig 5.59

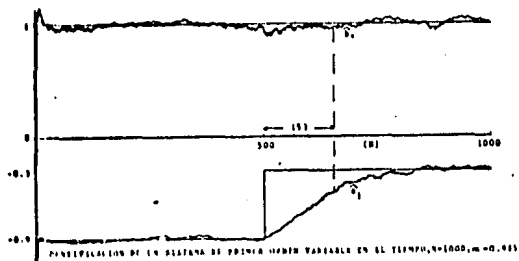


Fig 5.60

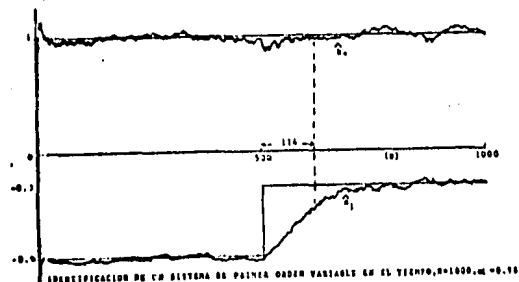


Fig 5.61

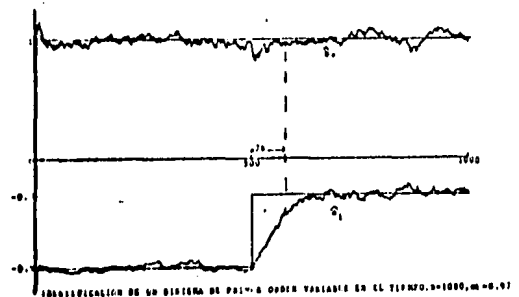


Fig 5.62

Puede verificarse que disminuyendo el valor de α se gana velocidad en el seguimiento de las variaciones de los parámetros, a costa de una mayor varianza en la estimación, lo que implica menor precisión.

No se puede establecer *a priori* qué es mejor, si mayor precisión o capacidad de seguimiento; la elección de α debe ser hecha tomando en cuenta en primer lugar el objetivo de la identificación y, tomando en cuenta la relación señal a ruido del sistema y la velocidad esperada en las variaciones de los parámetros.

En efecto: un nivel alto de S/N permite valores de α más pequeños para un valor fijo de la varianza de las estimaciones, mientras que variaciones lentas de los parámetros admiten valores altos de α sin afectar su seguimiento.

En aplicaciones de ingeniería, resulta entonces muy importante el tener en cuenta todo el conocimiento *a priori* del sistema en cuestión, y efectuar una selección *ad hoc* de α .

El efecto de incluir el factor de olvido α respecto la pérdida de precisión de la estimación se hace más patente si comparamos las últimas cuatro figuras con la figura (5.21) que representa la identificación del sistema de primer orden invariante en el tiempo:

$$y(t) \left[1 - 0.8 \right] q^{-1} = q^{-1} u(t) + \omega(t)$$

con la misma relación señal a ruido $S/N = 10$, pero sin factor de olvido (es decir $\alpha = 1$).

Un enfoque que puede dar algo de intuición al problema de la influencia de α sobre la varianza de la estimación, es considerar el modelo estadístico lineal (2.58):

$$Y = X\beta + e$$

con X y e independientes, y $E(ee^T) = \sigma^2 I$.

En ese caso la matriz de covarianzas del estimador de mínimos cuadrados pesados es:

$$\text{cov } \hat{O}_N = \sigma^2 \{E(X^T W X)^{-1} X^T W W X (X^T W X)^{-1}\}$$

donde W es la matriz de pesos:

$$W = \begin{bmatrix} \alpha^{N-1} & & & 0 \\ & \alpha^{N-2} & & \\ & & \ddots & \\ & & & \alpha \\ 0 & & & & 1 \end{bmatrix}$$

Consideremos matriz $X = \begin{bmatrix} x_1^T \\ \vdots \\ x_N^T \end{bmatrix}$ cuyas columnas son v.a. de una misma distribución, n. ej: cada columna representa muestras de un proceso estacionario ergódico. Entonces en virtud del teorema B.2, resulta:

$$\text{cov } \hat{O}_N = \sigma^2 \frac{\sum_{i=0}^{N-1} (\alpha^i)^2}{\left(\sum_{i=0}^{N-1} \alpha^i \right)^2} E(x_i x_i^T)^{-1}$$

esto es que:

$$\text{cov } \hat{O}_N = \sigma^2 \frac{1+\alpha}{1-\alpha} \frac{1-\alpha}{1+\alpha} E(x_i x_i^T)^{-1} \quad (5.180)$$

compárese con la matriz de covarianzas del estimador de mínimos cuadrados ordinario:

$$\frac{\sigma^2}{N} E(x_i x_i^T)^{-1}$$

Mientras que la matriz de covarianzas tiende a la matriz nula en el estimador de mínimos cuadrados ordinario, en el estimador con olvido exponencial

$$\text{cov}(\hat{O}_N) \xrightarrow{P} \sigma^2 \frac{1-\alpha}{1+\alpha} E(x_i x_i^T)^{-1} = \text{cov}(\hat{O}_\infty) \quad (5.181)$$

para $N \rightarrow \infty$

de esta forma, si $0 < \alpha < 1$ se obtiene una matriz de covarianzas \hat{O}_N para $N \rightarrow \infty$ equivalente a la que se obtendría para un número N^* finito de observaciones:

$$N^* = \frac{1+\alpha}{1-\alpha}$$

en el caso que tuviéramos una estimación de mínimos cuadrados ordinaria ($\alpha = 1$). Por ejemplo $\alpha = 0.97$ $N^* = 65.7$, $\alpha = 0.98$ $N^* = 99$, $\alpha = 0.985$ $N^* = 132.3$, $\alpha = 0.99$ $N^* = 199$.

b. Sistemas con término constante

El método de mínimos cuadrados estudiado en este capítulo puede ser extendido a representaciones paramétricas más amplias que la que hemos visto, por ejemplo a sistemas de varias entradas y varias salidas.

Cuando se extienden los resultados a varias entradas una salida, las condiciones del teorema 5.1 se extienden fácilmente para asegurar la existencia de

$$\lim_{N \rightarrow \infty} \frac{1}{N} X_N^T X_N \quad \text{y} \quad \lim_{N \rightarrow \infty} \frac{1}{N} X_N^T u_N$$

un caso particular que puede ser entendido así, es el de los sistemas que tienen un constante aditiva, es decir, que se rigen por una ley del tipo:

$$y(t) = -a_1 y(t-1) - \dots - a_n y(t-n) + b_0 u(t) + \dots + b_m u(t-m) + c + \sigma e(t) \quad (5.182)$$

Estrictamente hablando no se trata de un sistema lineal. Puede en -
tenderse este caso como el de un sistema con dos entradas (ver fig 5.63), una de las cuales permanece constante u (igual a 1).

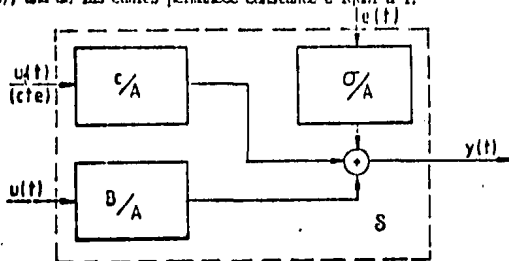


Fig 5.63

Escribiendo (5.182) en forma matricial para N observaciones de $y(t)$, $u(t)$ y $u_1(t)$ como en el modelo lineal (2.58):

$$Y_N = X_N \theta + \sigma_N$$

con

$$Y_N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}; \quad X_N = \begin{bmatrix} x_{11}^T \\ \vdots \\ x_{N1}^T \end{bmatrix}; \quad \begin{cases} x_{i1}^T = (-y(i-1), \dots, -y(i-n), u(i), \dots, \\ u(i-m), 1) \\ \sigma_N^T = (\sigma u(1), \sigma u(2), \dots, \sigma u(N)) \\ \theta^T = (a_1, \dots, a_n, b_0, \dots, b_m, c) \end{cases} \quad (5.183)$$

con el mismo criterio que antes, el estimador de mínimos cuadrados es:

$$\hat{\theta}_N = (X_N^T X_N)^{-1} X_N^T Y_N$$

La diferencia con el método empleado hasta ahora es mínima: una columna de unos en la matriz X_N ; el algoritmo ha sido implantado en los dos esquemas recursivos vistos en el capítulo con los programas:

MICU/C que realiza el algoritmo de Plackett

SROOT/C que realiza el algoritmo square-root.

Ambos programas están escritos en Pascal para la computadora B-6800 del CEC de la UNAM. Su listado puede encontrarse en el apéndice U. Se realizaron numerosas pruebas con ambos programas, todas fueron exitosas.

Como ejemplo se presenta una simulación. Se trata de un sistema de primer orden una relación señal a ruido $S/N = 1$.

$$y(t) = 0.6 y(t-1) + u(t-1) + 0.5 + \sigma e(t) \quad (5.184)$$

Fueron simuladas 1127 observaciones a partir de condiciones iniciales nulas, suprimiéndose los 127 primeros resultados.

Con los datos restantes se efectuó la identificación del sistema, con el programa SHOOT/C inicializado en forma exacta.

La evolución de la estimación $\hat{\theta}_N$ con $1 \leq N \leq 1000$ se muestra en la figura (5.64)

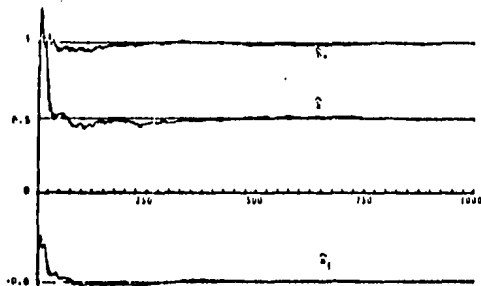


Fig 5.64

La estimación es excelente.

Este procedimiento permite identificar el sistema al mismo tiempo que el sesgo de la salida, sin necesidad de un tratamiento previo del estudio de promediar la salida para restar su valor medio o filtrar ésta con un filtro pasabajas.

Este término constante que parece agregar un simple detalle con el método de mínimos cuadrados, puede transformarse en un escollo importante en los métodos de correlación, especialmente cuando no es constante en el tiempo, aún cuando describa variaciones lentas.

Este caso podría atacarse directamente con el método de mínimos cuadrados agregando alguna clase de olvido que permita un seguimiento de sus variaciones.

5.12.8 Ruido correlacionado

Como visto en la sección 5.3 que es clave para que el estimador de mínimos cuadrados converja al parámetro real, que el proceso del ruido sea no correlacionado.

Más aún, la expresión (5.42) señala cuál es el sesgo asintótico de la estimación. Lo verificaremos con un ejemplo.

Sea el sistema de primer orden (5.145) donde la perturbación en lugar de ser e^* es $e(t)$ es: $e^*(t) = \lambda(u(t) + ru(t-1))$, con $e(t)$ ruido blanco gaussiano de media cero y varianza unitaria.

El sistema es:

$$[1 - 0.8q^{-1}]y(t) = q^{-1}u(t) + \lambda[1 + rq^{-1}]e(t)$$

de acuerdo a (5.142):

$$\hat{\theta}_N - \theta = E\left\{\frac{1}{N} X_N^T X_N\right\}^{-1} E\left\{\frac{1}{N} X_N^T e_N^*\right\} \quad (5.185)$$

$$\therefore \hat{\Theta}_m - \Theta = \begin{bmatrix} R_{yy}(0) & 0 \\ 0 & R_{uu}(0) \end{bmatrix}^{-1} \begin{bmatrix} -R_{yu}(1) \\ 0 \end{bmatrix} \quad (5.186)$$

como $R_{uu}(0) = 1$, si se consideran $\lambda = 1.64399$ y $r = -0.7$ entonces resulta $S/N = 1$, $R_{yy}(0) = 2 \times 2.7778 = 5.5556$.

Además, como $R_{yu}(1) = r\lambda + R_{yu} = -1.09109$

$$\therefore \hat{\Theta}_m - \Theta = \begin{bmatrix} +0.34054 \\ 0 \end{bmatrix} \quad \text{como } \Theta = \begin{bmatrix} -0.6 \\ 1 \end{bmatrix}$$

$$\rightarrow \hat{\Theta}_m = \begin{bmatrix} -0.4595 \\ 1 \end{bmatrix} \quad (5.187)$$

es decir que la estimación resultará bastante sesgada para el parámetro a_1 , el sesgo es aproximadamente el 43% de a_1 .

En cambio para el parámetro b_0 la estimación sigue siendo insesgada, esto se debe a que la matriz $E\left\{\frac{1}{N} \begin{pmatrix} x_N^T & x_N \end{pmatrix}\right\}^{-1}$ es diagonal, y el vector $E\left\{\frac{1}{N} \begin{pmatrix} x_N^T & e_N^T \end{pmatrix}\right\}$ tiene elementos distintos de cero sólo en el lugar superior.

Ya se había advertido en la observación de la sección 5.3 que aún cuando el ruido fuera correlacionado, las últimas ecuaciones del sistema (5.34) que de la solución de mínimos cuadrados, siguen siendo válidas.

Se simuló un proceso de la forma descrita con el sistema de primer orden, entrada $u(t) = \text{PRNR127}$, $e(t) = \text{RUID05}$. Se consideraron 1127 observaciones de las cuales se descartaron las 127 primeras.

Con estos datos se realizó la identificación del sistema con el algoritmo de mínimos cuadrados inicializado en forma exacta.

La evolución del estimador desde $N = 1$ hasta $N = 1000$ se dibuja en la figura (5.65). En ella se dibujan con línea llena los valores reales de a_1 (-0.6) y $b_0(1)$, y en línea punteada la estimación sesgada de a_1 que prevé la ecuación (5.186) cuya magnitud está dada por (5.187).

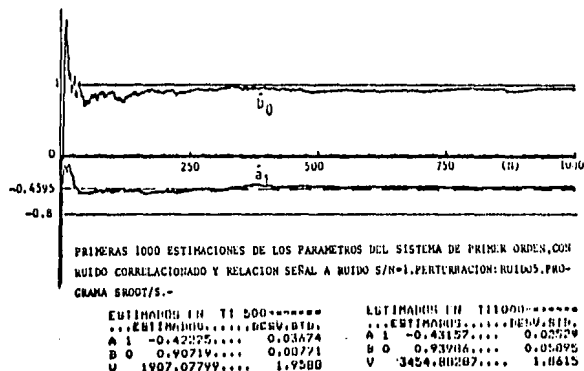


Fig 5.65

Efectivamente la estimación resulta sesgada para a_1 y para el parámetro b_0 aunque la estimación no es muy buena, puede pensarse que no hay sesgo.

Comparese la figura (5.65) con la figura (5.13) que corresponden al mismo sistema con la misma relación señal a ruido pero donde las perturbaciones eran independientes. La diferencia es notable, en primer lugar por el sesgo de a_1 y en segundo lugar porque en el caso que estamos discutiendo en este párrafo, la varianza de b_0 parece ser bastante más alta. Para ello basta observar que en la identificación descrita en la figura (5.13) la estimación de b_0 en 200 observaciones es sensiblemente mejor que en la de este caso con 1000 observaciones tanto por el valor de b_0 como por sus oscilaciones.

En este ejemplo se comprueba tanto la existencia del sesgo como la magnitud que prevé la teoría.

5.12.9 Perturbaciones no gaussianas

Hasta ahora en las simulaciones presentadas nos hemos limitado a perturbar los sistemas tratados con ruido blanco gaussiano (o al menos con señales que lo simulan).

Como se ha visto, la condición de no correlación de la perturbación en distintos instantes de tiempo es una restricción importante puesto que en el caso general es condición necesaria para obtener una estimación insesgada.

Por otro lado la forma de la distribución del ruido no es una limitación seria; más aún, las propiedades asintóticas se siguen cumpliendo para distribuciones no gaussianas.

Se ha elegido simular con señales gaussianas ya que el teorema del límite central mediante, es el caso más importante en las aplicaciones normales.

Consideremos por ejemplo el teorema 5.1 de convergencia del método de mínimos cuadrados. Supongamos que se cumplen sus hipótesis y mantengamos fija la varianza σ^2 del ruido variando su distribución. Si el ruido $e(t)$ es gaussiano, el estimador \hat{O}_N tiende a una ley normal

$$\hat{O}_N \xrightarrow[N \rightarrow \infty]{\text{ley}} \mathcal{N}(O, \frac{\sigma^2}{N} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i x_i^T)$$

si $u(t)$ es un proceso estacionario, la matriz asintótica de covarianzas es la inversa de la matriz de Fisher, y es por lo tanto la cota mínima de Cramer-Rao.

Sin embargo, para otras distribuciones ésta no es forzosamente la cota de Cramer-Rao y el teorema 2.10 asegura que asintóticamente el estimador de máxima verosimilitud es al menos no peor que el de mínimos cuadrados.

Por lo tanto, desde este punto de vista, el caso de ruido gaussiano es el peor en el sentido que la matriz de covarianzas asintótica:

$$\text{cov}(\hat{O}_N) = \frac{\sigma^2}{N} E \{x_i x_i^T\}^T \quad (5.188)$$

no puede ser mejorada con otro estimador (ver teorema 2.5), en cambio para el caso de otras distribuciones, si u es un proceso estacionario, la matriz $\text{cov}(\hat{0}_H)$ tiene el mismo valor puesto que depende sólo de las estadísticas de 2º orden, y sin embargo es factible mejorarla con otro estimador distinto que el de mínimos cuadrados.

De esto que hemos dicho vale la pena reaffirmar dos cosas:

- la primera es que la matriz de covarianzas $\text{cov}(\hat{0}_H)$ asintótica es la misma para $u(t)$ estacionario, o¹ etc, y el sistema bajo las hipótesis del teorema 5.1; es decir que asintóticamente la forma de la distribución de $e(t)$ no importa
- la segunda es que aún siendo factible mejorar la covarianza del estimador considerando otra forma de estimar, es enojosa la afirmación de que el caso gaussiano es el peor desde este punto de vista, ya que aún cuando la matriz de covarianzas sea la misma asintóticamente, no hemos dicho nada acerca de cómo se tiende a dicha matriz, y en el caso no gaussiano, dicha tendencia puede ser más lenta.

Para verificar la bondad del estimador de mínimos cuadrados con perturbaciones no gaussianas se han hecho algunas simulaciones. Se simularon los sistemas de primer y segundo orden con entrada PRBS127 y perturbaciones: uniforme (RUID07) y binaria pseudo-aleatoria (ENT1) con el programa SIMULA, para distintas relaciones señal a ruido.

Los primeros 127 datos se descartaron y sobre las siguientes se realizó la identificación del sistema correspondiente.

La figura (5.66) muestra la evolución del estimador de los parámetros del sistema de primer orden para $S/N = 10$ y perturbación ENT1 para 175 puntos de estimación.

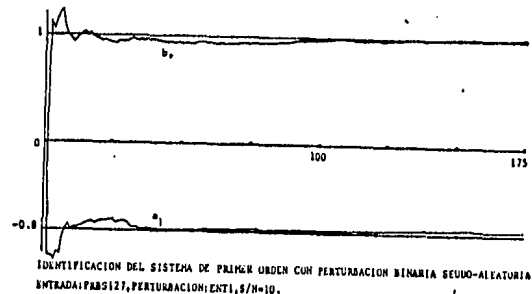


Fig 5.66

La estimación es buena. De los valores tabulados, las estimaciones de b_0 se encuentran dentro de 1 desv. std. (esperada); para a_1 la estimación es algo inferior, en $N = 100$ el estimador está a 1.462 desv. std. y para $N = 175$ a 2.7 desv. std.

Se puede comparar la estimación con la de la figura (5.11) que corresponde al mismo sistema con $S/N = 10$ y perturbación gaussiana, en aquel caso la estimación es algo mejor especialmente la del parámetro a_1 . Prolongando la estimación se ha encontrado que al alejamiento de \hat{a}_1

nostrato la desviación prevista por (5.156) va disminuyendo.

La figura (5.67) muestra la identificación del mismo sistema con perturbación uniforme, y una relación señal a ruido: $S/N = 1$.

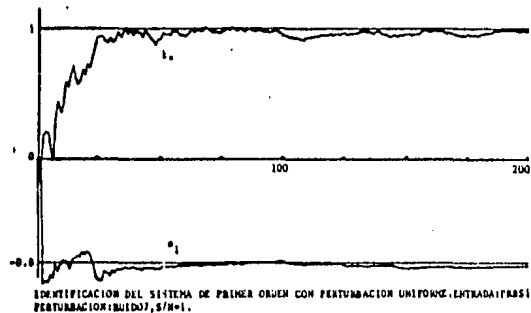


Fig 5.67

En este caso la estimación es mejor aún: todos los valores tabulados se encuentran en un entorno de ± 1 desv. std (dada por (5.156)). Comparando con la figura (5.43) que corresponde a la identificación del mismo sistema con $S/N = 1$ y perturbación gaussiana, parecería esta última el go mejor que la estimación de la figura (5.67).

La figura (5.68) muestra la identificación del sistema de segundo orden perturbado con una señal binaria pseudo aleatoria y $S/N = 1$

Para $N = 200$ los estimadores se encuentran dentro del entorno ± 1.96

desv. std. (de 95% de confianza), incluso \hat{b}_0 y \hat{b}_1 se encuentran dentro del entorno ± 1 desv. std.

Para $N = 500$ todos los estimadores excepto \hat{b}_1 , se encuentran dentro del entorno ± 1 desv. std; el estimador \hat{b}_1 se encuentra a 2.02 desv. std.

La figura (5.16) corresponde a la identificación del mismo sistema perturbado con una señal gaussiana y con el mismo valor de S/N . Comparándose ambas identificaciones, son similares.

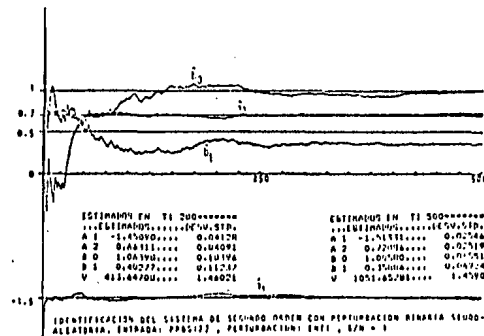


Fig 5.68

En el caso gaussiano resulta algo peor la estimación de a_1 y a_2 , son similares en ambos casos la de b_0 , y en el caso gaussiano es algo mejor la estimación de b_1 . Este parámetro es el peor estimado en ambos casos, y es donde, de acuerdo a (5.161) se esperaba una mayor varianza en la estimación.

En la figura (5.69) se muestra la identificación del mismo sistema con una relación señal a ruido $S/N = 1$, cuya perturbación es uniforme.

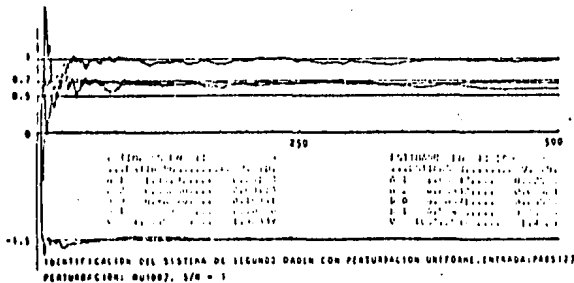


Fig 5.69

La estimación sigue de acuerdo a las previsiones. De los estimados tabulados (para $N = 200$ y $N = 500$) todos se encuentran dentro del entorno de ± 1.95 desv. std. (de 95% de confianza) alrededor del valor real.

Las desviaciones std. esperadas se han calculado con la matriz (5.161).

De todos los valores tabulados el más alejado respecto a su desv. std. es el estimador \hat{a}_2 para $N = 500$ que se encuentra a 1.485 desv. std. La estimación es muy similar a la figura (5.16) desde el punto de vista de la proximidad de los estimadores a los valores reales.

En los ejemplos presentados se ha observado una gran consistencia de los resultados respecto la tendencia asintótica esperada, aún para situaciones con distribuciones uniforme y binaria. Los resultados obtenidos son muy similares a los obtenidos en condiciones análogas para perturbaciones gaussianas.

5.12.10 Determinación de la respuesta a pulso

Estimación Indirecta

Aunque el método de mínimos cuadrados está orientado fundamentalmente a la determinación de la función de transferencia de un sistema en tiempo discreto, tal y como lo hemos visto, a través de él se puede obtener una estimación de la respuesta a pulso bastante más precisa que con algunos métodos orientados a este fin.

Suponiendo estimada la función de transferencia, alcanza con considerar como estimador de la sucesión de peso del sistema, la del sistema estimado.

Si se cumplen las condiciones necesarias para la convergencia del estimador de mínimos cuadrados, es fácil mostrar que también el estimador de la sucesión de peso converge a su valor real, esto es que si llamamos $\hat{h}_N(k)$ a la sucesión de peso estimada con N observaciones y $h(k)$ a la sucesión real, entonces:

$$\hat{h}_N(k) \xrightarrow{\text{Prob}} h(k) \text{ para } N \rightarrow \infty \text{ y } \forall k \text{ finito}$$

En efecto: $h(k)$ satisface la recursión (5.149), suponiendo $n = m$ sin pérdida de generalidad, para la parte determinística del modelo (5.2)

$$\begin{aligned}
 b_0 &= h(0) \\
 b_1 &= h(0) a_1 + h(1) \\
 b_2 &= h(0) a_1^2 + h(1) a_1 + h(2) \\
 &\vdots \\
 b_n &= h(0) a_1^n + h(1) a_1^{n-1} + h(2) a_1^{n-2} + \dots + h(n) \\
 0 &= \sum_{i=0}^n h(r-i) a_i, \text{ con } a_0=1, \forall r > n
 \end{aligned} \tag{5.189}$$

calculando la sucesión $\hat{h}_N(k)$ poniendo $\hat{b}_N i$ y $\hat{a}_N i$ en lugar de b_i y a_i respectivamente ($i=1, \dots, n$), donde los estimadores $\hat{a}_N i$, $\hat{b}_N i$ son los estimadores de a_i y b_i obtenidos por mínimos cuadrados para N observaciones.

Si se cumplen las condiciones del teorema 5.1, entonces:

$$\hat{O}_N^T = (\hat{a}_N 1, \hat{a}_N 2, \dots, \hat{a}_N n, \hat{b}_N 0, \hat{b}_N 1, \dots, \hat{b}_N n) \xrightarrow{m.c.} (a_1, \dots, a_n, b_0, \dots, b_n)$$

para $N \rightarrow \infty$. También converge en probabilidad.

Observamos que como $\hat{h}_N(k)$ satisface la recursión (5.189) es una función:

$$\hat{h}_N(k) = g(\hat{O}_N, k) \tag{5.190}$$

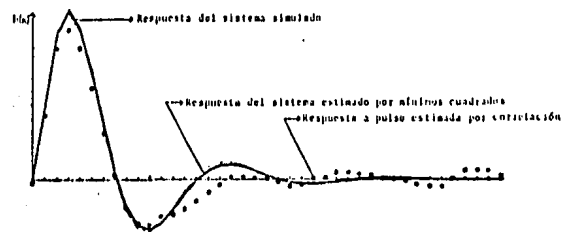
racional en los elementos del vector \hat{O}_N , luego, en virtud del teorema B.2 (ver apéndice B), se cumple que:

$$\hat{h}_N(k) \xrightarrow{\text{prob.}} h(k) \tag{5.191}$$

No discutiremos ahora la varianza de la estimación, para compararla con otros métodos de estimación de la respuesta a pulso nos limita-

remos a presentar un ejemplo.

Para el sistema de segundo orden con entrada PRBS127 y perturbado con RUIDO1, con $S/N = 4$ y $S/H = 1$ se le estimó la sucesión de peso $h(k)$ de esta forma, para 500 observaciones, con $0 \leq k < 40$. También se hizo la estimación de la sucesión de peso por métodos de correlación para el mismo número de observaciones, y se calculó la sucesión de peso real. La estimación de la sucesión de peso por correlación se llevó a cabo con el programa ESTIMA/PESO, en Pascal para la B6800 del CSC de la UNAM, su listado puede encontrarse en el apéndice D.



ESTIMACION DE LA RESPUESTA A PULSO DEL SISTEMA DE SEGUNDO ORDEN CON 500 PULSOS DE UNA REALIZACION CON ENTRADA PRBS127 Y $S/N=4$.

Fig. 5.70

La figura (5.70) muestra dichas estimaciones para una relación señal a ruido $S/N = 4$. Por el método de mínimos cuadrados el resultado de la identificación fue:

Parámetro	Valor estimado	Valor real
a_1	-1.50716	-1.5
a_2	0.71102	0.7
b_0	1.003	1
b_1	0.46368	0.5

En línea continua delgada se muestra la respuesta a pulso del sistema estimado, en línea gruesa la respuesta del sistema real, y por puntos la estimación por métodos de correlación.

La respuesta estimada indirectamente por mínimos cuadrados es lo suficientemente próxima a la real como para que sólo en algunas partes de la figura se lo pueda distinguir.

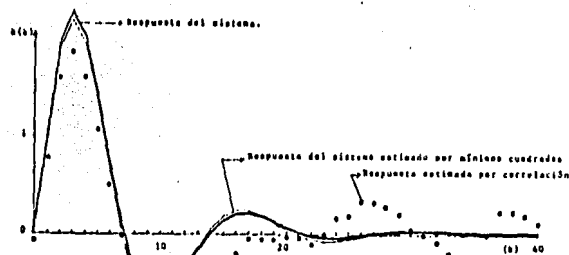
La estimación de $h(k)$ por correlación es bastante más pobre, y empiora muchísimo para $k > 12$.

En los primeros puntos no es muy inferior a la estimación de mínimos cuadrados, especialmente si se tiene en cuenta que en nuestro modelo hemos impuesto la condición $\hat{h}(0) = 0$ que es muy importante. En la figura (5.72) se muestra $\hat{h}(k)$ estimada por mínimos cuadrados sin imponer dicha restricción (en una simulación con $S/N = 1$).

En la figura (5.71) se repite lo mismo pero aumentando la potencia del ruido 4 veces, es decir $S/N = 1$.

La estimación por mínimos cuadrados ahora es:

Parámetro	Estimación	Valor real
a_1	-1.5123	-1.5
a_2	0.71812	0.7
b_0	1.000752	1
b_1	0.43027	0.5



ESTIMACIÓN DE LA RESPUESTA A PULSO DEL SISTEMA DE RUIDO BLANCO CON 500 DATOS DE UNA REALIZACIÓN CON POTENCIA DE RUIDO $S/N = 1$.

Fig 5.71

La estimación de $h(k)$ obtenida del modelo estimado por mínimos cuadrados es ligeramente peor que la anterior, en cambio la obtenida por correlación empírica algo para $k \leq 12$ y bastante más para $k > 12$.

Métodos directos

1. Entradas no aleatorias

En la sección 5.11 se describieron dos métodos de estimación de la respuesta a pulso aplicando el esquema de mínimos cuadrados. En el primero se modela el sistema como:

$$Y_N = U_N h + e_N$$

Debe distinguirse que las e_N no son las perturbaciones del modelo (5.2), son errores que se presentan incluso cuando no hay perturbaciones en el modelo (5.2), en tal situación serían errores de ajuste de un modelo de respuesta a pulso finita para un sistema que no lo es.

En ese caso puede pensarse que el modelo $B(q^{-1})/A(q^{-1})$ se aproxima con un modelo donde $A(q^{-1}) = 1$ aumentando el orden del polinomio $B(q^{-1})$.

Más aún, si existieran perturbaciones como en el modelo (5.2) tendríamos una estimación de la respuesta a pulso asintóticamente segura. En efecto, sea el sistema:

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})} u(t) + \frac{1}{A(q^{-1})} e(t)$$

si se puede aproximar la parte determinística como un filtro FIR, esto es que existe $B^1(q^{-1})$ (de orden finito), tal que $B(q^{-1})/A(q^{-1}) \approx B^1(q^{-1})$ hay que definir el criterio de proximidad; el de $\| \cdot \|_2$ p. ej. sería adecuado (más ha sentido con varianzas de salida similares para un mismo ruido blanco a la entrada).

Entonces

$$y(t) = B^1(q^{-1}) u(t) + \frac{1}{A(q^{-1})} e(t) \quad (5.192)$$

los coeficientes de $B^1(q^{-1})$ vienen a ser precisamente los mismos que buscamos. Nótese que al intentar identificar por mínimos cuadrados dichos pesos; estamos en presencia de un modelo del tipo (5.2) con perturbaciones correlacionadas; en general esto conduce a una estimación sesgada como hemos visto.

Sobre una simulación sin perturbaciones se estimó con este método la respuesta a pulso $h(k)$ $0 < k < 30$ del sistema de segundo orden aproximado con un filtro FIR de orden 30. Se consideraron 100 observaciones.

Los resultados se tabulan en la tabla 5.IX, son enormemente próximos a la respuesta real ya que la cola tiende rápidamente a cero. Si no fuera así no tendría sentido emplear el método para aproximar el sistema con un filtro FIR de pocos coeficientes.

Respuesta a pulso real Aproximación F.I.R. por mínimos cuadrados Estimación por correlación

HC 01)	0.00000	0.00395	0.203621
HC 1)	1.00000	1.00515	0.995379
HC 2)	2.00000	2.00624	1.900049
HC 3)	2.30000	2.30514	2.012599
HC 4)	2.05000	2.05344	1.726301
HC 5)	1.46500	1.46477	1.281707
HC 6)	0.76250	0.76107	0.753086
HC 7)	0.11825	0.11434	0.310566
HC 8)	-0.35637	-0.36034	0.113491
HC 9)	-0.61734	-0.62135	-0.096473
HC10)	-0.67654	-0.68149	-0.192271
HC11)	-0.58260	-0.58540	-0.457407
HC12)	-0.40044	-0.40017	-0.570555
HC13)	-0.19278	-0.19085	-0.713744
HC14)	-0.00887	-0.00788	-0.590404
HC15)	0.12165	0.12511	-0.323476
HC16)	0.18868	0.18976	-0.130785
HC17)	0.19786	0.19725	0.011782
HC18)	0.14472	0.14417	-0.009073
HC19)	0.10858	0.10570	-0.053216
HC20)	0.04758	0.04427	-0.107407
HC21)	-0.00466	-0.00567	-0.167580
HC22)	-0.04079	-0.04151	-0.149545
HC23)	-0.05717	-0.05547	-0.301723
HC24)	-0.05755	-0.05405	-0.515525
HC25)	-0.04631	-0.04319	-0.605400
HC26)	-0.02910	-0.02525	-0.599807
HC27)	-0.01135	-0.00975	-0.418717
HC28)	0.00340	0.00410	-0.223434
HC29)	0.01304	0.00834	-0.083763
HC30)	0.01719	0.01205	0.109004

En nuestro caso es completamente inútil efectuar la aproximación de esta forma, puesto que se trata de un sistema de orden finito que puede por lo tanto ser identificado en forma exacta con muchos más datos, y se hubiera tenido una excelente aproximación FIR truncando la cola de la respuesta a pulso que tiende rápidamente a cero.

Además el procedimiento es altamente ineficiente, porque requiere la inversión de una matriz de 31 x 31 elementos.

Una medida de la bondad del ajuste lo da el error cuadrático medio para las 100 salidas consideradas, es decir:

$$\text{e.c.m.} = \sqrt{\frac{1}{100} \sum_{N=1}^{100} |e_N|^2}$$

en la aproximación que se hizo esta cantidad fue: 0.017201. Recordando que la entrada es PRUS127 y que para el sistema de segundo orden con ruido blanco de varianza 1 a la entrada se espera una salida de varianza 10.88021, esto quiere decir que el error cuadrático medio a la salida representa aproximadamente el 0.4% de la amplitud esperada de la salida (en sentido cuadrático medio). La aproximación es muy buena para un rango muy amplio de propósitos.

En la tabla 5.IX se ha agregado para fines de comparación la estimación de la respuesta a pulso del sistema, $h(k)$, obtenida por correlación entre entrada y salida para las mismas 100 observaciones.

2. Entradas aleatorias

El segundo método de la sección 5.11 también apunta a resolver el problema de determinación de la respuesta a pulso del sistema, diferenciando del método anterior en varios aspectos, entre ellos:

- 1) La entrada y la perturbación deben ser procesos estocásticos. El proceso de entrada debe ser estacionario al menos en sentido amplio o débilmente estacionario, para que existan los momentos de segundo orden.

- 2) El método no es más que una extensión del método de correlación para el caso en que la señal de entrada no sea ruido blanco. Convergirá asintóticamente a una aproximación IIR del sistema, obviamente si existe τ_s tal que $h(i) \approx 0$ para $i > \tau_s$ y $k \leq 1$, entonces aproxima los primeros k puntos de la respuesta a pulso.

Se realizó una simulación con este método para el sistema de orden 2; la entrada fue PRBS127, la perturbación RUIDO; (que simula ruido blanco gaussiano) con relación señal a ruido S/N = 4.

Con 500 observaciones se estimaron $R_{uy}(i)$, $i = 0, 1, \dots, 126$, con el programa ESTIMA/PESO, puesto que como $\hat{R}_{uu}(0) = 1$ resulta que la suco- sión calculada por ESTIMA/PESO es $\frac{\hat{R}_{uy}(i)}{\hat{R}_{uu}(0)} = \hat{R}_{uy}(i)$.

Con el programa CORRELACION se estimaron las cantidades $R_{uu}(i)$, $i = -30, \dots, 126$ y se aplicó el método para ajustar el modelo (5.142);

$$R_{uy} = R_{uu} h + e'$$

con un orden $k = 30$; el resultado obtenido se compara contra la respuesta real y la respuesta del sistema de segundo orden estimado por mínimos cuadrados con los mismos 500 datos, en la figura (5.72).

En la figura (5.73) se lo compara con la respuesta real y con la estimada por los métodos de correlación.

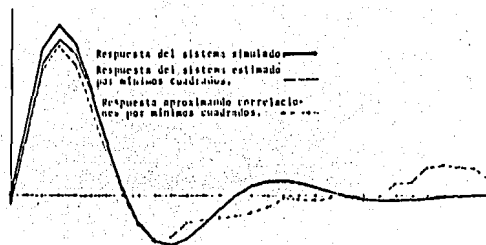


Fig. 5.72

Veamos primero la estimación que resulta de estimar el sistema por mínimos cuadrados. Esta vez se identificó el sistema con un modelo cuyo polinomio $B(q^{-1})$ en $B(q^{-1}) = b_0 + b_1 q^{-1} + \dots + b_m q^{-m}$, por lo que no se impone un retraso que obligue a ser $\hat{h}(0) = 0$.

La figura (5.72) muestra, si la comparamos con (5.71) que este tiene efecto sobre los primeros puntos de $\hat{h}(k)$, en particular para $k \leq 6$, pero la proximidad de $\hat{h}(k)$ para $k > 6$ es tan grande que no se pueden apreciar diferencias en la figura.

En cambio para la estimación que resulta de aplicar este último método, los primeros puntos son los mejores; como en el método de correlación, para $k > 12$ la estimación desmejora notablemente.

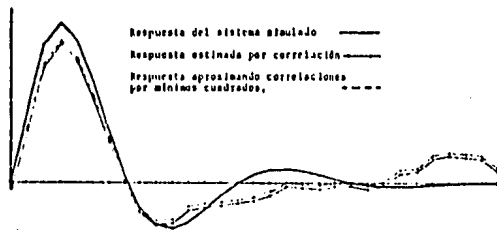


Fig 5.73

Comparando la estimación de $h(k)$ por correlación y con este último método no se aprecian diferencias muy importantes, esto es natural, Obsérvese cómo se obtuvieron ambos estimadores.

llamando \hat{h}_1 al vector $(\hat{h}_1(0), \dots, \hat{h}_1(k))^T$ obtenido por correlación y \hat{h}_2 al mismo obtenido por este método se cumple que:

$$\begin{cases} \hat{R}_{uy} = \hat{h}_1 \\ \hat{h}_2 = (\hat{R}_{uu}^T \hat{R}_{uu})^{-1} \hat{R}_{uu}^T \hat{R}_{uy} \end{cases} \quad (5.193)$$

luego

$$\hat{h}_2 = (\hat{R}_{uu}^T \hat{R}_{uu})^{-1} \hat{R}_{uu}^T \hat{h}_1 \quad (5.194)$$

pero la matriz \hat{R}_{uu} tiene la forma:

$$\begin{bmatrix} 1 & r & r & \dots & r \\ r & 1 & r & \dots & r \\ r & r & 1 & \dots & r \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r & r & r & \dots & 1 \end{bmatrix} = \hat{R}_{uu}$$

ya que la señal de entrada es PRBS127, puede verse en la sección 3.2.4 su señal de auto-correlación.

Entonces $r \ll 1$ (en este caso $r = -0.00787$), de donde $\hat{h}_2 \approx \hat{h}_1$. Obviamente cuanto más "blanca" sea la señal $u(t)$, más se parece R_{uu} a una matriz diagonal, y el método descrito coincide con el de correlación.

La conclusión general que podemos extraer es que, excepto cuando haya necesidad de buscar una aproximación FIR para el sistema, y se trate de una clase restringida de señales de entrada, es una mejor aproximación, más rápida y eficiente, la que resulta de estimar la respuesta a pulso a partir de la función de transferencia obtenida por mínimos cuadrados.

No debe perderse de vista que representar el sistema mediante funciones racionales es una forma eficiente de representar las distintas sucesiones de peso. En este sentido el conjunto de las funciones racionales es denso y tiene el atractivo de ser caracterizado por un número bajo de parámetros.

5.12.11 Mínimos cuadrados repetidos. Determinación del orden.

Determinación del retraso.

En la sección 5.9 se describió el procedimiento de mínimos cuadrados repetidos y la prueba F para la determinación del orden del modelo. Hemos visto que básicamente se trata de probar si la disminución de la función de costo es significativa al aumentar el número de parámetros del modelo.

El procedimiento se ha implantado a través de los programas SKOOT/NPE/S y SKOOT/REPE/C para sistemas sin y con término constante respectivamente.

En este párrafo se describen algunas simulaciones para ejemplificar el uso del procedimiento (y los programas), en el caso de sistemas con perturbaciones correlacionadas, y en la determinación del orden y retraso del modelo "una entrada una salida" del sistema a identificar.

Los dos programas mencionados son iguales en cuanto a los datos y la salida de resultados, la única diferencia es que el último maneja modelos con términos constantes. Así que trabajaremos con uno de ellos solamente: SKOOT/REPE/S, para el otro no hay diferencia.

Los datos del programa son el archivo de señales de entrada (ENT) y de salida (SAL) del sistema a identificar, los valores máximo y mínimo del retraso k , órdenes M y N .

El programa ajusta modelos de la forma:

$$y(t) (1+a_1q^{-1}+\dots+a_nq^{-n}) = q^{-k} [b_0+b_1q^{-1}+\dots+b_mq^{-m}] u(t) + e(t) \quad (5.195)$$

con $u(t)$ ruido blanco de varianza unitaria, para todas las combinaciones posibles de k , m , n que cumplan

$$\begin{cases} K \text{ MIN} \leq k \leq K \text{ MAX} \\ M \text{ MIN} \leq m \leq M \text{ MAX} \\ N \text{ MIN} \leq n \leq N \text{ MAX} \end{cases} \quad (5.196)$$

El resultado se entrega en forma de una serie de tablas como en la figura (5.74), cada una de ellas para un valor K_j de k . En la intersección de la columna m_j con la fila n_l se encuentran los valores estimados de los parámetros al ajustar por mínimos cuadrados un modelo de la forma (5.195) con $K = k_l$, $m = m_j$, $n = n_l$.

TABLA DE RESULTADOS $K = k_j$

	$M = m_{\text{MIN}}$	$M = m_j$	$M = m_{\text{MAX}}$
$N = n_{\text{MIN}}$			
$N = n_l$			
$N = n_{\text{MAX}}$			

Fig 5.74

Al final de cada una de estas tablas se encuentra una tabla de resultados de la prueba F para el orden del sistema. Estas tablas tienen la forma que se muestra en la figura (5.75)

Fig 5.75

Los asteriscos en las intersecciones de las columnas y filas recurrentes indican la estimación correspondiente que se da en la tabla anterior, y los números entre asteriscos corresponden a la cantidad de pruebas t , en la transición de un orden a otro. Por ejemplo en la figura (5.75) la cantidad encerrada en 1 representa el valor de t para medir la reducción de la función de costo al pasar del modelo con $k=k_1, m=n_1, n=n_2$ al modelo con $k=k_1, m=n_1+1, n=n_2$. De acuerdo a la ecuación (5.37) el valor de t en este caso se calcula como:

$$t = \frac{V_1 - V_2}{V_2} \cdot \frac{N - (m+1)n_2}{1}$$

Donde V_1 es la función de costo asociada al primer modelo y V_2 la asociada al segundo.

Ha existido algo de confusión en cuanto a la aplicación de esta prueba; debe destacarse que no sólo importa el número de parámetros sino también cuales son estos.

El segundo modelo debe contener todos los parámetros del primero, es decir su vector de parámetros debe contener el vector de parámetros del primer modelo. En el esquema de la figura (5.75) esto quiere decir que sólo se pueden probar transiciones hacia abajo o la derecha. ¿Cómo determinar el orden y el retraso?

La tabla nos da el resultado de aplicar la prueba al ir aumentando de uno en uno el número de parámetros. Supongamos conocido el retraso k y observamos la tabla correspondiente a dicho retraso. Asimismo que el modelo del sistema que genera los datos tiene $n = n_0$, $m = n_0$.

Entonces dadas las identificaciones correspondientes a $m \geq m_0$ y $n \geq n_0$ se encuentran bajo la hipótesis nula que queremos verificar a través de esta prueba. Sea t_0 el punto crítico elegido, entonces los bloques de la figura (5.75) que estén a la derecha o en la misma columna, y debajo o en la misma fila que el asterisco correspondiente a m_0, n_0 deben contener cantidades $t \leq t_0$ para el nivel de riesgo elegido.

En la figura (5.76) esta región se ha sombreado.

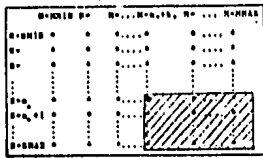
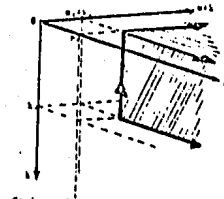


Fig. 5.76

Se puede determinar también el retraso K al mismo tiempo que m y n observando que disminuir el retraso K en una unidad equivale a agregar un coeficiente más al polinomio $H(q^{-1})$, con una potencia de q^{-1} un grado menor.

Entonces podríamos extender el esquema de la tabla para la prueba T a 3 dimensiones, cuyas coordenadas son n, m, k , y si los valores de n, m y k del sistema que genera los datos fueran n_0, m_0, k_0 respectivamente, entonces la región $(n \geq n_0, m \geq m_0, k \leq k_0)$ debería tener (con un nivel de riesgo prefijado) cantidades t de transición, no nulas (o t_0). Esto es una consecuencia inmediata del hecho de que los modelos contenidos en dicha región están bajo la hipótesis nula que se quiere verificar.

En la figura (5.77) se dibuja dicha región.



Conjunto de células que se encuentran bajo la hipótesis nula H_0 y direcciones de bloques de la matriz de mayor orden.

Fig. 5.77

Las tablas de la figura (5.75), que son resultado de los programas, son secciones $K = cte.$ del espacio dibujado en la figura (5.77), denotadas k unidades sobre el eje m .

Ejemplos de determinación de m, n, k

Distribución gaussiana

Mostraremos ahora algunos ejemplos de la determinación de m, n y k para el sistema de segundo orden (que sigue la ley (5.146)).

El sistema fue excitado con entrada PRBS127 y perturbación EUID01, con relación señal a ruido = 1 (es entonces la salida SALA10+).

Se realizó el procedimiento de mínimos cuadrados repetidos para

$$\begin{cases} 0 \leq K \leq 2 \\ 0 \leq M \leq 4 \\ 0 \leq N \leq 4 \end{cases}$$

*****MELANGE TIME*****

*****DE BUNAMI-REPARA 01***

Table with 4 columns: No 0, No 1, No 2, No 3. Rows contain numerical data for various categories, including 'DE BUNAMI-REPARA 01' and 'DE BUNAMI-REPARA 02'.

Table with 4 columns: No 0, No 1, No 2, No 3. Contains numerical data for categories 'DE BUNAMI-REPARA 01', 'DE BUNAMI-REPARA 02', and 'DE BUNAMI-REPARA 03'.

*****MELANGE TIME*****

*****DE BUNAMI-REPARA 01***

Table with 4 columns: No 0, No 1, No 2, No 3. Rows contain numerical data for various categories, including 'DE BUNAMI-REPARA 01' and 'DE BUNAMI-REPARA 02'.

*****DE BUNAMI-REPARA 03***

Table with 4 columns: No 0, No 1, No 2, No 3. Contains numerical data for categories 'DE BUNAMI-REPARA 01', 'DE BUNAMI-REPARA 02', and 'DE BUNAMI-REPARA 03'. A box highlights specific data points in the bottom right corner.

*****MELANGE TIME*****

La prueba determina entonces $k = 1$, $M = 1$, $N = 2$ con mucha certidumbre, y que son además los valores reales de k , M , N .

Hay muchos otros factores que pueden ayudar en la decisión, algunos que se desprenden del listado de las páginas anteriores, y otros que no, vamos primero lo que se puede extraer del propio listado:

Hemos tratado de ajustar un modelo

$$f(t) \{ (1 + a_1 q^{-1} + \dots + a_{N_{MAX}} q^{-N_{MAX}}) + (b_0 + b_1 q^{-1} + \dots + b_{M_{MAX} + K_{MAX}} q^{-(M_{MAX} + K_{MAX})}) \} u(t) + o(t) \quad \dots (5.195)$$

a un sistema de la forma

$$y(t) \{ (1 + a_1 q^{-1} + \dots + a_{n_0} q^{-n_0}) = q^{-k_0} [b_0 + b_1 q^{-1} + \dots + b_{m_0} q^{-m_0}] \} u(t) + o(t) \quad \dots (5.196)$$

al variar M, N, k en la identificación lo que en realidad se hace es imponer condiciones del tipo

$$\begin{aligned} a_j &= 0 \\ \beta_j &= 0 \end{aligned} \quad (5.197)$$

para algunas i y algunas j ; una forma de verificar la identificación de los valores de m_0 , n_0 y k_0 es que al levantar alguna de dichas restricciones el valor estimado del a_{i_0} o β_{j_0} sea cero o sea próxima a cero.

Si se observan las tablas de estimaciones de las páginas anteriores esto puede corroborarse en forma contundente: excepto para $k = 0$ donde el término b_0 de transmisión directa agregado dista del cero aproximadamente 1.64 veces su desv. estándar estimada, en el resto de los casos los coeficientes que se agregan distan del cero menos que una desv. estándar estimada.

Esto sucede así, no sólo al levantar una restricción, sino aún cuando levantamos varias restricciones al mismo tiempo. Debe decirse sin embargo, que en esto hubo suerte, pues cuando se levantan varias restricciones del tipo (5.197) al mismo tiempo, en lugar de tener coeficientes próximos a cero puede darse lugar a un fenómeno de cancelación de polos y ceros.

Esta es otra forma de ayudar a la determinación del orden especialmente para otros procedimientos de identificación distintos del de mínimos cuadrados, en que se quiera determinar la dinámica de las perturbaciones cuando éstas están correlacionadas. Puede estudiarse el empleo de este método en (52) que constituye una excelente recopilación y descripción de métodos de determinación del orden del modelo.

Otros factores que pueden tenerse en cuenta son por ejemplo: la estimación de la respuesta a pulso por correlación, y la verificación de las hipótesis del modelo (5.2).

Con la respuesta a pulso estimada por correlación u otro método, aun que no hay una gran precisión podemos tener una idea de la magnitud del retraso K y quizá del orden N .

Por ejemplo, en nuestro caso la identificación se llevó a cabo con los mismos datos con que se estimó por correlación la respuesta a pulso de la figura (5.71), de ella se puede apreciar en primer lugar que $k = 1$, y que $n \geq 2$ puesto que no se puede ajustar una exponencial.

Con verificar las hipótesis del modelo (5.2) queremos decir que sobre los residuos de la identificación debemos verificar independencia (o al menos no correlación), no correlación con la entrada, no correlación con la salida para retardos negativos, etc.

Por ejemplo, en el caso estudiado, si hubiera cabido alguna duda (que no hubo) acerca de si el modelo obtenido para $K = 1$, $M = 1$, $N = 2$ era mejor que el obtenido para $k = 2$, $M = 0$, $N = 2$, bastaba con estimar la función de correlación $R_{eu}(k)$ con los residuos de ambos modelos.

En las figuras (5.78) y (5.79) se muestran las estimaciones normalizadas, es decir $\hat{r}_{eu}(\tau) = \hat{R}_{eu}(\tau) / \sqrt{\hat{R}_{uu}(0) \hat{R}_{ee}(0)}$ para $-10 \leq \tau \leq 10$. En el modelo para $k = 2$ se viola una hipótesis fundamental, el residuo está correlacionado con las entradas y lo anteceden en un instante de muestreo. La razón es lógica, dicho modelo no contiene en el numerador términos en q^{-1} .

La comparación de las figuras (5.78) y (5.79) hubiera reportado un elemento decisivo, para llegar a la conclusión que el mejor modelo era el que se había obtenido con $K = 1$.



Se creyó conveniente en algún momento incorporar una sección que discutiera métodos de determinación del orden del modelo. A medida que se iba trabajando y simulando con cada uno de ellos más crecía la convicción de que no alcanzaría ni siquiera un capítulo, por lo que han sido introducidos a través de este ejemplo, en torno fundamentalmente a uno de ellos: la prueba F.

Otros métodos (además de los enunciados) tales como el de "convertimiento del determinante", graficación de las señales, y los métodos ya vistos, pueden ser estudiados en la referencia (52) y en H. Ubenhaan "Application of different statistical tests...", Proceedings 3rd, IFAC Symposium, The Hague/Delft, Netherlands 1973. En ambos se describen simulaciones para comparar las pruebas.

Todos los autores coinciden en que es de vital importancia usar varias de estas pruebas en forma complementaria, cuando se trata con datos no simulados; las conclusiones que se extraen de cada una dan lugar a resultados cualitativamente superiores cuando se les combina.

En el ejemplo que hemos visto, la prueba r da el modelo correcto, no sólo el orden sino además K y M en forma precisa. No nos olvidemos que trabajamos en condiciones ideales, con datos simulados, con señales sinusales, etc.

La perturbación utilizada fue RUID01, que simula una señal de ruido blanco de distribución gaussiana.

¿Qué sucede con otras distribuciones? Puede verse que si $n \rightarrow \infty$, la prueba sigue siendo consistente.

Vemos dos ejemplos: para el mismo sistema de segundo orden, con la misma señal de entrada FESB127, y la misma relación señal a ruido, se repite el mismo procedimiento para la perturbación RUID07 que simula una distribución uniforme) y EBT1 (que simula una distribución binaria); el número de datos sigue siendo 500.

Distribución uniforme

En las siguientes figuras se presentan los listados de la prueba r para $k = 0, 1, 2$ en la aplicación del procedimiento de mínimos cuadrados repetidos sobre la realización del sistema de segundo orden perturbada con la señal RUID07.

Se listan sólo las tablas de la prueba r pues en este ejemplo nos interesa sólo determinar el orden del modelo, por lo que se suprimieron los listados de los estimados de los parámetros para distintos órdenes.

Se observa el mismo tipo de resultados que en el caso anterior, según estos listados para $k = 0$ es $M=2$, $N=2$; para $k=1$ es $M=1$, $N=2$; para $k=2$ es $M=0$, $N=2$. Las tres tablas coinciden en que $M=2$, $N=2$, falta determinar k .

Si probamos la hipótesis $k=2$ contra $k>2$ calculando el valor de t asociado a la transición $k=2$, $M=0$, $N=2 \rightarrow k=1$, $M=1$, $N=2$ resulta $t = 278 \gg 3.9$, no pasa la prueba. Es que la función de costo se reduce mucho, de aproximadamente 1535 a 1059.3

En cambio, si probamos la hipótesis $k=1$ contra $k>1$ calculando el valor de t asociado a la transición $k=1$, $M=1$, $N=2 \rightarrow k=0$, $M=2$, $N=2$, éste es: $t = 0.176$. La función de costo se redujo poco; aproximadamente de 1050.3 a 1049.9.

La prueba vuelve a dar en forma inequívoca $k=1$, $M=1$, $N=2$ que son los valores reales.

Retraso K:0

	M=0	M=1	M=2	M=3	M=4
M=0	11.7107	60.5030	104.8166	121.5111	
	2111.4314	2411.7548	2804.9661	2527.3545	1907.0540
M=1	71.3185	137.9078	57.7500	13.6052	
	572.5390	804.4673	660.4207	550.2210	524.7032
M=2	194.6487	72.2047	2.0633	2.3744	
	4.5362	6.7035	2.9025	1.9021	1.4009
M=3	196.9776	67.6175	1.0630	1.8037	
	0.8146	0.3500	1.7035	1.1926	0.0345
M=4	195.5408	68.9635	1.2661	0.7225	

Retraso K:1

	M=0	M=1	M=2	M=3	M=4
M=0	60.7375	107.0073	121.7261	92.5972	
	2416.5249	2012.6403	2532.5479	1992.0175	1595.9609
M=1	138.3230	57.8619	13.6907	1.2004	
	805.9567	669.3610	550.0921	525.2741	524.1971
M=2	72.3401	2.0205	2.3338	1.7141	
	6.0307	2.7830	1.8180	1.3271	1.2683
M=3	67.4870	1.0553	1.8443	1.6573	
	0.3403	1.6914	1.1241	0.0264	0.0620
M=4	68.7402	1.2017	0.7435	1.6097	

Retraso K:2

	M=0	M=1	M=2	M=3	M=4
M=0	101.7191	115.1176	87.2064	37.7701	
	2350.4349	2079.1955	1670.1470	1304.1706	1210.0165
M=1	45.0127	10.0660	0.3027	11.0019	
	454.0737	376.7197	359.2097	357.9440	336.0037
M=2	0.9039	0.8009	0.4162	0.1424	
	0.7488	0.7212	0.4701	0.4238	0.3979
M=3	0.0759	0.5511	0.3710	0.1180	
	1.6122	0.9462	0.7750	0.9906	0.9641
M=4	0.7107	0.3009	0.5000	0.0954	

Comparamos las tablas de la prueba F para esta realización con las que se obtienen para la realización con perturbación gaussiana.

En las últimas tablas los valores de t en las regiones que pasan la prueba son algo mayores que en las anteriores, aunque ninguno de ellos llega a alcanzar 3, el mayor es 2.9025.

En las tablas de la realización con perturbación gaussiana el valor máximo de t en estas regiones es 1.511.

En estos casos el valor crítico $t_0 = 3.92$ está sobrado por encima, es decir que hubiéramos obtenido el mismo resultado aún con valores de t_0 mayores. En las últimas tablas se hubiera necesitado $t_0 > 6.8$ para modificar la forma de las regiones, en las tablas anteriores se hubiera necesitado $t_0 > 32$ para provocar dichos cambios.

Las regiones que pasan y no pasan la prueba están pues, bien diferenciadas.

Distribución binaria

Las siguientes tablas resumen los resultados de la prueba F para el mismo sistema en las mismas condiciones, pero perturbado con la señal ENI que simula una sucesión binaria aleatoria.

Igual que en los casos anteriores, las tablas indican como probables valores de M y N :

Sistema H=0					
N=0	N=1	N=2	N=3	N=4	
N=0	0 11.0563 0	48.5102 0	76.5629 0	66.7730 0	
	1775.6806	1918.0150	2162.9244	1949.7557	1471.4034
N=1	0 39.4397 0	88.3102 0	32.4991 0	4.7301 0	
	610.0192	972.8948	816.3421	734.6402	725.1451
N=2	0 222.4508 0	26.7404 0	0.9700 0	2.3341 0	
	2.0190	0.3400	1.3330	0.3584	0.6006
N=3	0 223.1975 0	23.2994 0	0.0050 0	2.5731 0	
	0.1101	0.4434	1.1417	1.1930	0.4455
N=4	0 224.7711 0	23.9344 0	0.0702 0	1.8411 0	

Sistema H=1					
N=0	N=1	N=2	N=3	N=4	
N=0	0 48.6163 0	76.6594 0	66.8129 0	34.7886 0	
	1985.3858	2141.9344	1950.6481	1473.2117	1550.5990
N=1	0 89.0046 0	32.9159 0	4.9133 0	4.5815 0	
	977.3212	814.9409	736.3950	727.1789	710.6360
N=2	0 26.8505 0	1.0403 0	2.3057 0	1.1116 0	
	4.7476	0.3592	0.3166	0.3423	0.5603
N=3	0 23.2214 0	0.0049 0	2.4008 0	1.1293 0	
	0.4934	0.2251	1.2741	0.4768	0.4424
N=4	0 23.0914 0	0.0702 0	1.8004 0	1.1128 0	

Sistema H=2					
N=0	N=1	N=2	N=3	N=4	
N=0	0 72.5751 0	62.7786 0	32.4633 0	9.8940 0	
	1892.5489	1687.6737	1444.3105	1342.1714	1348.7050
N=1	0 25.4647 0	2.7837 0	3.8049 0	18.4082 0	
	573.3375	471.7890	466.8470	457.7811	424.4700
N=2	0 0.1477 0	1.2298 0	2.0714 0	1.7132 0	
	0.3582	0.4129	0.6779	0.7349	0.5498
N=3	0 0.0034 0	1.4944 0	2.1270 0	1.5273 0	
	0.0478	0.0559	0.0107	0.0274	0.0256
N=4	0 0.0114 0	1.4441 0	2.1391 0	1.5273 0	

$$\left\{ \begin{array}{l} \text{para } k=0 : H=2, N=2 \\ \text{para } k=1 : H=1, N=2 \\ \text{para } k=2 : H=0, N=2 \end{array} \right.$$

Las tres tablas coinciden en que $H=3, H+k=2$. Probámonos ir disminuyendo el retraso manteniendo constante $H+k$:

Al pasar de $k=2, H=0, N=2$ a $k=1, H=1, N=2$ la función de costo pasa de 1553.4623 a 1051.6528, luego $t = 235.72$ es $\gg 3.92$, no pasa la prueba.

Al pasar de $k=1, H=1, N=2$ a $k=0, H=2, N=2$ la función de costo pasa de 1051.6528 a 1050.2798, luego $t = 0.6445 < 3.92$, pasa la prueba.

Otra vez la prueba Γ señala precisamente los valores reales de K, M, N ; $k=1, H=1, N=2$

Aplicación a sistemas con perturbaciones correlacionadas

Lo que hemos hecho hasta ahora con mínimos cuadrados repetidos es aplicar la idea de la sección 5.9 cuando la autoregresión de las perturbaciones

$$n(t) = \frac{1}{H(q^{-1})} e(t) \quad (e(t) \text{ ruido blanco})$$

es tal que $H(q^{-1}) = 1$, en ese caso podemos obtener directamente el orden (y retraso) de la parte determinística del sistema, además de poder probar con la prueba Γ la hipótesis de que algunos coeficientes sean nulos.

Veamos algunos ejemplos de identificación de sistema con perturbaciones correlacionadas; es decir, dado el sistema:

$$A(q^{-1})y(t) = B(q^{-1})u(t) + n(t) \quad (5.198)$$

trataríamos de ajustar un modelo de la forma (5.136):

$$A(q^{-1})H(q^{-1})y(t) = B(q^{-1})H(q^{-1})u(t) + e(t) \quad (5.199)$$

aproximando la perturbación como una autorregresión:

$$n(t) = \frac{1}{H(q^{-1})} e(t)$$

Como se explicó en la sección 5.9, puede obtenerse así un buen predictor en el sentido de mínimos cuadrados para los errores de predicción, pero se estima AH y BH (no se estima A y B).

De la estimación de los polinomios AH y BH podría estimarse la dinámica de la parte determinística del proceso (los polinomios A y B), y la de la parte estocástica (polinomio H). Esto puede hacerse por ejemplo buscando ceros comunes de AH y BH, pero en general esto es difícil por las razones explicadas en la sección 5.9.

Veamos dos ejemplos con sistemas de primer orden.

1) Consideremos el ejemplo utilizado para mostrar en el párrafo

5.12.8 el sesgo del estimador de mínimos cuadrados en el caso que las perturbaciones fueran correlacionadas. El sistema era:

$$(1-0.8q^{-1})y(t) = (q^{-1})u(t) + \lambda(1-0.7q^{-1})e(t) \quad (5.200)$$

Este sistema fue identificado con el método de mínimos cuadrados en el párrafo 5.12.8 habiéndose arribado a una estimación sesgada del parámetro $a_1 = -0.8$; $\hat{a}_1 = -0.4595$.

El valor de $\lambda = 1.64399$ fue elegido para que la relación señal a ruido S/N sea 1.

Al ver el resultado de la estimación por mínimos cuadrados se aprecia que para 1000 muestras se obtiene un modelo con sesgo en los parámetros y una desv. std. de los residuos de 1.8615. La desv. std. de la perturbación es 1.66667. ¿No puede entonces obtenerse un mejor resultado, al menos como predictor?

La respuesta es sí.

Ahora se aplicará el procedimiento de mínimos cuadrados repetidos al mismo sistema, simulando de la misma manera. En lugar de utilizar 1000 observaciones utilizaremos la mitad: 500.

El listado que sigue es el resultado del programa SREGOT/REPE/PL aplicado a las señales de entrada y salida del sistema. Se han suprimido las filas y columnas correspondientes a $M, N = 0, 1$ por razones de espacio, ya que no agregan nada nuevo. Se asumió $K=1$ como conocido.

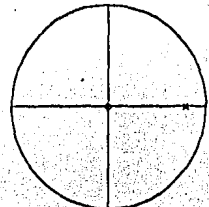
Para este valor de k la prueba F señala como modelo adecuado el que corresponde a $M=4$, $N=5$ puesto que la reducción de la función de costo que resulta de agregarle parámetros no es significativa.

CONTINUA DE LA TABLA N.º 1000 DEL ANEXO I

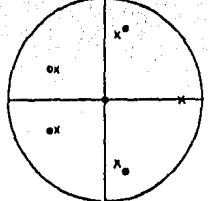
Polos	Ceros
0.13311 + j0.63484	0.20618 + j0.70393
0.13311 - j0.63484	0.20618 - j0.70393
0.50001 + j0.29502	0.57326 + j0.3033
0.50001 - j0.29502	0.57326 - j0.3033

Polos Ceros

0.13311 + j0.63484	0.20618 + j0.70393
0.13311 - j0.63484	0.20618 - j0.70393
0.50001 + j0.29502	0.57326 + j0.3033
0.50001 - j0.29502	0.57326 - j0.3033



PATRON DE POLOS Y CEROS DEL SISTEMA.



PATRON DE POLOS Y CEROS DEL SISTEMA ESTIMADO.

Fig 5.80

El resultado de la estimación para dichos valores de K, H, N está recuadrado en la tabla superior, a primera vista se observa que hay muchos parámetros y no se encuentra similitud alguna con los valores de los parámetros del modelo. Sin embargo lo que no puede ser visto en una forma de representar el sistema resulta clarísimo en otra.

Los polos y ceros de la función de transferencia estimada con $\hat{A}H$ y $\hat{B}H$ se tabulan a continuación del listado de mínimos cuadrados revisados. A partir de ellos se dibujan los patrones de polos y ceros del sistema simulado y del sistema estimado con mínimos cuadrados revisados. El resultado sorprendería más si no supiéramos lo que buscábamos.

Se observa el comportamiento que se esperaba: al aproximar las perturbaciones con una autorregresión, el método agrupa polos y ceros muy próximos (que asintóticamente, para $N \rightarrow \infty$ se cancelan), que son resultado de la factorización del proceso autorregresivo de las perturbaciones.

Poseyendo otros elementos de juicio, o. ej. una aproximación a la respuesta a pulso, para confirmar que se trata de un sistema de primer orden, se puede deducir una estimación de la parte determinística y de la parte estocástica del sistema.

Tomando en cuenta que lo que se tiene son estimaciones aproximadas $\hat{A}H$ y $\hat{B}H$, una posibilidad (que no hemos implantado) es encontrar no lineamientos $\hat{\lambda}, \hat{\beta}, \hat{\eta}$ tales que $\hat{\lambda}H$ y $\hat{\beta}H$ aproximen $\hat{A}H$ y $\hat{B}H$ respectivamente en el sentido de mínimos cuadrados de las diferencias de sus coeficientes.

Este enfoque permitiría incluir las estimaciones de las desviaciones standard de los coeficientes de $\hat{A}H$ y $\hat{B}H$ como los pesos de una suma de cuadrados ponderada.

Más allá de cómo encontrar la mejor aproximación en el caso general está claro que en este caso la aproximación es buena puesto que el denominador $\hat{A}H$ puede factorizarse como: $(1 - 0.7889q^{-1})(1 + 0.7338q^{-1} + 0.4916q^{-2} + 0.331q^{-3} + 0.1418q^{-4})$ el numerador $\hat{B}H$ puede factorizarse como: $0.9619q^{-1}(1 + 0.7342q^{-1} + 0.4659q^{-2} + 0.4434q^{-3} + 0.2263q^{-4})$ donde el polo libre de cancelación es $z = 0.78894$, próximo al real 0.8, y el promedio móvil del ruido es:

$$1 - 0.7q^{-1} = 1/1 + 0.7q^{-1} + 0.49q^{-2} + 0.343q^{-3} + 0.2401q^{-4} + 0.1681q^{-5} + \dots \text{ (com-} \\ \text{párese con las factorizaciones).}$$

- 2) En el ejemplo anterior fue necesaria una aproximación autorregresiva para el proceso del ruido con cuatro coeficientes. En realidad es muy dependiente del proceso del ruido, la cantidad de coeficientes autorregresivos necesarios para llegar a una buena aproximación. Consideremos un ejemplo

Para el sistema (5.202):

$$(1+0.8q^{-1})y(t) = (q^{-1})u(t) + \lambda(1+0.3q^{-1}+0.075q^{-2})e(t) \quad (5.202)$$

condo $e(t)$ es ruido blanco de media cero y varianza unitaria.

Para el proceso del ruido (que es de promedios móviles) en forma de proceso auto-regresivo:

$$(1+0.3q^{-1}+0.075q^{-2})e(t) = \frac{1}{1-0.3q^{-1}+0.015q^{-2}+0.018q^{-3}+0.006q^{-4}+\dots} e(t) \quad \dots (5.203)$$

Los dos primeros coeficientes contienen el 99.95% de la suma de cuadrados de los coeficientes de la expansión. Se puede pensar entonces un aproximar el proceso de promedios móviles como:

$$(1+0.3q^{-1}+0.075q^{-2})e(t) \approx \frac{1}{1-0.3q^{-1}} e(t) \quad (5.204)$$

Una forma de ver qué tan buena es la aproximación es considerar la salida de ambos procesos, como en la figura (5.81). Se puede ver que la varianza de $n_1(t)$ es $\text{var}(n_1(t)) = 1.09503$, y la salida de la comparación de estos procesos $n_2(t)$ tiene una varianza $\text{var}(n_2(t)) = 0.10161$. Esto es que el valor esperado de la potencia de la diferencia de los dos procesos es 1064 veces menor que la potencia del proceso original $n_1(t)$.

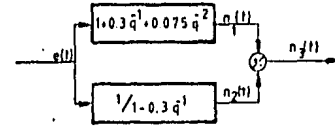


Fig. 5.81

Este sistema fue simulado excitándolo con entrada PRBS127, y perturbación RUIDO1, a partir de condiciones iniciales nulas. Se simularon 645 instantes de muestreo. Se descartaron los primeros 127 datos.

La relación señal a ruido simulada fue $S/N = 1$ al fijar $\lambda = 0.765251$ para tal efecto. A los primeros 500 datos se les aplicó el procedimiento de mínimos cuadrados repetidos, y los resultados se listan a continuación.

La prueba F señala para $k = 1$ un modelo de orden $N = 2$, con 1er orden en el numerador ($M = 1$).

En el listado superior se recuadró la estimación de los parámetros por mínimos cuadrados para $M = 1$, $N = 2$, $K = 1$. Las raíces del polinomio $\hat{A}(z)$ son: 0.32972 y 0.75458; la del polinomio $\hat{B}(z)$ es: 0.31634.

En la figura (5.82) se ha dibujado el patrón de ceros y polos del sistema original y el del sistema identificado de esta forma. El sistema estimado puede ponerse como:

Se puede inferir así una estimación gruesa del sistema como

$$y(t) (1 - 0.7545q^{-1}) = (1.0002q^{-1})u(t) + \frac{1}{1 - 0.37972q^{-1}} e(t)$$

que es bastante aproximada. Nótese que en la figura (5.82) se muestra un polo y un cero casi cancelándose cerca de donde se había previsto estaría el polo de la aproximación autorregresiva del ruido.

En este caso, por la forma del proceso del ruido, aunque los pronósticos móviles que lo generan tienen más términos, se logró una aproximación autorregresiva de menor orden que en el caso anterior.

En los ejemplos vistos, se aproximó el proceso de las perturbaciones $n(t)$ con un proceso autorregresivo:

$$n(t) = \frac{1}{\hat{H}(q^{-1})} e(t) \quad (5.205)$$

donde $e(t)$ es ruido blanco con media cero.

Es fácil de demostrar que bajo las condiciones generales de convergencia del método de mínimos cuadrados, si (5.205) es exacta, entonces los estimadores que hemos llamado $\hat{A}(q^{-1})$, $\hat{B}(q^{-1})$ convergerán realmente a $A(q^{-1})$, $B(q^{-1})$ y $\hat{H}(q^{-1})$ respectivamente.

Si en cambio (5.205) es simplemente una aproximación, puede demostrarse que para valores grandes de N los estimadores $\hat{A}(q^{-1})$ y $\hat{B}(q^{-1})$ convergen para $N \rightarrow \infty$ a una proximidad de $A(q^{-1})$ y $B(q^{-1})$ respectivamente, donde $\tilde{H}(q^{-1})$ es el polinomio que minimiza:

$$\min_N E \{ \hat{H}(q^{-1}) n(t) \}^2$$

Elencamos

$$n(t) = C(q^{-1}) e(t)$$

$$A(q^{-1}) y(t) = B(q^{-1}) u(t) + n(t) \quad (6.2)$$

Si $n(t)$ fuera una secuencia no correlacionada estaríamos frente al problema de mínimos cuadrados. En general si C no es una constante, $n(t)$ es una secuencia de variables aleatorias con matriz de covarianzas no diagonal.

6. MINIMOS CUADRADOS GENERALIZADOS Y ESTIMADOS DE MARKOV

6.1 Introducción

Trataremos de atacar con dos procedimientos el problema de identificación cuando las perturbaciones del sistema están correlacionadas. Hemos visto que la solución de mínimos cuadrados es asintóticamente óptima, y ese hecho está vinculado a la correlación de las perturbaciones.

Consideremos el caso general de un sistema discreto con perturbaciones aleatorias de espectro racional visto en el Capítulo 4.

$$A(q^{-1}) y(t) = B(q^{-1}) u(t) + C(q^{-1}) e(t) \quad (6.1)$$

con $e(t)$ no correlacionado.

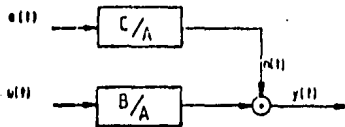


Fig 6.1

6.2 Estimación de Markov

Retomemos el modelo lineal del Capítulo 2:

$$Y = X\theta + n \quad (6.3)$$

En el teorema 2.21 fue probado que el estimador lineal insesgado de varianza mínima para θ es el estimador de Markov:

$$\hat{\theta}_M = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \quad (6.4)$$

que minimiza la función de costo:

$$J = \hat{n}^T V^{-1} \hat{n} \quad (6.5)$$

con

$$\begin{cases} \hat{n} = Y - X\hat{\theta} \\ V = E[\hat{n}\hat{n}^T] \end{cases} \quad (6.6)$$

($\hat{\theta}$ es un estimador cualquiera), (V = matriz de covarianzas de \hat{n}).

esta estimación es mejor que la de mínimos cuadrados cuando el vector

de errores n tiene elementos no independientes.

En el teorema 2.21 fue supuesto X y n independientes; como V es simétrica y positiva definida, admite una descomposición:

$$V^{-1} = R^T R \quad \text{ó} \quad V = R^{-1} (R^{-1})^T \quad (6.7)$$

donde R es una matriz triangular inferior; luego hacemos el mismo razonamiento que en el teorema 2.21:

$$\hat{\theta}_M = (X^T R^T N X)^{-1} X^T R^T N Y \quad (6.8)$$

El estimador de Murov es el mismo que obtendríamos de aplicar mínimos cuadrados al sistema (6.3) filtrado por la matriz N_f en efecto: pre-multiplicando (6.3) por N_f

$$N_f Y = N_f X \theta + N_f n \quad (6.9)$$

Como en este modelo los errores no son correlacionados pues:

$$E\{(N_f n)(N_f n)^T\} = E\{N_f n n^T N_f^T\} = N_f E\{n n^T\} N_f^T = N_f V N_f^T = I \quad (\text{por (6.7)})$$

entonces la estimación de Murov es la mejor en el sentido de varianzas mínimas.

N_f es entonces una matriz "blanqueadora" de la sucesión de elementos del vector n .

Pongamos como en el Capítulo 5, nuestro sistema dinámico (6.2) en la forma (6.3), con Y el vector de salidas, n el de perturbaciones, X matriz de entradas y salidas, y θ el vector de parámetros de la parte determinística del proceso. Es decir:

$$\left\{ \begin{array}{l} Y = (y(1), y(2), \dots, y(n))^T \\ X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix}; \text{ con } x_i = \begin{bmatrix} -y(i-1) \\ \vdots \\ -y(i-n) \\ u(i) \\ \vdots \\ u(i-m) \end{bmatrix} \\ \theta = (a_1, \dots, a_n, b_1, \dots, b_m)^T \\ n = (n(1), n(2), \dots, n(n))^T \end{array} \right.$$

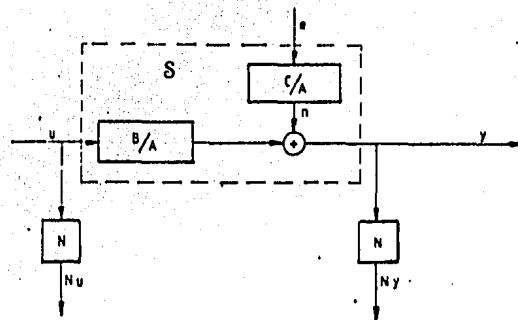


Fig 6.2

Se puede pensar en la matriz M como un filtro de blanquea las perturbaciones. La gran ventaja es que como ahora x y u no son independientes (que era la hipótesis del Capítulo 2), el aplicar mínimos cuadrados sobre los datos filtrados Mx y $M'y$ (a través de $M'Y$ y $M'X$) conduce a una estimación insesgada de θ en virtud de los resultados del Capítulo 5.

Obsérvese que la forma triangular inferior de M permite hacer el filtrado Mx , $M'y$ como una suma ponderada de las señales presentes y pasadas, es realizable. La estimación que resulta es asintóticamente insesgada (y eficiente para perturbaciones gaussianas), pero su limitación es obvia: se requiere conocer previamente M , es decir la correlación de las perturbaciones, ó el polinomio $c(q^{-1})$ en el modelo (6.1).

6.3 Mínimos Cuadrados Generalizados

6.3.1 El método

Para salvar esta dificultad, Clarke (1967), [15] propone el método de mínimos cuadrados generalizados (en la literatura: GLS), que consiste esencialmente en aplicar en forma iterativa el procedimiento de mínimos cuadrados para identificar por separado la dinámica de interés del sistema, y el espectro de las perturbaciones.

Volvamos al modelo (6.1). Fijado $c(q^{-1})$ podemos aproximar tanto como se quiera la densidad espectral del proceso $n(t)$ con un proceso

de la forma $n(t) = \frac{1}{\Gamma(q^{-1})} u(t)$. Basta para ello considerar polinomios Γ de grado r lo suficientemente grande.

De hecho se está aproximando el proceso con un modelo que contiene solo polos, es decir un proceso autorregresivo. Y se hace la aproximación

$$\Gamma(q^{-1}) \cong \frac{1}{c(q^{-1})} \quad (6.10)$$

$$\Gamma(q^{-1}) = 1 + F_1 q^{-1} + \dots + F_r q^{-r}$$

El modelo aproximado es:

$$A(q^{-1})y = B(q^{-1})u + \overbrace{\frac{n(t)}{c(q^{-1})}} \quad (6.11)$$

Puesto de otra forma

$$A(q^{-1})y^h(t) = B(q^{-1})u^h(t) + e(t) \quad (6.12)$$

donde

$$\begin{cases} u^h(t) = \Gamma(q^{-1}) u(t) \\ y^h(t) = \Gamma(q^{-1}) y(t) \end{cases}$$

Conocer $\Gamma(q^{-1})$ equivale a resolver el problema de la estimación de Yarovkov. Compárese la ecuación (6.12) con (5.136) de mínimos cuadrados repetidos. En aquel método se identificaban M' y M' , pero no se podía distinguir la dinámica de S_1 (ver fig (6.3)), del proceso de las perturbaciones.

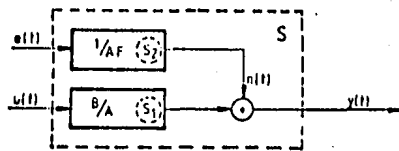


Fig 6.3

La idea básica del método es ir construyendo $\Gamma(q^{-1})$ en forma iterativa y poner el problema de identificar A y B en términos de resolver por mínimos cuadrados la identificación de (6.12).

Para ello se estiman A y B por mínimos cuadrados en (6.11). La estimación será sesgada en general, pero es una primera aproximación. Luego se estima la perturbación $n(t)$ como:

$$\hat{n}(t) = \hat{A} y(t) - \hat{B} u(t)$$

con esta se ajusta un proceso autorregresivo a $n(t)$:

$$\hat{n}(t) = \frac{1}{\Gamma(q^{-1})} \hat{o}(t)$$

Obteniéndose $\hat{\Gamma}(q^{-1})$, con que se filtran las entradas y salidas:

$$\begin{cases} u^*(t) = \hat{\Gamma}(q^{-1}) u(t) \\ y^*(t) = \hat{\Gamma}(q^{-1}) y(t) \end{cases}$$

Con los datos filtrados se hace una nueva estimación de A y B y se repite el proceso hasta encontrar convergencia (Es una aproximación cuasi-lineal a un problema no lineal).

Se puede realizar este procedimiento en dos formas: una en que Γ se aproxima por una sucesión de filtros, y la otra en que se va aproximando Γ de orden fijo. Veamos las dos.

Para una discusión más amplia de los dos algoritmos pueden consultarse las referencias (8), (15), (26), (47), (50) y (51).

6.3.2 Algoritmo 1:

En la iteración i :

- 1) Filtrar los datos:

$$\begin{cases} u_i^*(t) = \hat{\Gamma}_{i-1}^{\wedge} u_{i-1}^*(t) \\ y_i^*(t) = \hat{\Gamma}_{i-1}^{\wedge} y_{i-1}^*(t) \end{cases}$$
- 2) Identificar A y B del modelo:

$$A y_i^* = B u_i^* + e_i \Rightarrow \hat{A}_i, \hat{B}_i$$
- 3) Estimar los residuos:

$$\hat{e}_i = \hat{A}_i y_i^* - \hat{B}_i u_i^*$$
- 4) Estimar una autorregresión para los residuos:

$$\hat{e}_i = \frac{1}{\Gamma_i(q^{-1})} e \Rightarrow \hat{\Gamma}_i$$

5) Si converge parar. El resultado es $\begin{cases} \hat{A} = \hat{A}_i \\ \hat{B} = \hat{B}_i \\ \hat{\Gamma} = \hat{\Gamma}_i \end{cases}$
 Si no converge, volver a 1)

con la inicialización

$$\begin{cases} u_0^* = u \\ y_0^* = y \\ \hat{\Gamma}_0 = 1 \end{cases}$$

Donde

$$\begin{cases} A(q^{-1}) = 1 + a_1 q^{-1} + \dots + a_n q^{-n} \\ B(q^{-1}) = b_0 + b_1 q^{-1} + \dots + b_m q^{-m} \\ \hat{F}_i(q^{-1}) = 1 + \hat{f}_1^i q^{-1} + \dots + \hat{f}_r^i q^{-r}; \hat{\Gamma}(q^{-1}) = \prod_{k=0}^i \hat{F}_k(q^{-1}) \end{cases}$$

En este algoritmo se estiman por mínimos cuadrados los polinomios A y B, con los datos filtrados.

Se ajusta una regresión sobre los residuos de la identificación. Esta autorregresión se invierte como proceso de promedios móviles (Mh) para volver a filtrar los datos que se usarán en la siguiente iteración.

En la iteración i, la estimación de A y B se hace sobre el modelo:

$$A(q^{-1}) \underbrace{\hat{F}_0(q^{-1}) \hat{F}_1(q^{-1}) \dots \hat{F}_{i-1}(q^{-1})}_{\hat{\Gamma}_i(q^{-1})} y(t) = B(q^{-1}) \underbrace{\hat{F}_0(q^{-1}) \dots \hat{F}_{i-1}(q^{-1})}_{\hat{\Gamma}_i(q^{-1})} u(t) + e_i(t) \quad \dots (6.13)$$

donde

$$e_i(t) = \frac{\hat{F}_0(q^{-1}) \hat{\Gamma}(q^{-1}) \dots \hat{F}_{i-1}(q^{-1})}{\hat{\Gamma}(q^{-1})} e(t) \quad (6.14)$$

si el algoritmo converge correctamente $\hat{F}_0(q^{-1}) \dots \hat{F}_{i-1}(q^{-1}) + \Gamma$ luego $e_i(t) + e(t)$ y los polinomios $\hat{A}_i + A, \hat{B}_i + B$

Se puede interpretar graficamente el algoritmo en la figura (6.4), como un filtrado en cascada de los datos.

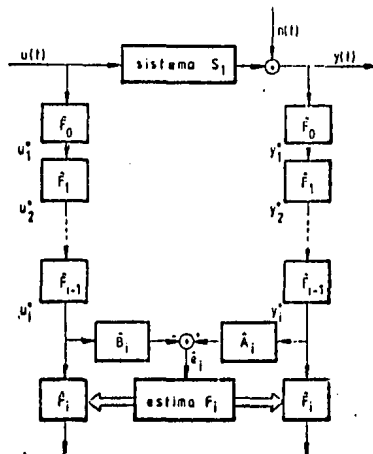


Fig 6.4

El análisis de los residuos $\hat{e}_i(t)$ en la iteración i construye el nuevo filtro $\hat{\Gamma}_i$, que se encadenará para la iteración $i+1$.

Concúrese con el estimado de Markov en la figura (6.2).

No hemos encontrado pruebas de la convergencia de este algoritmo.

(Clarke (15) señala que, aunque aún no hay pruebas sobre convergencia en general, el procedimiento mostró ser bastante bueno con datos simulados y datos reales. En particular, en la referencia que se cita hay reportada una aplicación exitosa al control de una columna de destilación binaria.

Fosterstrom (1973), (46), explica que es muy dificultoso llegar a condiciones generales para la convergencia, y muestra un ejemplo donde se pueden llegar a dos puntos de convergencia para el caso donde el ruido sea suficientemente grande.

4.3.3 Implantación del algoritmo 1: el programa MICU/GENE1

El algoritmo 1 fue implantado en la computadora B-6800 del C.S.C. de la IBM con el programa MICU/GENE1 escrito en PASCAL para dicha máquina. El listado del programa se encuentra en el apéndice D.

Como los otros programas de identificación, este tiene 2 tipos de datos: i) las señales y ii) los parámetros de la identificación.

Las señales se toman de archivos ENT y SAL de disco, con las entradas y salidas respectivamente, en forma secuencial desde $t = 1$.

Los parámetros de la identificación (grados de $A(q^{-1})$, $B(q^{-1})$, $\Gamma_i(q^{-1})$: N , M y R respectivamente, retardo K , largo de las observaciones L) se dan al programa durante la corrida, por terminal.

La figura (6.5) muestra el diagrama de flujo del programa, en sus operaciones principales.

Aunque es un procedimiento iterativo para funcionar fuera de línea y el tiempo no es crucial como en otros casos, se buscó evitar la pérdida excesiva de tiempo que implicaría un acceso reiterado a los archivos de disco, a costa de memoria, almacenando entradas, salidas y residuos en arreglos en memoria principal. De esta forma además, la aplicación del algoritmo no destruye los archivos de entrada y salida.

Nº. de operaciones

En cada iteración se hacen dos estimaciones por mínimos cuadrados, una para los parámetros de $A(q^{-1})$ y $B(q^{-1})$: $M+N+1$ parámetros, y la otra para los de $\Gamma_i(q^{-1})$: R parámetros. Las estimaciones se hacen con el algoritmo SROOT sobre L parejas de observaciones (aproximadamente).

Vemos que este algoritmo es algo más lento que el de Plackett, pero más robusto numéricamente.

Además se realiza el filtrado de entrada y salida con filtro Γ_i de orden R sobre señales de largo L . Esto hace el total de multiplicaciones y sumas (sin considerar la inicialización):

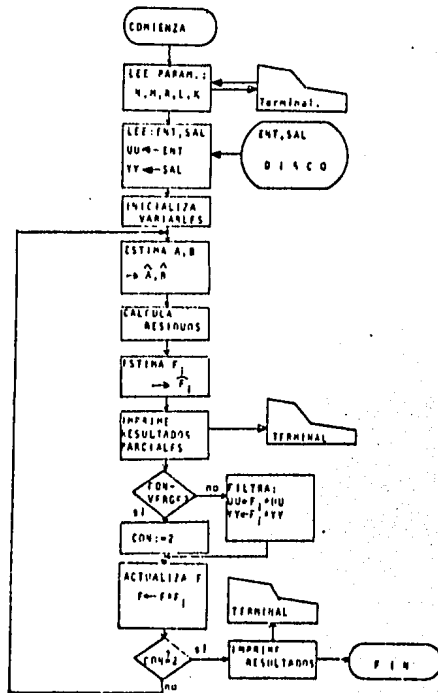


Fig 6.5

llamando $L = n^2$ de muestras, $p = MH+1$, entonces en cada iteración se realizan

- 1) Productos: por estimación de \hat{A} y \hat{B} : $\frac{L}{2} (5p^2 + 19p)$
 por filtrado de u , y : $L (2R)$
 por estimación de residuos $\hat{f}_i (p)$
 por estimación de \hat{F}_i : $\frac{L}{2} (5R^2 + 19R)$

$$\text{Total: } L(2.5p^2 + 10.5p + 2.5R^2 + 11.5R)$$

- 2) Sumas: por estimación de \hat{A} y \hat{B} : $\frac{L}{2} (3p^2 + 5p)$
 por filtrado de u , y : $L (2R)$
 por estimación de residuos $\hat{f}_i (p)$
 por estimación de \hat{F}_i : $\frac{L}{2} (3R^2 + 5R)$

$$\text{Total: } L(1.5p^2 + 3.5p + 1.5R^2 + 4.5R)$$

Condición de convergencia

Clarcke (15) propone detener el algoritmo frente a una de tres condiciones:

- i) $E_n^2 =$ (medida de los errores) diverge
- ii) el número de iteraciones ha alcanzado un máximo prefijado
- iii) la medida de los errores converge dentro de un porcentaje prefijado

La primer condición es para impedir la prosecución del procesamiento en forma inútil. No se ha incluido en el programa porque éste se usó en forma limitada, para pruebas, y no dentro de otro esquema general.

La segunda condición fue incluida en el programa como condición (a).

La tercer condición se ha incluido indirectamente al ser sustituida - por una condición de convergencia de los parámetros en norma 1, definiendo:

$$\|\hat{0}_i\|_1 = \|\hat{a}_1^i, \hat{a}_2^i, \dots, \hat{a}_n^i, \hat{b}_1^i, \dots, \hat{b}_m^i\|_1 = |\hat{a}_1^i| + \dots + |\hat{b}_m^i|$$

(donde \hat{a}_j^i es el estimado de a_j en la iteración i).

Imponemos las dos condiciones:

$$(3) \left\{ \begin{array}{l} 1) \|\hat{0}_{i+1} - \hat{0}_i\|_1 / \|\hat{0}_i\|_1 < \text{constante prefijada} \\ 2) \|\hat{r}_i - 1\|_1 < \text{constante prefijada} \end{array} \right.$$

Con estas dos condiciones aseguramos convergencia de los parámetros.

Se ha agregado una tercer condición (γ): si los residuos son muy pequeños, ya su efecto es despreciable sobre la estimación de los parámetros

a_i, b_i . Entonces:

$$(\gamma) \sum_t e^2(t) / \sum_t y^2(t) < \text{constante prefijada}$$

El programa se detiene si (a), (b) o (γ) se cumplen.

Hemos encontrado que estas tres condiciones han funcionado en forma efi-

ciente en las simulaciones.

6.3.4 Algoritmo 2

En la iteración i :

$$1) \text{ Filtrar los datos } \begin{cases} u_i^h(t) = \hat{r}_{i-1} u(t) \\ y_i^h(t) = \hat{r}_{i-1} y(t) \end{cases} \quad (6.16)$$

$$2) \text{ Para el modelo } A y_i^h = B u_i^h + a_i \\ \text{estimar por mínimos cuadrados } A \text{ y } B \Rightarrow \hat{A}_i, \hat{B}_i$$

3) Estimar las perturbaciones:

$$\hat{n}_i = \hat{A}_i y - \hat{B}_i u \quad (6.17)$$

4) Ajustar una autorregresión a las perturbaciones

$$\hat{n} = \frac{1}{1 - q^{-1}} e \Rightarrow \hat{r}_i$$

$$5) \text{ Si converge parar. Resultados } \begin{cases} \hat{A} = \hat{A}_i \\ \hat{B} = \hat{B}_i \\ \hat{r} = \hat{r}_i \end{cases}$$

Si no converge regresar a 1.

Inicialización: $r_0 = 1$

Este algoritmo difiere del anterior en que busca iterativamente una nueva estimación de la autorregresión de los residuos no como una autorregresión en cascada con la anterior, sino una nueva del mismo orden. Por eso el filtro estimado r es de orden r , cualquiera sea el n^o de iteraciones.

Parce más robusto que el anterior ya que en cada iteración no se modifican los datos sino que se busca sobre estos mismos una mejor aproximación.

En el algoritmo 1, en cada nuevo filtrado hay dos posibilidades: o se consideran condiciones iniciales arbitrarias (fue lo que hicimos, tomando como condiciones iniciales los datos en instantes previos, sin filtrar), o se toman secuencias cada vez más reducidas (utilizando las colas iniciales como condiciones iniciales), con la consiguiente pérdida de información.

Fueo arquirse, por supuesto, que el considerar condiciones iniciales "casi" arbitrarias no influirá prácticamente en el resultado final, pero en este segundo algoritmo no es necesario.

Otra ventaja es que podemos tener una gruesa estimación de la varianza de la estimación de los coeficientes de F .

Una interpretación gráfica del algoritmo es la de la figura (6.6). En cada iteración i , entradas y salidas están filtradas por $F_{i-1}(q^{-1})$, para estimar $A(q^{-1})$ y $B(q^{-1})$, mientras que la perturbación $\hat{n}(t)$ estimada sirve para calcular el nuevo filtro \hat{F}_i que reemplazará al anterior. Compárese con las figuras (6.2) y (6.4).

El algoritmo puede pensarse como un esquema de punto fijo:

$$(\hat{A}_{i+1}, \hat{B}_{i+1}, \hat{F}_{i+1}) = f(\hat{A}_i, \hat{B}_i, \hat{F}_{i-1}) ; \text{ con } \hat{A}_1, \hat{B}_1 \text{ de mín. cuad. y } \hat{F}_0 = 1$$

donde la funcional f es no lineal y dependiente de las observaciones U y Y y consideradas como parámetros de la funcional.

También puede interpretarse como una búsqueda multidimensional del mínimo de la función de costo por relajación de grupos de parámetros. Dicha función de costo es $\sum e^2(t)$ con $e(t)$ definida en 6.11.

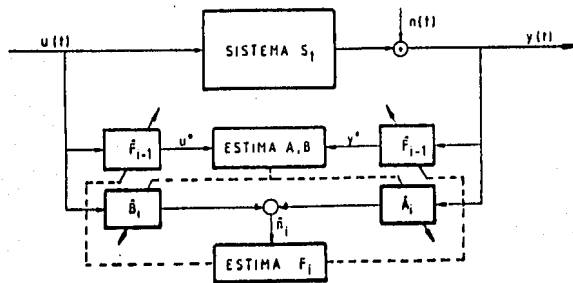


Fig 6.6

Aplicaciones reportadas

Clarke (1967), (15) reportó buenos resultados con datos simulados y reales. Otros autores, Åström y Eyckhoff (1971), (8), Strojce (1977) (10), 1980 (51) señalan resultados similares, aunque este último indica que en la práctica los estimados son buenos para altas relaciones señal a ruido, cuando esta es baja el procedimiento puede arrojar resultados incorrectos. Como veremos, esto puede explicarse por conclusiones obtenidas por Söderström que se verán en seguida.

Hasting-James & Sage (1969) (28) precisan que obtuvieron convergencia a una vecindad de los valores reales de los parámetros en todos los experimentos con datos simulados.

Söderström (1973) (46), muestra ejemplos de aplicación en la identificación del modelo de una columna de destilación binaria, en la identificación del modelo de un reactor nuclear, y en un proceso de difusión térmica de laboratorio.

En Goodwin & Payne (26), se puede encontrar la extensión a sistemas de varias entradas y varias salidas.

6.3.5 Convergencia del algoritmo 2

Söderström, arriba a resultados sobre la convergencia del algoritmo 2. Las pruebas de estos resultados pueden verse en Söderström, 1972 (45), y un resumen general en Söderström 1973 (46).

Se presentarán aquí algunas de las conclusiones más importantes. Estas se basan en un lema que muestra que el método optimiza una función de costo (para cualquier número de muestras), y luego una serie de teoremas sobre posibles puntos de convergencia cuando el número de muestras $n \rightarrow \infty$, para distintos valores de la relación señal a ruido S/N . Consideraremos $n = m = r$.

El sistema que genera las observaciones es de la forma (6.2):

$$A(q^{-1})y(t) = B(q^{-1})u(t) + n(t) \quad (6.18)$$

con

$$\begin{cases} A(q^{-1}) = 1 + a_1q^{-1} + \dots + a_nq^{-n} \\ B(q^{-1}) = b_1q^{-1} + \dots + b_nq^{-n} \\ n(t) = C(q^{-1})e(t) \end{cases} \quad (6.19)$$

donde $C(q^{-1})$ es algún filtro estable de orden finito y $e(t)$ es ruido blanco de varianza σ^2 .

Lema

El algoritmo 2 de mínimos cuadrados generalizados minimiza la función de costo

$$W(\hat{\theta}) = \frac{1}{2N} \sum_{t=1}^N e^2(t) \quad (6.20)$$

donde

$$\begin{cases} \hat{O}^T = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n, \hat{b}_1, \dots, \hat{b}_n, \hat{f}_1, \dots, \hat{f}_n) \\ e(t) = \hat{\Gamma}(q^{-1}) \left[\hat{A}(q^{-1}) y(t) - \hat{B}(q^{-1}) u(t) \right] \\ \hat{\Gamma}(q^{-1}) = 1 + \hat{\Gamma}_1 q^{-1} + \dots + \hat{\Gamma}_n q^{-n} \end{cases}$$

Más aún, si existe $\Gamma(q^{-1})$ tal que:

$$n(t) = \frac{1}{\Gamma(q^{-1})} e(t) \quad (6.21)$$

Con $e(t)$ ruido blanco gaussiano, entonces la solución del método es la de máxima verosimilitud aplicada al sistema (6.18).

Observaciones:

1. Si se cumple (6.21) entonces la estimación tiene propiedad de consistencia, y es asintóticamente eficiente y normal cuando $n \rightarrow \infty$.
2. $W(\hat{O}_k)$, con $k = 1, 2, \dots$ (k es el número de iteraciones del algoritmo) forma una sucesión monótona decreciente y acotada por lo que es convergente. Los posibles puntos límites de \hat{O}_k deben ser puntos estacionarios de la función $W(\hat{O})$. Se demuestra que para que estos puntos sean estables (esto es que si se comienza en un entorno de dicho punto lo suficientemente pequeño, se converge hacia él), deben ser mínimos relativos de $W(\hat{O})$. Un punto silla, por ejemplo, no es estable.

En el desarrollo que sigue se hacen las siguientes suposiciones:

a) la entrada $u(t)$ es:

$$u(t) = u_1(t) + G(q^{-1}) e_1(t) \quad (6.22)$$

donde

$$\begin{cases} u_1(t) \text{ es determinística y periódica} \\ G(q^{-1}) \text{ es un filtro estable de orden finito} \\ e_1(t) \text{ es ruido blanco} \end{cases}$$

b) $u(t)$ y $e_1(t)$ son independientes.

La función $W(\hat{O})$ minimizada depende del n^2 de observaciones N . Pero con las dos suposiciones anteriores puede ser probado que

$$W(\hat{O}) \xrightarrow{\text{prob. 1}} V(\hat{O}) \text{ para } N \rightarrow \infty \quad \forall \hat{O}, \text{ donde}$$

$$\begin{aligned} V(\hat{O}) = & \frac{1}{2} E \left[\frac{\hat{\Gamma}(q^{-1})}{\Lambda(q^{-1})} \left(\hat{A}(q^{-1}) B(q^{-1}) - \Lambda(q^{-1}) \hat{B}(q^{-1}) \right) u(t) \right]^2 + \\ & + \frac{1}{2} E \left[\frac{\hat{A}(q^{-1}) \hat{B}(q^{-1})}{\Lambda(q^{-1})} n(t) \right]^2 \end{aligned} \quad (6.23)$$

donde $E \left[u_1(t) \right]^2$ se entiende como $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N u_1^2(i)$

La función $V(\hat{O})$ queda definida entonces sobre R^{2n} ; está claro que no todo este espacio es de interés. Se restringe el conjunto de \hat{O} factibles para simplificar un poco el análisis con las siguientes suposiciones adicionales:

- c) $\hat{A}(z)$ tiene todos sus ceros fuera del círculo $|z| = r > 1$ (r próximo a 1)

1) $\hat{\Gamma}(z)$ tiene todos sus ceros fuera del círculo $|z| = r > 1$ (r próximo a 1)

e) $\hat{b}_i, 1 \leq i \leq n$, son acotados.

Se investigan ahora los mínimos relativos de $V(\hat{\theta})$, que son los posibles puntos de convergencia del algoritmo al minimizar $W(\hat{\theta})$ para $N \rightarrow \infty$.

Se distinguen dos situaciones: relación señal a ruido grande y relación señal a ruido pequeña.

Relación señal a ruido grande

Teorema 6.1

Sea el sistema (6.18), de orden n , controlable. La entrada $u(t)$ es persistentemente excitante de orden $2n$ y el orden del modelo es n . Los posibles valores de $\hat{\theta}$ se encuentran en un conjunto Ω arbitrario, compacto.

Entonces:

Existe una constante S_0 tal que si $S_0 \leq S/N < \infty$, la función de costo (6.23) tiene solamente un punto estacionario en $\hat{\theta}$. Este punto es un mínimo local de (6.23) y satisface:

$$\begin{cases} \hat{a}_i = a_i + O(1/S), & S \rightarrow \infty & i = 1, 2, \dots, n \\ \hat{b}_i = b_i + O(1/S), & S \rightarrow \infty & i = 1, 2, \dots, n \\ \hat{f}_i = \bar{f}_i + O(1/S), & S \rightarrow \infty & i = 1, 2, \dots, n \end{cases} \quad (6.24)$$

Donde, llamando $\bar{F}(q^{-1}) = 1 + \bar{f}_1 q^{-1} + \dots + \bar{f}_n q^{-n}$, se cumple que $(\bar{f}_1, \bar{f}_2, \dots, \bar{f}_n)$ es el punto que minimiza la función

$$V_2(\bar{f}_1, \dots, \bar{f}_n) = E \left[\left\{ \bar{F}(q^{-1}) n(t) \right\}^2 \right] \quad (6.25)$$

Observaciones:

1) Nótese que el teorema no asegura convergencia asintótica de los polinomios $\hat{A}(q^{-1})$ y $\hat{B}(q^{-1})$ a $A(q^{-1})$ y $B(q^{-1})$, sino a una proximidad de estos.

La diferencia de los parámetros a_i y b_i con sus estimadores (asintóticos) es un infinitésimo de orden por lo menos $1/S$ ($S \rightarrow \infty$). Si se asegura $\hat{a}_i \rightarrow a_i, \hat{b}_i \rightarrow b_i$ si $N \rightarrow \infty, S \rightarrow \infty$.

2) La restricción sobre la clase de señales de entrada es muy importante; si no se cumple, la función de costo $V(\hat{\theta})$ puede tener varios mínimos locales.

3) La suposición de controlabilidad del sistema es debida a que si existen factores comunes entre $A(q^{-1})$ y $B(q^{-1})$, la ecuación (6.18) puede ser dividida entre esos factores, y se llega a un sistema de menor orden que el anterior, con otra correlación de las perturbaciones. Esta situación se discute en el siguiente teorema.

4) La condición (6.25) de minimización que cumple $\bar{F}(q^{-1})$, puede ser puesta, de acuerdo a (6.19), como que el polinomio $\bar{F}(q^{-1}) C(q^{-1})$ es el que mejor aproxima a 1 en $\| \cdot \|_2$; en efecto, sea:

$$D(q^{-1}) = \bar{F}(q^{-1}) C(q^{-1}) = 1 + d_1 q^{-1} + \dots + d_{2n} q^{-2n}$$

entonces, $(\bar{f}_1, \dots, \bar{f}_n)$ es tal que

$$(d_1)^2 + (d_2)^2 + \dots + (d_{2n})^2 = \text{un mínimo}$$

Teorema 6.2

Sea el sistema (6.18), controlable, de orden n . Se asume un modelo de orden $n+p$, $p > 0$ y $u(t)$ persistentemente excitante de orden $2n+p$. Se consideran estimados $\hat{\theta}$ en un conjunto $\hat{\Omega}$ compacto arbitrario. Entonces:

Existe una constante S_0 tal que si $S_0 \leq S/N < \infty$, se cumple que

1) todos los mínimos locales de $V(\hat{\theta})$ satisfacen

$$\begin{cases} \hat{A}(q^{-1}) = A(q^{-1}) + o(1), & S \rightarrow \infty \\ \hat{B}(q^{-1}) = B(q^{-1}) + o(1), & S \rightarrow \infty \end{cases} \quad (6.25)$$

con

$$L(q^{-1}) = 1 + \ell_1 q^{-1} + \dots + \ell_p q^{-p};$$

además

$$\begin{cases} \hat{L}(q^{-1}) = L(q^{-1}) + o(1), & S \rightarrow \infty \\ \hat{\Gamma}(q^{-1}) = \bar{\Gamma}(q^{-1}) + o(1), & S \rightarrow \infty \end{cases} \quad (6.26)$$

donde

$\hat{\theta}_1, \dots, \hat{\theta}_p, \hat{f}_1, \dots, \hat{f}_{n+p}$ es un punto estacionario de

$$V_2(\hat{\theta}_1, \dots, \hat{\theta}_p, \hat{f}_1, \dots, \hat{f}_{n+p}) = E \{ [L(q^{-1}) \hat{\Gamma}(q^{-1}) n(t)]^2 \} \quad (6.27)$$

La matriz de derivadas de segundo orden (Hessiana) de V_2 en dicho punto debe ser positiva definida o positiva semidefinida.

ii) Si la matriz de segundas derivadas de V_2 es positiva definida en $(\hat{\theta}_1, \dots, \hat{f}_{n+p})$ entonces existe un único mínimo local, y los infinitesimos $o(1)$ en las fórmulas (6.25) y (6.26) pueden ser reemplazadas por infinitesimos de orden mayor o igual que 1 respecto $1/S$ o $1/N$.

Además la matriz V'' es positiva definida en este punto.

Observación

De (6.25) se desprende que los mínimos tienen la propiedad:

$$\frac{\hat{B}(q^{-1})}{\hat{A}(q^{-1})} = \frac{\hat{B}_2(q^{-1})}{\hat{A}_2(q^{-1})} + o(1), \quad S \rightarrow \infty$$

Cuando la relación señal a ruido es pequeña, se puede ver que en general no existe un mínimo único de la función $V(\hat{\theta})$.

Dirigirse que el ruido $n(t) = C(q^{-1}) e(t)$ satisface la "condición de ruido", si existen al menos dos parejas de polinomios $\hat{A}_1(q^{-1})$, $\hat{\Gamma}_1(q^{-1})$ y $\hat{A}_2(q^{-1})$, $\hat{\Gamma}_2(q^{-1})$ tales que:

$$V_0(a_1, \dots, a_n, f_1, \dots, f_n) = E \left\{ \frac{\hat{A}(q^{-1}) \hat{\Gamma}(q^{-1}) \hat{C}(q^{-1})}{\hat{A}(q^{-1})} e(t) \right\}^2 \quad (6.28)$$

tiene mínimos locales con matrices de derivadas de segundo orden positivas definidas en $(a_1^1, \dots, a_n^1, f_1^1, \dots, f_n^1)$ y $(a_2^2, \dots, a_n^2, f_2^2, \dots, f_n^2)$

Teorema 6.3

Sea $u(t)$ persistentemente excitante de orden n , y $n(t)$ satisface la condición del ruido. Entonces existe un número $S_1 > 0$ tal que si $0 < S/N \leq S_1$, la función $V(0)$ tiene más de un mínimo local. Dos de ellos tienen la forma:

$$\begin{aligned}\hat{A}(q^{-1}) &= \hat{A}_i(q^{-1}) + 0(s), s \rightarrow 0 \\ \hat{F}(q^{-1}) &= \hat{F}_i(q^{-1}) + 0(s), s \rightarrow 0\end{aligned}\quad (6.29)$$

con \hat{A}_i, \hat{F}_i arajas de polinomios que se derivan de la definición de "condición del ruido". Los puntos mínimos dan residuos $c_1(t), c_2(t)$ que satisfacen:

$$c_1(t) - c_2(t) = 0 (S^{1/2}) \quad (6.30)$$

6.3.6 La implantación del algoritmo 2: el programa MICU/GENE 2

El algoritmo 2 fue implantado en la computadora B-6800 del CSC de la UAM mediante el programa MICUGENE2, escrito en Pascal para dicha máquina.

Es muy similar el programa MICU/GENE1 tanto en los datos requeridos como en la forma de entregar resultados.

La única gran diferencia es que las señales de entrada y salida se vuelven a filtrar sin modificarse, en cada iteración por el filtro \hat{F}_i estimado en cada iteración.

El diagrama de flujo se muestra en la figura (6.7). El listado puede encontrarse en el apéndice J.

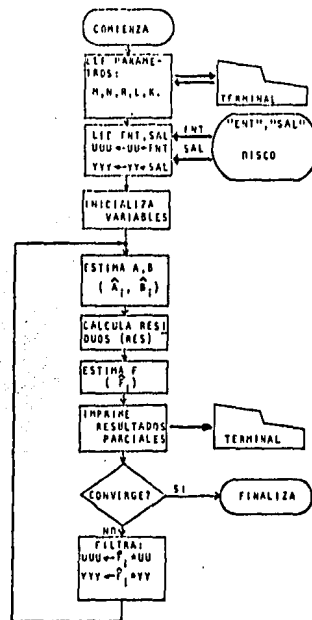


DIAGRAMA DE FLUJO DEL PROGRAMA MICU/GENE2 ..

El número de operaciones es casi exactamente el mismo en el programa MICU/GENE1, la única diferencia es que MICU/GENE2 no hay que realizar la actualización del filtro F, de $R^2 \times T^2/2$ sumas y productos ($I = n^\circ$ de iteraciones).

6.4 Experimentos con datos simulados

6.4.1 Las simulaciones

Para observar si las propiedades de los métodos descritos se cumplen para un número N de observaciones, finito, y para ejemplificar dichas propiedades, se han hecho algunos experimentos con datos provenientes de sistemas simulados.

Se simularon dos sistemas de primer orden con perturbaciones correlacionadas. Las características de las simulaciones son las siguientes:

Sistema E:

Se rige por la ecuación en diferencias:

$$\{1 - 0.8q^{-1}\}y(t) = \{q^{-1}\}u(t) + \lambda \{1 - 0.7q^{-1}\}e(t) \quad (6.31)$$

con $e(t)$: ruido blanco de varianza 1, media cero. Si $u(t)$ es blanco de varianza 1, media cero, los valores de λ para distintos S/N son:

S/N = 10	$\lambda = 0.519875$
S/N = 4	$\lambda = 0.421994$
S/N = 1	$\lambda = 1.643988$
S/N = 0.1	$\lambda = 5.19875$

Estos valores de λ fueron calculados bajo las hipótesis mencionadas, con el programa VARIANZA/SALIDA que ejecuta el algoritmo descrito en el Capítulo 4.

Sistema f:

Se rige por la ecuación en diferencias:

$$(1-0.8q^{-1})y(t) = (q^{-1})u(t) + \frac{\lambda}{(1-0.3q^{-1})}e(t) \quad (6.32)$$

con $e(t)$ ruido blanco de media cero y varianza unitaria. Si $u(t)$ es ruido blanco, media cero y varianza unitaria, entonces los valores de λ para distintas relaciones señal a ruido S/N, son:

S/N = 10	$\lambda = 0.230166$
S/N = 4	$\lambda = 0.373411$
S/N = 1	$\lambda = 0.746821$
S/N = 0.1	$\lambda = 2.36166$

Antes sistemas fueron simulados con el programa SIMULA, con entrada $u = \text{FRAS127}$, perturbación $e = \text{RUIDOS}$ que simula una señal de ruido blanco, media cero, varianza unitaria y distribución normal.

En todos los casos se simularon 1127 observaciones de la salida, de las cuales se descartaron las primeras 127. Las 1000 restantes se guardaron en archivos en disco de la computadora B-6800 del CSC de la IZAM para ser usadas en diversas pruebas.

Las simulaciones se hicieron para distintos valores de S/N; los nombres de los archivos de salida de cada una de estas simulaciones se indican en la siguiente tabla:

S/N	10	4	1	0.1
Sistema E	SALE502	SALF503	SALF504	SALE505
Sistema F	SALF502	SALF503	SALF504	SALF505

Estas señales serán usadas a lo largo de este capítulo y el siguiente para identificar los sistemas que las generaron a través de distintos algoritmos.

En los siguientes párrafos se procederá a identificar los sistemas con los algoritmos de mínimos cuadrados generalizados 1 y 2.

En los primeros dos párrafos se comparan los resultados de dichos algoritmos, y en los siguientes se comprueban las propiedades del algoritmo 2.

6.4.2 Algoritmo 1

Se identificó el sistema E suponiendo conocidos: el retraso (un instante de muestreo), y los órdenes $M(M=0)$ y $N(N=1)$. Este proceso se repitió para distintos valores del orden R de los Filtros Γ_i de cada iteración ($R = 1, 2, 3, 4$) y para distintos niveles de ruido: S/N = 0.1, 1, 4 y 10.

En las siguientes tablas se muestran los valores obtenidos para los estimadores \hat{a}_1 , \hat{b}_1 , la función de costo $V(0)$, y el número de iteraciones que fueron necesarias para arribar a la convergencia.

Recordamos que la función de costo es:

$$V(\hat{\theta}) = \sum e^2(t)$$

donde los residuos estimados $\hat{e}(t)$ son:

$$\hat{e}(t) = \hat{A}(q^{-1}) (\hat{F}(q^{-1}) y(t)) - \hat{B}(q^{-1}) (\hat{F}(q^{-1}) u(t)) \quad (6.33)$$

siendo

$$\left\{ \begin{array}{l} 1/\hat{F}(q^{-1}) = 1/\hat{r}_0! \dots \hat{r}_1 \text{ la aproximación autorregresiva del proceso del ruido} \\ \hat{A}(q^{-1}) \text{ el estimado del polinomio } A(q^{-1}) \\ \hat{B}(q^{-1}) \text{ el estimado del polinomio } B(q^{-1}) \end{array} \right.$$

Las condiciones de convergencia fueron:

$$\left\{ \begin{array}{l} a: n^{\circ} \text{ de iteraciones máximo} = 50 \\ B.1: \|\hat{\theta}_{i+1} - \hat{\theta}_i\| / \|\hat{\theta}_i\| < 10^{-3} \\ B.2: \|\hat{F}_1 - 1\| \leq R \times 10^{-3} \\ \gamma: \sum_{t=1}^n \hat{e}^2(t) / \sum_{t=1}^n y^2(t) \leq 10^{-3} \end{array} \right. \quad (6.34)$$

Tabla 6.1 Número de iteraciones en la identificación del sistema E con el algoritmo 1

R \ S/N	1	2	3	4
10	7	6	5	5
4	8	6	6	6
1	9	9	10	10
0.1	19	27	35	45

Tabla 6.11. Estimación de los parámetros del sistema E para distintos valores de S/N y R, con el algoritmo 1.

S/N	Par.	R=1	2	3	4
10	a_1	-0.7846	-0.7841	-0.7841	-0.7821
	b_0	0.9913	0.989	0.991	0.9922
4	a_1	-0.7603	-0.7889	-0.7934	-0.7943
	b_0	0.9915	0.9854	0.9872	0.9847
1	a_1	-0.7151	-0.772	-0.7826	-0.7823
	b_0	1.0028	0.981	0.9794	0.9894
0.1	a_1	-0.5713	0.7168	-0.7885	0.7462
	b_0	1.0374	0.9621	0.9441	0.9459

Valores reales: $a_1 = -0.8$
 $b_0 = 1$

Tabla 6.111. Función de costo obtenida en la identificación del sistema E, para distintos R y S/N.

R \ S/N	1	2	3	4
10	307.08	281.63	275.63	275.19
4	755.99	699.64	688.35	689.41
1	2927.25	2778.4	2753.25	2761.54
0.1	28084.73	27593.82	27506.05	27539.13

la velocidad de convergencia la discutiremos más adelante, veamos primero cómo son las estimaciones \hat{a}_1 y \hat{b}_1 en relación al nivel de ruido (a través de S/N) y del orden R de los factores que aproximan la autoregresión del ruido.

Dependencia del orden R

En estas simulaciones se han encontrado que la elección del orden R es importante, pero que no arroja ninguna ventaja aumentando indefinidamente.

Las figuras (6.8) y (6.9) muestran los resultados de la estimación \hat{a}_1 y \hat{b}_1 contra el orden R , para distintas relaciones señal a ruido.

Recordemos que los valores reales de a_1 y b_1 son -0.8 y 1 respectivamente.

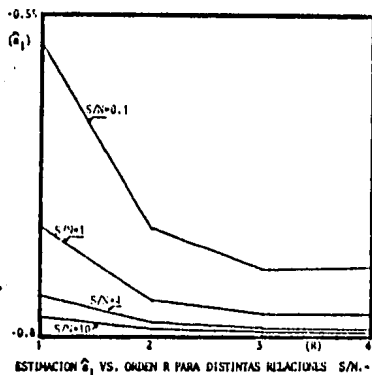


Fig 6.8

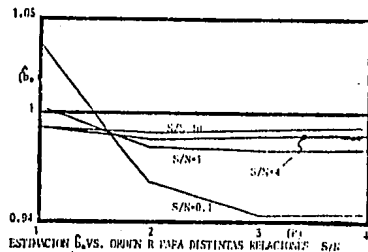


Fig 6.9

Se puede observar que independientemente de la relación señal a ruido, la estimación \hat{a}_1 mejora al aumentar R .

Se obtiene una mejor estimación en $R = 2$ que en $R = 1$, y una mejor estimación en $R = 3$ que en $R = 2$.

Sin embargo seguir aumentando R no mejora la estimación de a_1 . Es obvio que la mejora se obtiene puesto que órdenes mayores en los factores de $\Gamma(q^{-1})$ dan más grados de libertad en la aproximación.

Pero seguir aumentando R otorga más grados de libertad en los coeficientes de órdenes superiores, que en un filtro estable tienden a ser cada vez menos importantes. Por eso es que se consigue una mejora cada vez menor al incrementar R ; más aún, para $R = 4$ la estimación de a_1 con $S/N = 0.1$ es ligeramente inferior a la obtenida en $R = 3$ (hecho que se compensa al mejorar ligeramente la estimación de b_1).

La estimación de b_0 es en general mejor que la de a_1 (nótese que las escalas de las figuras (6.8) y (6.9) son muy distintas).

Circaamente la estimación de b_0 para $R = 1$ es algo mejor que para $R = 2, 3, 4$ con $S/N = 0.1, 1, 4$.

De $R = 2$ en adelante la estimación de b_0 parece depender mucho más del nivel de ruido que de el orden R , aunque para $S/N = 10$ y $S/N = 4$ la estimación es mejor en $R = 3$ que en $R = 2$ y en $R = 4$ que en $R = 3$.

En $R = 4$ todas las estimaciones son mejores que en $R = 3$, aunque la tendencia general parece ser, como en el caso de \hat{a}_1 , que no hay ventajas del incremento de R . Más aún, las diferencias $|\hat{a}_1 - a_1|$ $|\hat{b}_0 - b_0|$ no solo permanecen bastante estables al aumentar R , sino que además, para cada valor de S/N son bastante similares.

La observación hecha sobre el efecto de aumentar el orden R puede comprobarse una vez más observando la variación de la función de costo de la identificación. En la figura (6.10) se han dibujado las funciones de costo obtenidas en la identificación para $S/N = 0.1, 1, 4, 10$ contra el orden R de cada identificación. Fueron normalizadas multiplicándolas por el factor $((S/N) \times 10^{-3})$ para poder ser dibujadas con una misma escala.

Nótese la franca disminución de aumentar R hasta 3 y el franco estancamiento al aumentarlo de 3 a 4. Con los datos utilizados, el valor real de la suma de los cuadrados de las perturbaciones $u(t)$ corresponde en la figura al valor 2.76235. En la siguiente tabla se listan los valores de las funciones de costo normalizadas obtenidas con $R = 3$ y $R = 4$.

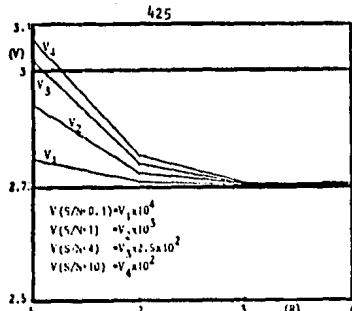
Tabla 6.IV. Valores normalizados de la función de costo, para $R = 3$ y $R = 4$

S/N	función normal.	R = 3	R = 4
10	V_4	2.7563	2.7519
4	V_3	2.7534	2.7576
1	V_2	2.7533	2.7615
0.1	V_1	2.7506	2.7537

Dependencia del nivel de ruido

En la figura (6.8) se puede apreciar que cualquiera sea el valor de R , la estimación de a_1 es mejor cuando menor sea la potencia del ruido (mayores valores de S/N). En la figura (6.9) puede verse que excepto para $R = 1$, se cumple también que la estimación de b_0 es mejor a menor potencia del ruido.

En la figura (6.11) se ilustra mejor la situación para $R = 1$. Se grafican las estimaciones de a_1 y b_0 contra S/N , para $R = 1$.



Funciones de costo de las identificaciones del sistema E vs. orden R, para relaciones señal a ruido: S/N=0,1,1,4 y 10.

Fig 6.10

La estimación de a_1 mejora al disminuir la potencia del ruido.

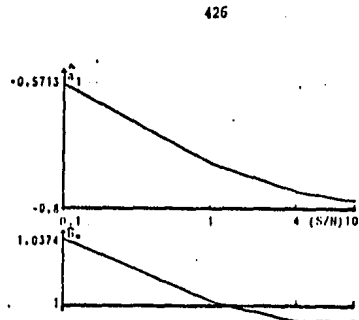
Para b_0 , al pasar \hat{b}_0 de un valor por exceso a uno por defecto del valor real, se para en S/N = 1 por una estimación muy cercana al valor real b_0 .

Las figuras (6.12) y (6.13) muestran la evolución de los estimadores \hat{a}_1 y \hat{b}_0 , al variar S/N de 0.1 a 10, parametrizado en el orden R (R = 1, 2, 3, 4).

En la figura 6.12 puede observarse claramente como converge \hat{a}_1 al valor real a_1 , para S/N $\rightarrow \infty$ cualquiera sea el valor de R. Por ejemplo para R = 1 el error de la estimación sigue muy aproximadamente una ley:

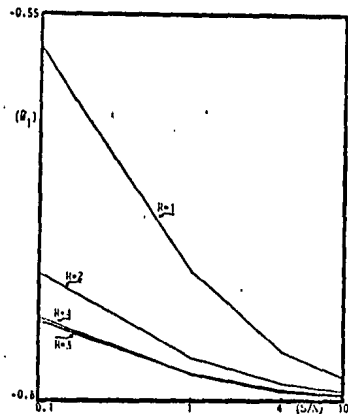
$$|\hat{a}_1 - a_1| = 0.844(S/N)^{-0.71}$$

Puede verse además que no se obtienen diferencias apreciables al aumentar R de 3 a 4.



Estimación \hat{a}_1, \hat{b}_0 , vs. S/N para R=1, algoritmo 1.

Fig 6.11



ESTIMACION \hat{a}_1 VS. RELACION S/N PARA DISTINTOS ORDENS R.

Fig 6.12

En la figura (6.14) se amplía la parte inferior derecha de la figura (6.12) para distinguir las curvas correspondientes a $R = 3$ y $R = 4$ que aparecen encimadas en dicha figura.

En la figura (6.13) se observan las estimaciones \hat{b}_0 para $R = 1, 2, 3, 4$ graficadas contra la relación señal a ruido S/N .

Puede apreciarse la tendencia convergente de \hat{b}_0 y b_0 para S/N creciente. Las líneas correspondientes a $R = 1, 3$ y 4 siguen una tendencia bastante similar.

En particular para $R = 3$ y $R = 4$ el error absoluto en la estimación sigue bastante aproximadamente una ley:

$$|\hat{b}_0 - b_0| = 0.022 (S/N)^{-0.4}$$

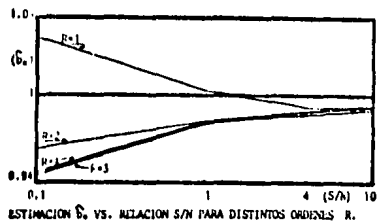


Fig 6.13

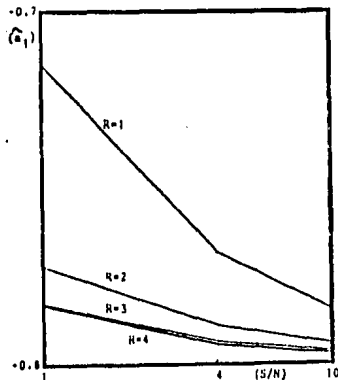


Fig 6.14

Estimación \hat{a}_1 vs. S/N para distintos órdenes R.

En la figura (6.15) se amplía la parte izquierda de la figura (6.13) (correspondiente a $S/N = 1, 4, 10$), para distinguir las líneas que se superponen en dicha figura.

En resumen, se ha podido apreciar una tendencia de convergencia de los parámetros estimados a sus valores reales, cuando la potencia del ruido tiende a disminuir.

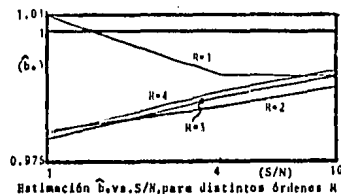


Fig 6.15

Identificación del proceso de ruido

Al mismo tiempo que se estimaron los parámetros a_1 y b_1 , fue aproximado el proceso del ruido como $1/F(q^{-1}) = 1/F_0(q^{-1}) \dots F_R(q^{-1})$ siendo i el n.º de iteraciones.

En el sistema simulado que ha sido identificado, el proceso del ruido es:

$$n(t) = \lambda(1 - 0.7 q^{-1}) e(t)$$

con $e(t)$ ruido blanco de media cero y varianza unitaria (simulando una distribución normal).

Se pretende aproximar el proceso M.A.1 $C(q^{-1}) = (1 - 0.7 q^{-1})$ con un proceso A.R.1 $C(q^{-1}) = \frac{1}{F(q^{-1})}$; la aproximación es exacta si

$$F(q^{-1}) = 1 + 0.7q^{-1} + 0.49q^{-2} + 0.343q^{-3} + 0.2401q^{-4} + 0.1681q^{-5} + \dots \quad (6.35)$$

La expansión $F(q^{-1}) = \sum_{i=0}^{\infty} f_i q^{-i}$ contiene infinitos términos. Los primeros 6 coeficientes (escritos en (6.35)) son tales que $\sum_{i=0}^6 f_i^2 > 90.62\% \left(\sum_{i=0}^{\infty} f_i^2 \right)$, es decir que los coeficientes no escritos en (6.35) pesan muy poco respecto a los seis primeros. Por eso los tomamos en cuenta para juzgar la bondad de la estimación $\hat{F}_i(q^{-1})$, y para distintos valores de la relación señal a ruido de la simulación.

Puede observarse que la aproximación de f_i ($i = 1, 2, \dots$) depende fuertemente del orden R , más que de la potencia del ruido. Debo recomen-

darse que se ha impuesto la condición $\hat{f}_0 = f_0 = 1$.

Para un orden R dado, se obtienen buenas aproximaciones de $f_1 \dots f_R$, siendo bastante pobres las estimaciones de $f_{R+1}, f_{R+2} \dots$

En la tabla V se han recuadrado las estimaciones de f_i con $i \leq R$. En cuanto a la influencia del nivel de ruido, puede apreciarse que dentro de los estimados \hat{f}_i con $i \leq R$ siempre es \hat{f}_R el estimador que empeora más rápidamente al bajar la relación señal a ruido.

Para los otros estimadores \hat{f}_i $i < R$ se obtiene una buena estimación dentro de los niveles de ruido correspondientes a $S/N = 10, 1, 1$, pero la estimación se empeora enormemente para $S/N = 0.1$, en forma análoga con la peor estimación de b_0 y a_1 .

6.4.3 Identificación del sistema E con el algoritmo 2

El sistema E fue identificado con el algoritmo 2 para relaciones señales a ruido $S/N = 10, 4, 1, 0.1$ y órdenes de aproximación del proceso del ruido: $R = 1, 2, 3, 4$.

En las siguientes tablas se presentan el resultado de la estimación de los parámetros a_1 y b_0 , y la función de costo asociada a la identificación.

Más adelante, donde se discute la estimación sobre las perturbaciones, se incluyen los resultados de la estimación de los parámetros del proceso autoregresivo $1/F(q^{-1})$.

Tabla 6.V. Estimación del proceso del ruido del sistema E con el algoritmo 1, para $R = 1, 2, 3, 4$ y $S/N = 0.1, 1, 4, 10$.

K	S/N	f_1 0.7	f_2 0.49	f_3 0.343	f_4 0.2401	f_5 0.1681
1	10	0.684	0.1252	0.0073	0.0001	
	4	0.6672	0.134	0.0095	0.0002	
	1	0.6176	0.1421	0.0152	0.0008	
	0.1	0.4075	0.1095	0.0146	0.0013	
2	10	0.716	0.4881	0.1531	0.0507	0.0062
	4	0.7101	0.4793	0.1697	0.0532	0.0101
	1	0.6241	0.4547	0.1812	0.0556	0.0143
	0.1	0.6417	0.3800	0.1511	0.0476	0.0123
3	10	0.7202	0.4206	0.3131	0.1262	0.0472
	4	0.7166	0.4858	0.3057	0.1343	0.0416
	1	0.7057	0.4697	0.2909	0.1300	0.0317
	0.1	0.6734	0.4234	0.2406	0.1051	0.0311
4	10	0.7205	0.4943	0.3179	0.1411	0.0204
	4	0.7174	0.4901	0.3129	0.1772	0.0251
	1	0.7056	0.4711	0.2416	0.1561	0.0147
	0.1	0.6712	0.4305	0.2175	0.1072	0.0383

En la identificación del sistema se supuso conocido el retraso ($k=1$), el orden $M(N=0)$ y el orden $N(N=1)$.

Tabla 6.VI. Estimación de los parámetros a_1 y b_0 del sistema E, con el algoritmo 2.

S/N	parám.	$k = 1$	2	3	4
10	a_1	-0.7746	-0.7677	-0.7430	-0.7255
	b_0	0.7302	0.9431	0.9431	0.9307
4	a_1	-0.7450	-0.7759	-0.7859	-0.7916
	b_0	0.9912	0.9480	0.9361	0.9473
1	a_1	-0.6496	-0.7264	-0.7611	-0.7787
	b_0	1.0067	0.9963	0.9654	0.9810
0.1	a_1	-0.4442	-0.6009	-0.6089	-0.7407
	b_0	1.0474	1.0196	0.9774	0.9403

Tabla 6.VII. Función de costo asociada a la estimación del sistema E, algoritmo 2, para distintos valores de S/N y R .

S/N	R = 1	2	3	4
10	316.67	290.23	279.74	275.45
4	782.91	722.92	668.49	608.3
1	3025.89	2858.27	2783.36	2750.12
0.1	28364.67	27872.01	27589.03	27433.84

El número de iteraciones necesario para arribar a la convergencia en cada uno de estos casos se indica en la tabla 6.VIII.

Tabla 6.VIII. N° de iteraciones empleadas por el algoritmo 2 para converger en la identificación del sist. E

S/N	R = 1	2	3	4
10	4	4	4	5
4	4	5	5	6
1	6	7	8	10
0.1	12	22	39	52

Las condiciones utilizadas en la identificación para detectar con convergencias fueron similares a las (6.34) utilizadas para el algoritmo 1:

$$\left\{ \begin{array}{l} \alpha: n_0^2 \text{ de iteraciones máxima} = 60 \\ \beta_1: \|\hat{O}_{i+1} - \hat{O}_i\|_1 / \|\hat{O}_i\|_1 \leq 10^{-3} \\ \beta_2: \|\hat{r}_{i+1} - \hat{r}_i\|_1 / \|\hat{r}_i - 1\|_1 \leq 10^{-3} \\ \beta_3: \sum_t \hat{c}(t)^2 / \sum_t y^2(t) \leq 10^{-4} \end{array} \right. \quad (6.36)$$

El programa se detiene al cumplirse α , ó dos de las condiciones β . \hat{O}_i es el vector de parámetros estimado en la iteración i : $\hat{O}_i = \{\hat{O}_i^1, \dots, \hat{O}_i^P\}^T$; \hat{r}_i es el vector de parámetros del proceso del ruido en la iteración i :

$$\hat{r}_i = \{1, f_i^1, \dots, f_i^R\}^T$$

y definimos

$$\|\hat{O}_i\|_1 = \sum_{j=1}^P |\hat{O}_i^j|$$

Dependencia del orden R

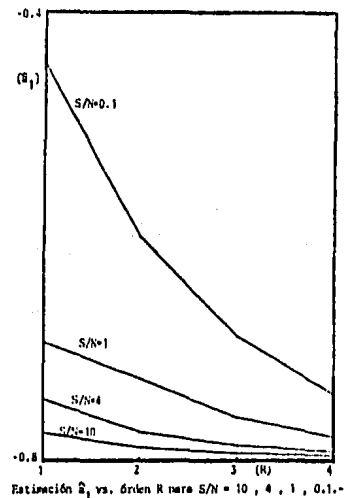
Discutiremos primero como dependen \hat{a}_1 y \hat{b}_0 calculadas en estas estimaciones respecto R y S/N , más adelante veremos las aproximaciones al proceso del ruido.

Se espera que al aumentar R aumenta la precisión con que se aproxima el proceso de las perturbaciones, e indirectamente entonces de tener una mejor aproximación de la parte determinística.

En la tabla 6.VI pueden verse las estimaciones de a_1 y b_0 . En las figuras (6.16) y (6.17) se dibujan las estimaciones de a_1 y b_0 contra el orden R para distintos valores de la relación señal a ruido.

Para la estimación de a_1 puede comprobarse que para cualquiera de las relaciones S/N , la estimación se aproxima monotonamente al valor real (-0.8), para R crecientes. La mejora (absoluta) de aumentar el orden R en una unidad es cada vez menor para valores de R cada vez mayores.

En cuanto a la estimación de b_0 , ésta parece mejor en el sentido de que para todos los valores de R y S/N experimentados, el valor de \hat{b}_0 nunca se aleja ni al 6% de distancia de su valor real (1); esto es que está comprendido dentro del intervalo (0.94, 1.05).



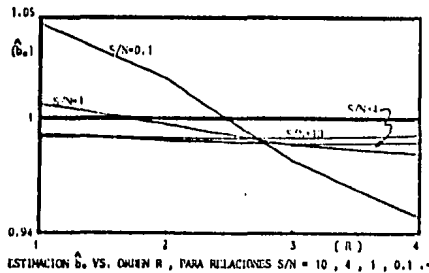


Fig 6.17

Para los valores más altos de S/N , la estimación \hat{b}_0 prácticamente no depende de R , y es próxima al valor real (oscila alrededor de 0.989).

Para valores más bajos de (S/N) , especialmente para $S/N = 0.1$ es cuando se obtienen las peores estimaciones. Laxyo veremos que precisamente $S/N = 0.1$ esta cerca del valor s_1 mencionado en el teorema 6.3, a partir del cual se obtienen varios puntos de convergencia y la identificación se vuelve muy pobre y poco confiable.

De todas maneras se repitió la identificación para $S/N = 0.1$ y valores más altos de R para ver si se mantiene la tendencia de alejamiento observada en la figura (6.17), pero se comprobó un estancamiento de \hat{b}_0 al aumentar R , también para $S/N = 0.1$

Claramente no se obtiene una estimación mejor de b_0 al seguir aumentando el valor de R .

Con esto y los resultados observados para \hat{a}_1 puede comprobarse la observación de Söderström et al (47), en el sentido de que no tiene efecto apreciable seguir aumentando el orden R del polinomio $\Gamma(q^{-1})$

Esta afirmación puede comprobarse una vez más observando la variación de la función de costo: $V(\hat{\theta}, \hat{\Gamma})$

$$V(\hat{\theta}, \hat{\Gamma}) = \sum_{t=1}^T \hat{e}(t)^2$$

donde

$$\hat{e}(t) = \hat{A}(q^{-1}) \{\Gamma(q^{-1}) y(t)\} - \hat{B}(q^{-1}) \{\hat{\Gamma}(q^{-1}) u(t)\} \quad (6.37)$$

En la figura (6.18) se dibujan las funciones de costo normalizadas, de cada relación S/N , contra el orden R .

Las funciones normalizadas son:

$$V_1 = V(S/N) \times S/N \times 10^{-3}$$

Puede observarse que la mejora obtenida en aumentar R en una unidad es cada vez menor para R creciente.

La disminución en aumentar R de 4 a 5, apenas podrá observarse a la escala de la figura (6.18).

Sin embargo la disminución relativa todavía es más apreciable que en el algoritmo 1 para los mismos valores de R . Compárese con la figura (6.10), notando que las escalas no son iguales.

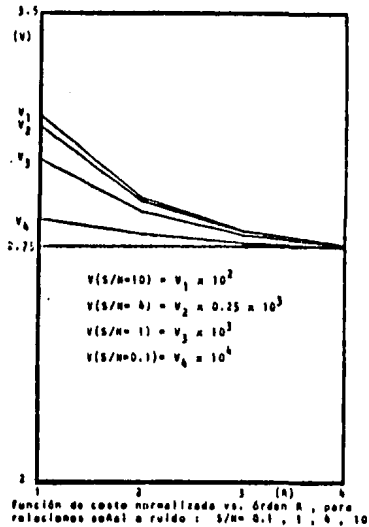


Fig 6.18

Dependencia del nivel de ruido

De acuerdo al teorema (6.1), se espera que para el número de observaciones $N \rightarrow \infty$, se cumplan las ecuaciones (6.24), donde el polinomio $\hat{F}(q^{-1})$ se obtiene minimizando (6.25).

Por supuesto que al extender este resultado para un orden K del polinomio $F(q^{-1})$ distinto a los órdenes de $A(q^{-1})$, $B(q^{-1})$, los infinitésimos $O(1/(S/N))$, $S/N \rightarrow \infty$ dependen de R . En particular se ha observado que los que corresponden a \hat{a}_1 en nuestras simulaciones son decrecientes con R .

Se espera pues unas distancias $\delta a_1 = |\hat{a}_1 - a_1|$ y $\delta b_0 = |\hat{b}_0 - b_0|$, decrecientes para $S/N \rightarrow \infty$ al menos de orden 1 en el infinitésimo $1/(S/N)$, $S/N \rightarrow \infty$.

Las figuras (6.19) y (6.20) muestran estas distancias contra S/N computadas para $\{S/N\} = 0.1, 1, 4, 10$, y para las identificaciones hechas con órdenes $R = 1, 2, 3, 4$.

Para \hat{a}_1 se observa una disminución monótonica con S/N .

Las escalas son logarítmicas, y se ha dibujado en línea punteada una línea de pendiente $1/(S/N)$, que es la que se espera.

La tendencia no alcanza esta pendiente, pero se aproxima cada vez más para valores de (S/N) cada vez mayores.

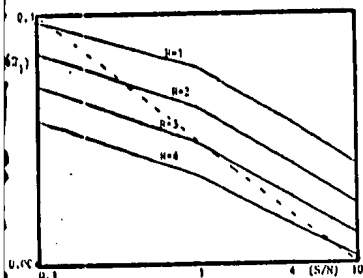
Error de estimación \hat{b}_0 vs. S/N para $R = 1, 2, 3, 4$.

Figura (6.19)

Para la distancia \hat{b}_0 , la tendencia es también decreciente en términos generales; pero la tendencia no alcanza el valor de $1/(S/N)$ señalada por la dirección de la línea punteada.

De todos modos \hat{b}_0 es en general bastante menor que \hat{a}_1 , como ya había sido observado.

Se comprueba entonces que tanto \hat{a}_1 como \hat{b}_0 se aproximan a a_1 y b_0 respectivamente para S/N crecientes.

Estimación del proceso del ruido

De acuerdo al teorema 6.1 para relaciones señal a ruido grandes, la estimación $\hat{F}_1 + \bar{F}$ para S/N $\rightarrow \infty$, donde el polinomio \bar{F} es tal como se da por la minimización de (6.25).

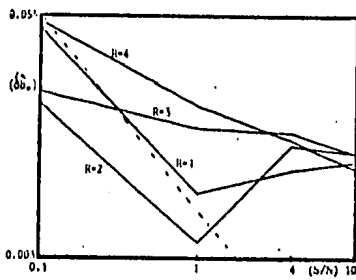
Error de estimación \hat{b}_0 vs. S/N, para órden $R = 1, 2, 3, 4$.

Figura (6.20)

En el caso que estudiamos este equivale a que:

$$V_2(\bar{F}_1 \dots \bar{F}_R) = E(\bar{F}(q^{-1}) \cdot C(q^{-1}) e(t))^2 = \text{mínimo} \quad (6.38)$$

como

$$\bar{F}(q^{-1}) = 1 + \bar{F}_1 q^{-1} + \bar{F}_2 q^{-2} + \dots + \bar{F}_R q^{-R}$$

Supongamos que $C(q^{-1}) = (1 - a q^{-1})$, entonces minimizar (6.38) cuando $e(t)$ es un proceso estacionario caracterizable como ruido blanco, implica minimizar una función cuadrática en las variables \bar{F}_1 , que resulta de la suma de cuadrados de los coeficientes del polinomio producto: $\bar{F}(q^{-1}) C(q^{-1})$.

Entonces:

$$\begin{aligned} -a + (1+a^2) \bar{F}_1 - a \bar{F}_2 &= 0 \\ a \bar{F}_1 + (1+a^2) \bar{F}_2 - a \bar{F}_3 &= 0 \\ a \bar{F}_2 + (1+a^2) \bar{F}_3 - a \bar{F}_4 &= 0 \\ &\vdots \\ a \bar{F}_{R-1} + (1+a^2) \bar{F}_R &= 0 \end{aligned} \quad (6.39)$$

Lo que equivale a resolver el sistema Toeplitz triangular:

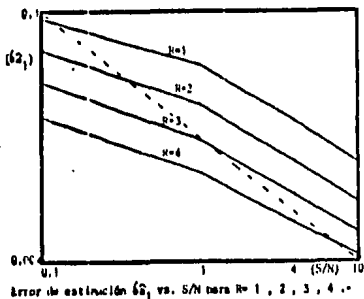


Figura (6.19)

Para la distancia $\hat{\delta}_0$, la tendencia es también decreciente en términos generales; pero la tendencia no alcanza el valor de $\hat{i}/(S/N)$ señalada por la dirección de la línea punteada.

De todos modos $\hat{\delta}_0$ es en general bastante menor que $\hat{\delta}_1$, como ya había sido observado.

Se comprueba entonces que tanto $\hat{\delta}_1$ como $\hat{\delta}_0$ se aproximan a a_1 y b_0 respectivamente para S/N crecientes.

Estimación del proceso del ruido

De acuerdo al teorema 6.1 para relaciones señal a ruido grandes, la estimación $\hat{F}_1 \rightarrow \bar{F}$ para $S/N \rightarrow \infty$, donde el polinomio \bar{F} es dado por la minimización de (6.25).

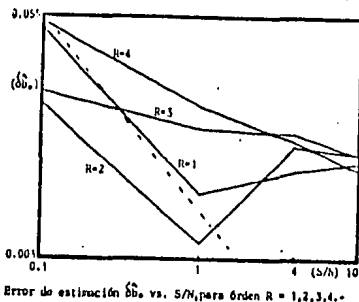


Figura (6.20)

En el caso que estudiamos este equivale a que:

$$V_2 (\bar{F}_1 \dots \bar{F}_R) = E(\bar{F}(q^{-1}) \cdot C(q^{-1}) e(t))^2 = \text{mínimo} \quad (6.30)$$

como

$$\bar{F}(q^{-1}) = 1 + \bar{F}_1 q^{-1} + \bar{F}_2 q^{-2} + \dots + \bar{F}_R q^{-R}$$

Supongamos que $C(q^{-1}) = (1 - a q^{-1})$,

entonces minimizar (6.30) cuando $e(t)$ es un proceso estacionario caracterizable como ruido blanco, implica minimizar una función cuadrática en las variables \bar{F}_1 , que resulta de la suma de cuadrados de los coeficientes del polinomio producto: $\bar{F}(q^{-1}) C(q^{-1})$.

Entonces:

$$\begin{aligned} -a + (1+a^2) \bar{F}_1 - a \bar{F}_2 &= 0 \\ a \bar{F}_1 + (1+a^2) \bar{F}_2 - a \bar{F}_3 &= 0 \\ a \bar{F}_2 + (1+a^2) \bar{F}_3 - a \bar{F}_4 &= 0 \\ &\vdots \\ a \bar{F}_{R-1} + (1+a^2) \bar{F}_R &= 0 \end{aligned} \quad (6.31)$$

Lo que equivale a resolver el sistema Toeplitz triangular:

$$\begin{bmatrix} 1+a^2 & -a & 0 & \dots & 0 & 0 \\ -a & 1+a^2 & -a & \dots & 0 & 0 \\ 0 & -a & 1+a^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -a & 1+a^2 \end{bmatrix} \begin{bmatrix} \hat{F}_1 \\ \hat{F}_2 \\ \hat{F}_3 \\ \vdots \\ \hat{F}_R \end{bmatrix} = \begin{bmatrix} a \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (6.40)$$

en el sistema E que estamos simulando $a = 0.7$; luego, sustituyendo su valor y resolviendo (6.40) para $R = 1, 2, 3, 4$ encontramos los valores teóricos a los que debe converger la estimación \hat{F} del filtro, y que se tabulan a continuación en la Tabla 6.IX.

TABLA 6.IX. Parámetros f_i , $i = 1, \dots, R$ Teóricos.

Parámetro	R=1	2	3	4
f_1	0.4698	0.6029	0.6554	0.6778
f_2	-	0.2832	0.3951	0.4449
f_3	-	-	0.1856	0.2682
f_4	-	-	-	0.1260

De acuerdo al teorema 6.1, la diferencia entre estos valores y los estimados, es, para $S/N \rightarrow \infty$ un infinitésimo de orden al menos 1 en el infinitésimo $1/(S/N)$, $S/N \rightarrow \infty$.

En la siguiente tabla se dan los valores estimados \hat{f}_i para dig

tintos S/N .

TABLA 6.X. Estimaciones \hat{f}_i , $i=1, \dots, R$, para $R = 1, 2, 3, 4$ y $S/N = 0.1, 1, 4, 10$.

R	Parámetro	S/N=0.1	1	4	10
1	\hat{f}_1	0.3265	0.4314	0.4683	0.4784
2	\hat{f}_1	0.5056	0.5910	0.6184	0.6257
	\hat{f}_2	0.1936	0.2632	0.2872	0.2939
3	\hat{f}_1	0.6062	0.6634	0.6811	0.6860
	\hat{f}_2	0.3257	0.3909	0.4121	0.4181
	\hat{f}_3	0.1321	0.1744	0.1884	0.1924
4	\hat{f}_1	0.6652	0.6986	0.7091	0.7123
	\hat{f}_2	0.4108	0.4546	0.4688	0.4731
	\hat{f}_3	0.2254	0.2639	0.2767	0.2805
	\hat{f}_4	0.0924	0.1153	0.1229	0.1252

Efectivamente las estimaciones están bastante cerca de lo que se esperaba. Pero analizando la evolución de las estimaciones para S/N crecientes hemos encontrado que se trata precisamente en las estimaciones \hat{f}_i donde podemos comprender claramente el efecto que tiene la finitud del número de observaciones (que en todos estos casos fue $N = 1000$). Veamos esto.

Aunque los valores tabulados en la Tabla 6.X se hallan próxi-

mos a los valores tabulados en la Tabla 6.IX que son los asintóticos para $S/N \rightarrow \infty$ (y $N \rightarrow \infty$), en todos los casos excepto para $R = 4$ el estimador \hat{f}_4 , se nota el mismo tipo de comportamiento, para S/N creciente \hat{f}_4 se aproxima a su valor asintótico por defecto, pero luego lo supera.

hemos ejemplificado éste graficando en la figura (6.21.a), la variación de \hat{f}_1 correspondiente a $R = 2$ contra S/N . Luego, la diferencia con el valor tabulado en b.IX no se parece en nada a un infinitésimo de orden 1 en S/N . Para el caso del ejemplo, esta diferencia se grafica en la figura (6.21.b).

Frente a este resultado proseguimos las simulaciones para $S/N = 100$, para ver si efectivamente este error disminuye, y, por lo contrario el resultado fue un estancamiento de los estimadores \hat{f}_1 a un valor próximo al obtenido para $S/N = 10$.

Lo que sucede es que el filtro asintótico $\bar{F}(q^{-1})$, para $S/N \rightarrow \infty$ es efectivamente el que resulta de la minimización (6.25):

$$\min(V_2(\hat{f}_1, \dots, \hat{f}_R)) = \min(L[\bar{F}(q^{-1}) \cdot n(t)]^2)$$

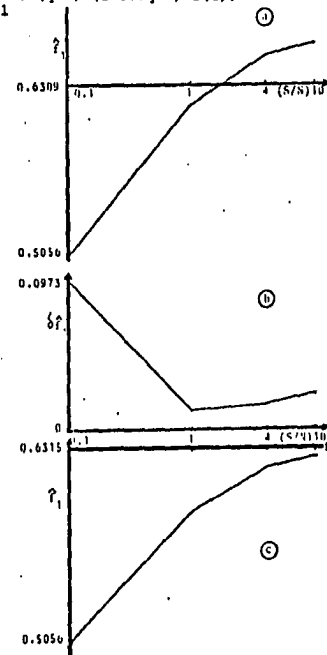
donde $n(t)$ es el proceso de las perturbaciones, pero para un número finito de observaciones, $\bar{F}(q^{-1})$ es el que minimiza:

$$V_4(\hat{f}_1, \dots, \hat{f}_R) = \frac{1}{N} \sum_{t=1}^N |\bar{F}(q^{-1}) n(t)|^2 \quad (6.41)$$

Lo llamaremos $\bar{F}(q^{-1})$.

Con los mismos datos numéricos hemos minimizado

$$\frac{1}{N} \sum_{t=1}^N \{ \bar{F}(q^{-1}) (1-0.7q^{-1}) e(t) \}^2 \quad (6.42)$$



Estimación \hat{f}_4 para $R=2$ vs. S/N y error de estimación.

Figura (6.21)

con el algoritmo de mínimos cuadrados estudiado anteriormente. Los resultados son los de la Tabla 6.XI.

Para el ejemplo de la figura (6.21.d), se ha graficado un (6.21.c) el mismo estimador \hat{f}_1 de $R = 2$ contra S/N , comparándolo con f_1^* obtenido con la minimización (6.42)

Compárense a y c, la convergencia de \hat{f}_1 a f_1^* hace mucho más sentido que a \bar{f}_1 .

TABLA 6.XI. Valores de f_i^* , $i=1...R$, que resultan de minimización (6.42)

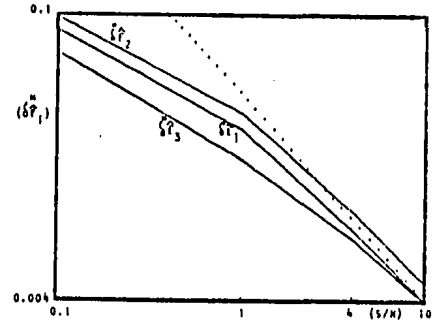
Parámetro	R=1	R=2	R=3	R=4
f_1^*	0.4864	0.6315	0.6900	0.3154
f_2^*	-	0.2989	0.4230	0.4776
f_3^*	-	-	0.1964	0.2857
f_4^*	-	-	-	0.1294

Se ha comprobado el mismo comportamiento en todos los casos de las estimaciones \hat{f}_i de la Tabla 6.X. Estos muestran tendencia a converger al valor f_i^* correspondiente, para S/N creciente.

A modo de ejemplo se grafican en la figura (6.22) los errores $\delta^* \hat{f}_i$

$$\delta^* \hat{f}_i = |f_i^* - \hat{f}_i|$$

para el caso $R = 3$.



Distancia $|f_i^* - \hat{f}_i|$ vs. relación señal a ruido, $i=1,2,3$, $R=3$.

Figura (6.22)

Relación señal a ruido pequeño

Hemos querido verificar con algún ejemplo el resultado principal dado por el teorema 6.3, es decir que para relaciones señal a ruido lo suficientemente grandes, la función $V(\hat{\theta})$ dada por (6.23) tiene más de un mínimo relativo. Esta función es el límite con probabilidad 1 de la función minimizada por el algoritmo 2 de mínimos cuadrados generalizados, para $N \rightarrow \infty$.

Los resultados que se presentan siguen siendo obtenidos para $N = 1000$.

Se hicieron reiteradas identificaciones sobre el sistema E con el algoritmo 2 para $R = 3$ y $S/N = 1, 0.1, 0.01$. Con $S/N = 1$ y $S/N = 0.1$ se encontró solo un mínimo relativo de la función de costo minimizada. En cambio para $S/N = 0.01$ se encontraron dos de estos puntos.

Para ilustrar esta situación, se repitió la identificación con los valores de R y S/N mencionados, pero en lugar de inicializar el algoritmo con una estimación de mínimos cuadrados, se lo inicializó con valores más arbitrarios todavía. Los valores elegidos de forma que los resultados gráficos fueran visualizables son: $(\hat{a}_1, \hat{b}_0)_0 = (-1, 1.5), (-0.6, 1.5), (-0.2, 1.5), (0.2, 1.5), (0.6, 1.5), (-1, 0.7), (-0.5, 0.7), (-0.3, 0.7), (0.6, 0.7), (-1, -0.3), (-0.6, -0.3), (-0.2, -0.3), (0.2, -0.3), (0.6, -0.3)$.

En las figuras (6.23), (6.24), (6.25) se presentan los resultados de las identificaciones con $S/N = 1, 0.1, 0.01$ respectivamente, inicializadas en los puntos mencionados.

Las trayectorias se obtuvieron uniendo con segmentos de recta las estimaciones sucesivas, iteración a iteración. Los puntos iniciales están enmarcados en circulitos, y con cruces se señalan el parámetro real θ y los puntos de convergencia $\hat{\theta}$. Todas las trayectorias convergen a algún $\hat{\theta}$. En las figuras donde hay un solo punto de convergencia, a él arriban; donde hay más de un punto de convergencia, cada trayectoria arriba a uno y

uno sólo de estos puntos.

Las flechas indican la dirección en que se siguieron las trayectorias de los estimadores.

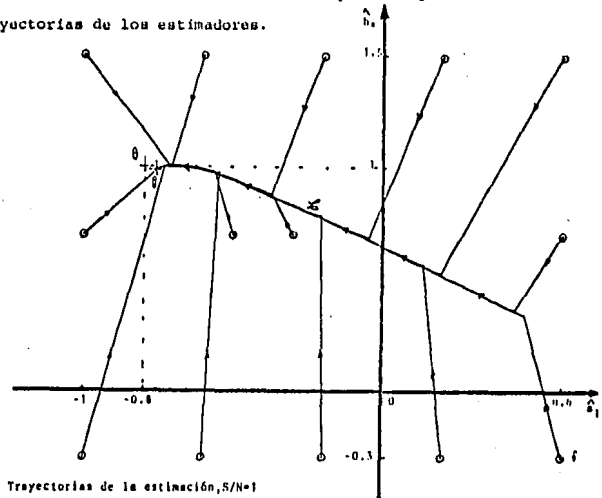
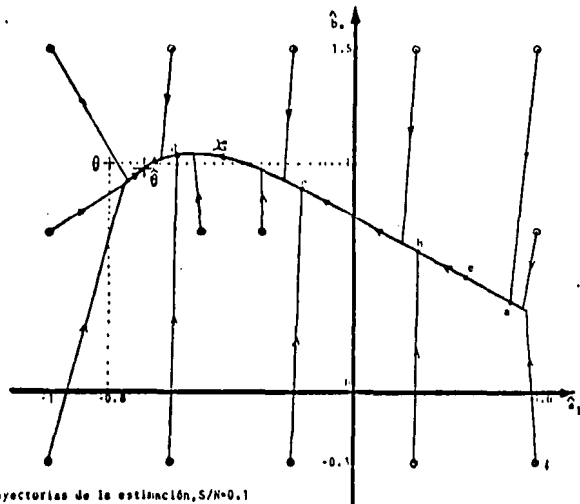
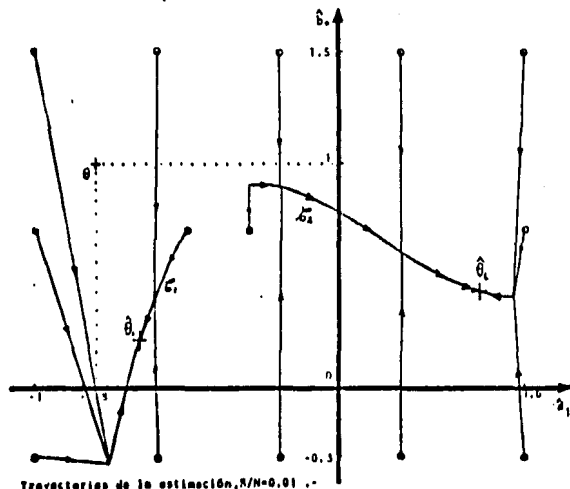


Figura (6.23)

Nótese que luego de la primera iteración, todos los siguientes dan estimadores sobre una curva a la que hemos llamado \mathcal{C} en las figuras (6.23) y (6.24), veremos porqué.

Primero convenimos por simplicidad en trabajar en el caso asintótico ($N \rightarrow \infty$), sino basta con considerar promedios en el tiempo de la forma $\frac{1}{N} \sum_{t=1}^N$, que puede verse que es lo que se

Traectorias de la estimación, $S/N=0.1$ Traectorias de la estimación, $S/N=0.01$

hace en realidad para un número finito de muestras. Si convenimos como antes en llamar $E U^2(t) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N U^2(t)$, y mantenemos las condiciones a) y b) de la sección 6.3.5.:

Llamemos $\theta_1 = (a_1, \dots, a_n, b_0, \dots, b_n)^T$ (no incluye los parámetros f_1), entonces puede verse que la curva $C(\theta_1)$ es la variedad definida por:

$$(\theta_1) = \{ \hat{\theta}_1 | E\{F(q^{-1})[\hat{A}(q^{-1}) y(t) - \hat{B}(q^{-1}) u(t)]^2\} = \text{mínimo} \text{ sobre } \theta_1 \text{ para algún conjunto } f_1, \dots, f_R \} \quad (6.43)$$

Donde hemos convenido en llamar:

$$F(q^{-1}) = 1 + f_1 q^{-1} + \dots + f_R q^{-R}$$

conjunto de valores admisibles para f_1, \dots, f_R

$\hat{A}(q^{-1}), \hat{B}(q^{-1})$, estimaciones de $A(q^{-1}), B(q^{-1})$ con los parámetros de $\hat{\theta}_1$.

Esto es que $C(\theta_1)$ es el conjunto de puntos $\theta_1 (f_1, \dots, f_R, \{y(t)\}, \{u(t)\})$ que minimizan la función de costo de la definición para los puntos f_1, \dots, f_R del conjunto admisible \mathcal{F} , dadas las observaciones $\{u(t)\}, \{y(t)\} t = 1, 2, \dots$

Las trayectorias de la estimación por mínimos cuadrados generalizados se mantienen dentro de C pues el método es de relajación. En efecto, en cada iteración, primero se realiza una

minimización sobre los parámetros f_i (la de la iteración anterior) obteniéndose un filtro, digamos $F(q^{-1}) = 1 + f_1 q^{-1} + \dots + f_N q^{-N}$ con el que se filtran entradas y salidas.

El valor de $\hat{\theta}_1$ que se obtiene en seguida ($\hat{\theta}_1$) es aquel que minimiza la función definida en (6.43) para $F(q^{-1})$ ya dado.

Por lo tanto en cualquier iteración la estimación $\hat{\theta}$ es tal que los primeros parámetros son los del vector $\hat{\theta}_1$ que satisface (6.43) para $(f_1 \dots f_N) = (\hat{f}_1 \dots \hat{f}_N)$ de la estimación anterior.

Supongamos ahora que existe un mínimo relativo de $V(\hat{\theta})$ definida en (6.23). Forzosamente éste debe estar contenido en C :

En efecto, siendo $n(t)$ y $u(t)$ independientes puede verse que:

$$V(\hat{\theta}) = \frac{1}{2} E \left| \hat{F}(q^{-1}) \hat{\Lambda}(q^{-1}) y(t) - \hat{B}(q^{-1}) u(t) \right|^2 \quad (6.45)$$

Supongamos que el mínimo corresponde a:

$$\hat{\theta} = (\hat{a}_1, \dots, \hat{a}_N, \hat{b}_0, \dots, \hat{b}_N, \hat{f}_1, \dots, \hat{f}_N)$$

luego si fijamos $(f_1 \dots f_N) = (\hat{f}_1 \dots \hat{f}_N)$, el punto

$\hat{\theta}_1 = (\hat{a}_1, \dots, \hat{a}_N)$ debe ser un mínimo relativo de la función definida en (6.43), $\hat{\theta}_1 \in C$

Nótese que la única diferencia entre (6.43) y (6.45) es que en (6.45) los parámetros $f_1 \dots f_N$ son libres (dentro de algún compacto) y en (6.43) no.

Desarrollemos otra vez (6.45):

$$V(\hat{\theta}) = \frac{1}{2} E \left| \frac{\hat{F}(q^{-1})}{\hat{\Lambda}(q^{-1})} (\hat{\Lambda}(q^{-1}) B(q^{-1}) - \Lambda(q^{-1}) \hat{B}(q^{-1})) u(t) \right|^2 + \frac{1}{2} E \left| \frac{\hat{\Lambda}(q^{-1}) \hat{F}(q^{-1})}{\hat{\Lambda}(q^{-1})} n(t) \right|^2 \quad (6.46)$$

intuitivamente puede verse que en la minimización de (6.46), hay dos sumandos no independientes pero que tienen un peso relativo que depende de la relación señal a ruido S/N. Cuando $S/N \rightarrow \infty$ la varianza de $n(t) \rightarrow 0$, al mantener $u(t)$. Resulta lógico entonces el resultado del teorema (6.1) que asegura que $\hat{\Lambda}(q^{-1})$ y $\hat{B}(q^{-1})$ son próximos a $\Lambda(q^{-1})$ y $B(q^{-1})$ (en el sentido de la norma de su diferencia), luego en el segundo término $\hat{\Lambda}(q^{-1})/\Lambda(q^{-1}) \rightarrow 1$ ($S/N \rightarrow \infty$) y la optimización respecto $\hat{F}(q^{-1})$ es para $\hat{F}(q^{-1})$ próximo al valor que minimiza $E |\hat{F}(q^{-1}) n(t)|^2$.

Cuando $S/N \rightarrow 0$ el segundo sumando de (6.46) tiene un peso relativo bastante mayor, y precisamente la existencia de mínimos locales para este término fue lo que se definió como "condición de ruido" cuando se trató el problema de relación señal a ruido pequeña.

En las simulaciones realizadas se confirma todo lo anterior.

En primer lugar, al disminuir S/N de 1 a 0.1 la función de costo medida sobre C se va aplanando y la convergencia es más lenta.

Por ejemplo, mientras en la figura 6.23, partiendo desde f , se converge a $\hat{\theta}$ en 22 iteraciones, en las identificaciones de la figura (6.24) sucede que partiendo del mismo punto f se upera apenas el punto e en unas 70 iteraciones.

En las identificaciones de la figura (6.24) para S/N = 0.1 el segmento de C comprendido entre a y b se vuelve particularmente plano. Por ejemplo: mientras pasar desde a hasta b implican más de 120 iteraciones y una reducción de la función de costo de 67.6 (lo que representa un 0.236%), el pasaje de c a d se hace con una baja de la misma función de 638.9 (lo que representa un 2.4%), y desde c se necesitan 38 iteraciones para superar el punto d y alcanzar convergencia en $\hat{\theta}$.

El aplanamiento de C se hace especialmente fuerte en las proximidades del punto señalado como e .

Pero si las diferencias al disminuir S/N de 1 a 0.1 son cuantitativas, al disminuir de 0.1 a 0.01 estas diferencias dan un salto en calidad, aparecen dos mínimos relativos de $W(\hat{\theta})$ sobre C : $\hat{\theta}_1$ y $\hat{\theta}_2$ en la figura (6.25). Estos tienen valores:

$$\hat{\theta}_1 = (-0.6567, 0.2066)^T \quad \hat{\theta}_2 = (0.4583, 0.4230)^T$$

ambos pueden considerarse para un rango muy amplio de aplicaciones como malas estimaciones del parámetro θ .

La curva C aparece dividida en dos curvas: C_1 y C_2 de acuerdo al punto de convergencia al cual converge el algoritmo ($\hat{\theta}_1$ ó $\hat{\theta}_2$), inicializado sobre la curva.

Entre los dos tramos C_1 y C_2 dibujados, hay un segmento de C que no está dibujado y que puede probarse fácilmente debe contener al menos un extremo relativo de $W(\hat{\theta})$, y al menos un máximo relativo sobre C (pudiendo ser el mismo). El máximo relativo en C es un punto de $W(\hat{\theta})$.

Nada asegura sin embargo que no haya más mínimos relativos sobre C .

De los dos puntos de convergencia hallados, ambos corresponden a sistemas discretos realizables, pero $\hat{\theta}_2$ podría ser deseado por ejemplo si se tratara de un sistema muestreado (no puede haber polos sobre el semieje negativo, aunque se encuentren dentro del círculo unitario) de primer orden.

Pueden entonces sugerirse reglas prácticas para la búsqueda de mínimos de $W(\hat{\theta})$ en forma más eficiente:

En primer lugar limitar el conjunto factible de soluciones a partir de consideraciones físicas (realizabilidad, estabilidad, etc.), y construir la variedad C dentro de esta región, por tramos en forma interactiva buscando numéricamente (p.ej. por extrapolación) puntos próximos a C en lugares óptimos en el sentido de minimizar tiempo de máquina.

Esto es aplicable especialmente cuando $S/N \rightarrow \infty$ (valores grandes), estamos buscando un solo extremo relativo y queremos minimizar el número de iteraciones ya que resulta extremadamente costoso para muestras grandes, teniendo claro, que para $S/N \rightarrow \infty$ de todos modos la convergencia es cada vez más rápida.

6.4.4 Algunos resultados comparativos en la identificación del sistema E

En las Tablas 6.I y 6.VIII se señalaron el número de iteraciones empleadas por los algoritmos 1 y 2 respectivamente, de mínimos cuadrados generalizados, para llegar a la convergencia en la identificación del sistema E, con distintos valores de las relaciones señal a ruido simuladas, y orden R en la identificación.

Las condiciones impuestas para detener los algoritmos: (6.34) y (6.35) son muy similares.

Los dos algoritmos emplearon un número semejante de iteraciones.

Para el algoritmo 1 se observa que el número de iteraciones crece mucho al aumentar la potencia del ruido ($S/N \rightarrow \infty$), y, para valores altos de S/N decrece ligeramente al aumentar el orden R, y crece al aumentar R para los valores más bajos simulados de S/N .

Para el algoritmo 2 se ha encontrado que en estas simulaciones el número de iteraciones crece tanto al aumentar R como al disminuir S/N .

Para bajos niveles de ruido, el aumento del número de iteraciones al aumentar R es muy pequeño, en cambio en el nivel de ruido más alto que fue simulado ($S/N = 0.1$), el número de iteraciones es aproximadamente proporcional a R.

La dependencia del nivel de ruido es muy fuerte.

Comparativamente, como hemos dicho los dos algoritmos emplearon números similares de iteraciones para llegar a la convergencia; en casi todos los casos el algoritmo 2 empleó un número ligeramente inferior o igual que el algoritmo uno, excepto para el nivel más alto de ruido y con los valores más altos de $R(S/N = 0.1$ con $R = 3$ y R_4), donde el número de iteraciones

del algoritmo 2 es mayor que el del algoritmo 1.

En cuanto al resultado de la estimación, el parámetro b_0 tiene estimaciones similares con los dos algoritmos para valores iguales de S/N y R. Esto es previsible puesto que ya habíamos visto que los algoritmos no mejoran mucho la estimación de b_0 respecto a mínimos cuadrados, debido a la estructura del modelo. (Referirse por ejemplo a la sección donde se discutió la estimación sesgada del estimador de mínimos cuadrados cuando las perturbaciones están correlacionadas).

En cambio para el estimador \hat{a}_1 se pueden hacer algunas observaciones.

En las identificaciones con ambos algoritmos se ha verificado que en ambos se obtiene una mejor estimación de a_1 al aumentar el orden R puesto que hay una mejor aproximación del proceso de las perturbaciones.

También en ambos la mejora que resulta un \hat{a}_1 de aumentar R en una unidad es menor cuanto más grande sea R.

Los dos algoritmos dan mejores resultados de \hat{a}_1 para mayores valores de S/N.

Pero para iguales valores de S/N y R el algoritmo 1 arroja una

mejor estimación \hat{a}_1 que el algoritmo 2.

Esto se debe a que el algoritmo 2 estima el proceso de las perturbaciones con un filtro de mayor orden que el algoritmo 2.

Esta situación puede comprobarse en las Tablas 6.II y 6.VI que muestran las estimaciones \hat{a}_1 y \hat{b}_0 de ambos algoritmos.

Se encontró sin embargo que hay una mayor proximidad entre las estimaciones \hat{a}_1 obtenidas con un orden R a través del algoritmo 1, y un orden R+1 a través del algoritmo 2.

Puede observarse esto en las tablas mencionadas; para ilustrar este hecho se han dibujado en las figuras (6.26) y (6.27) las estimaciones \hat{a}_1 con los algoritmos 1 y 2 vs. R para dos relaciones señal a ruido particulares: S/N = 10 y 1.

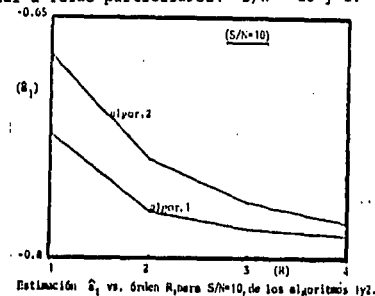


Figura (6.26)

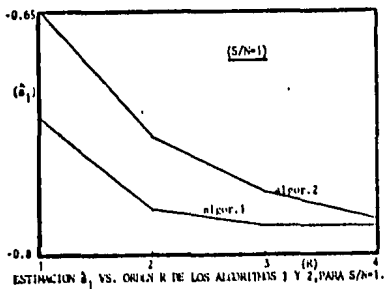


Figura (6.27)

En la figura (6.28) se muestra la estimación de a_1 y b_0 con ambos algoritmos contra el número de iteraciones para $S/N = 10$ y $R = 2$.

Ambos algoritmos comienzan con la misma estimación inicial (la de mínimos cuadrados), y luego se itera hasta la convergencia.

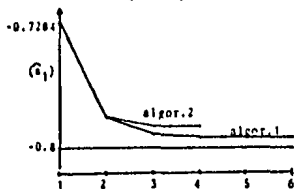
Estimaciones a_1, b_0 vs. número de iteraciones.

Figura (6.28)

Esta simulación muestra el comportamiento que fué típico en las identificaciones hechas sobre el sistema E:

- mientras para \hat{a}_1 se obtienen resultados distintos, para \hat{b}_0 son semejantes (en este caso demasiado).
- el algoritmo 2 llega más rápido a la convergencia, pero a un valor a_1 más alejado del real \hat{a}_1 , que el algoritmo 1.

6.4.5 Identificación del sistema F con el algoritmo 2

Con las señales obtenidas de la simulación del sistema F, que se describen en el párrafo 6.4.1 se identificó dicho sistema con el algoritmo 2, con $R = 1$, $N = 1$, $M = 0$ y retraso $k = 1$.

En la Tabla 6.XII se muestran los valores de los estimadores \hat{a}_1 y \hat{b}_0 , \hat{f}_1 , el número de iteraciones i , y el valor de la función de costo $W(\hat{\theta})$, para cada una de las relaciones señal a ruido consideradas: $S/N = 10, 4, 1, 0.1$.

TABLA 6.XII. Resultados de la identificación del sistema F con el algoritmo 2, $R = 1$, $S/N = 10, 4, 1, 0.1$

S/N	\hat{a}_1	\hat{b}_0	\hat{f}_1	i	$W(\hat{\theta})$
10	-0.7977	0.9991	-0.2865	3	58.8562
4	-0.7979	0.999	-0.2862	4	142.154
1	-0.8017	0.9999	-0.2813	5	568.647
0.1	-0.8179	1.0026	-0.2611	6	5683.7

A simple vista puede apreciarse la enorme proximidad de las estimaciones a sus valores reales (recordemos que $a_1 = -0.8$, $b_0 = 1$, $f_1 = -0.3$).

Deben considerarse que el sistema F tiene la estructura exacta para el algoritmo 2: orden finito y perturbaciones que resultan de una autoregresión o proceso AR, y además hemos supuesto conocido el orden exacto del sistema.

Para infinitas muestras, esperaríamos, de acuerdo al teorema 6.1 que para $S/N \rightarrow \infty$ los estimadores \hat{a}_1 , \hat{b}_0 , \hat{f}_1 convergen a sus valores reales.

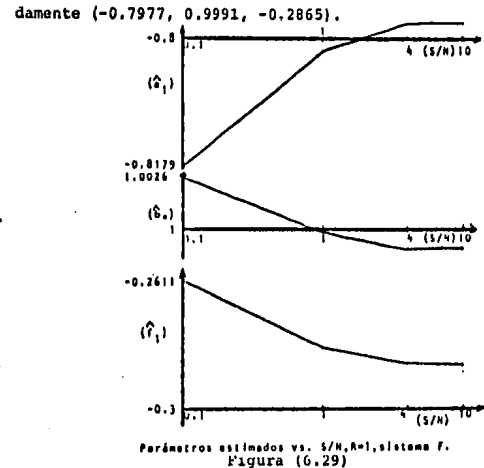
Para un número finito de muestras hemos encontrado convergencia a una proximidad de los valores reales; este punto de convergencia, a su vez, tiende a los valores reales para $N \rightarrow \infty$.

En este caso también vuelve a verificarse esta afirmación.

En la figura (6.29) se dibujan las estimaciones \hat{a}_1 , \hat{b}_0 , \hat{f}_1 contra la relación señal a ruido.

Puede verse que para S/N crecientes dichos estimadores tienden a converger a un valor cercano al parámetro real.

Para $\hat{\theta}^T = (\hat{a}_1, \hat{b}_0, \hat{f}_1)$ este valor de convergencia es aproximadamente



Para este sistema en particular, con estos datos simulados, la proximidad del sistema identificado respecto al sistema vectorial (desde el punto de vista paramétrico), es enorme. Nótese que aún para la relación señal a ruido más baja de las cuatro consideradas, los parámetros están próximos a sus valores reales.

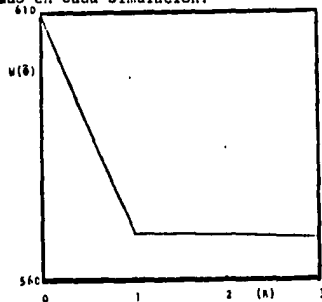
Para ejemplificar cómo se podría atacar el problema de la determinación del orden R cuando este es desconocido hemos realizado la identificación del sistema F con el mismo algoritmo para $S/N = 1$ pero variando R : $R = 0, 1, 2, 3$.

La identificación para $R = 0$ es la de mínimos cuadrados ordinarios, que así puede considerarse un caso sumamente particular de mínimos cuadrados generalizados. El valor obtenido para la función de costo de la identificación en cada uno de los cuatro casos se muestra en la Tabla 6.XIII.

TABLA 6.XIII

R	$\hat{W}(\hat{\theta})$
0	609.398
1	568.647
2	568.264
3	567.528

En la figura (6.30) se grafican estas cantidades contra el orden R empleado en cada simulación.



Función de costo de la identificación vs. R
Sistema F , 1000 observaciones, $S/N=1$.

En esta figura puede apreciarse la enorme disminución de $\hat{W}(\hat{\theta})$

al pasar de $R = 0$ a $R = 1$, que representa una diferencia de valores de $\hat{W}(\hat{\theta})$ de 40.751 frente a 0.383 y 0.736 que son las disminuciones de $\hat{W}(\hat{\theta})$ al pasar de $R = 1$ a $R = 2$, y de $R = 2$ a $R = 3$ respectivamente.

Esto sugiere extender los criterios vistos para mínimos cuadrados ordinarios a mínimos cuadrados generalizados.

En el caso representado en la figura resulta bastante obvio el resultado $R = 1$, pero para datos no simulados en general no es tan fácil.

7. MÍNIMOS CUADRADOS GENERALIZADOS RECURSIVOS

7.1 El método

Basados en el método de mínimos cuadrados generalizados propuesto por Clarke, Ljung-Jens y Sage (1969), (28) presentaron un algoritmo recursivo que aproxima el método visto en el capítulo anterior. Para ello se propone realizar alternadamente la estimación de la dinámica de la parte determinística y el proceso de las perturbaciones, recursivamente, conforme se recibe cada nueva pareja de datos, en lugar de hacerlo con todos los datos en cada iteración.

Se aprovecha para eso el esquema recursivo de mínimos cuadrados.

Supongamos otra vez el sistema de la figura (6.3), regido por la ecuación

$$A(q^{-1})y(t) = B(q^{-1})u(t) + n(t) \quad (7.1)$$

donde se hace la aproximación:

$$n(t) = \frac{1}{\Gamma(q^{-1})} e(t) \quad (7.2)$$

$$\text{con: } \begin{cases} e(t) \text{ ruido blanco de media cero} \\ A(q^{-1}) = 1 + a_1 q^{-1} + \dots + a_n q^{-n} \\ B(q^{-1}) = b_0 + b_1 q^{-1} + \dots + b_m q^{-m} \\ \Gamma(q^{-1}) = 1 + f_1 q^{-1} + \dots + f_r q^{-r} \end{cases}$$

Supongamos que hay datos (entradas y salidas del sistema) $u(t)$, $y(t)$ hasta $t=k$; entonces llamando

$$\begin{aligned} u^k(k) &= u(k) + \sum_{j=1}^r f_j u(k-j) \\ y^k(k) &= y(k) + \sum_{j=1}^m b_j y(k-j) \end{aligned} \quad (7.3)$$

$$x_k^{\Delta T} = (-y^k(k-1), \dots, -y^k(k-n), u^k(k), \dots, u^k(k-m)) \quad (7.4)$$

$$\hat{\theta}_1^{\Delta T} = (a_1, \dots, a_n, b_0, b_1, \dots, b_m) \quad (7.5)$$

$$x_k^{\Delta} = \begin{bmatrix} x_1^{\Delta T} \\ x_2^{\Delta T} \\ \vdots \\ x_k^{\Delta T} \end{bmatrix}; \quad Y_k^{\Delta} = \begin{bmatrix} y(1)^{\Delta} \\ y(2)^{\Delta} \\ \vdots \\ y(k)^{\Delta} \end{bmatrix} \quad (7.6)$$

Se cumple entonces:

$$Y_k^{\Delta} = X_k^{\Delta} \hat{\theta}_1 + e_k \quad (7.7)$$

$$\text{donde } e_k = (e(1), \dots, e(k))^T \quad (7.8)$$

es de elementos no correlacionados.

Llamando ahora:

$$c_k^T = (-n(k-1), \dots, -n(k-r)) \quad (7.9)$$

$$O_k^T = (f_1, f_2, \dots, f_r) \quad (7.10)$$

$$\Gamma_k = \begin{bmatrix} c_1^T \\ c_2^T \\ \vdots \\ c_k^T \end{bmatrix}; \quad H_k = \begin{bmatrix} n(1) \\ n(2) \\ \vdots \\ n(k) \end{bmatrix} \quad (7.11)$$

se cumple entonces:

$$N_k = E_k O_2 + e_k \quad (7.12)$$

para (7.12) no es más que la ecuación (7.2) puesto en forma matricial para $t = 1, 2, \dots, k$.

Ahora bien, conociendo X_k^A, Y_k^A, N_k, E_k se puede estimar O_1 y O_2 por mínimos cuadrados:

$$\begin{cases} \hat{O}_1(k) = (X_k^{AT} X_k^A)^{-1} X_k^{AT} Y_k^A \\ \hat{O}_2(k) = (L_k^T L_k)^{-1} L_k^T N_k \end{cases} \quad (7.13)$$

Pensando ahora en un esquema recursivo, cuando arriban los nuevos datos: $u(k+1)$, $y(k+1)$, $n(k+1)$, se puede actualizar la estimación de O_1 y O_2 con el algoritmo de Plackett visto en el capítulo 5:

$$\begin{cases} \hat{O}_1(k+1) = \hat{O}_1(k) - \frac{P_k X_{k+1}^{AT} (X_{k+1}^{AT} \hat{O}_1(k) - Y_{k+1}^A)}{1 + X_{k+1}^{AT} P_k X_{k+1}^A} \\ P_{k+1} = P_k - \frac{P_k X_{k+1}^{AT} P_k X_{k+1}^A}{1 + X_{k+1}^{AT} P_k X_{k+1}^A} \end{cases} \quad (\text{donde } P_k = (X_k^{AT} X_k^A)^{-1}) \quad (7.14)$$

$$\begin{cases} \hat{O}_2(k+1) = \hat{O}_2(k) - \frac{O_k E_{k+1}^T (E_{k+1} \hat{O}_2(k) - n(k+1))}{1 + E_{k+1}^T O_k E_{k+1}} \\ O_{k+1} = O_k - \frac{O_k E_{k+1}^T E_{k+1} O_k}{1 + E_{k+1}^T O_k E_{k+1}} \end{cases} \quad (\text{donde } O_k = (L_k^T L_k)^{-1}) \quad (7.15)$$

La observación que hay que hacer aquí es obvia:

Para estimar O_1 hubo que conocer X_k^A, Y_k^A lo que implica haber podido conocer las entradas y salidas filtradas: $u^A(i)$, $y^A(i)$. Para esto, conocidas las entradas y salidas $u(i)$, $y(i)$ habría que procesar el filtrado (7.3). Esto es que para estimar O_1 supusimos conocido O_2 .

Análogamente, para estimar O_2 necesitamos conocer N_k y H_k , lo que de acuerdo a (7.9) y (7.11) implica conocer $n(i)$, el proceso de las perturbaciones. Para conocer el proceso de las perturbaciones a partir del conocimiento de entradas y salidas es preciso conocer los parámetros $a_1, \dots, a_n, b_1, \dots, b_m$ del sistema. En resumen: para estimar O_2 supusimos conocida O_1 .

Nuestro problema de identificación lo hemos puesto en términos de estimar los parámetros θ_1 y θ_2 , desconocidos ambos. Lo que se propone entonces es aplicar el procedimiento anterior, modificándolo.

Para estimar θ_1 , en lugar de utilizar θ_2 , utilizar el estimado actual de este parámetro: $\hat{\theta}_2$.

Para estimar θ_2 , en lugar de utilizar θ_1 , utilizar el estimado actual $\hat{\theta}_1$.

Esto se hace a través de las siguientes modificaciones: aproximaremos u^* , y^* , n como sigue:

$$\begin{cases} \hat{u}^*(k+1) = u(k+1) + \sum_{j=1}^r \hat{f}_j u(k+1-j) \\ \hat{y}^*(k+1) = y(k+1) + \sum_{j=1}^r \hat{f}_j y(k+1-j) \end{cases} \quad \text{con } \hat{f}_j \text{ estimado en } t=k \quad \dots (7.16)$$

$$\hat{n}(k+1) = \hat{A}(q^{-1}) y(k+1) - \hat{B}(q^{-1}) u(k+1) = y(k+1) - \pi_{k+1}^T \hat{\theta}_1(k+1) \quad \dots (7.17)$$

donde el estimado $\hat{\theta}_1(k+1)$ es el obtenido en $t = k+1$.

Se sustituyen estos valores $\hat{u}^*(k+1)$, $\hat{y}^*(k+1)$, $\hat{n}(k+1)$ por $u^*(k+1)$, $y^*(k+1)$, $n(k+1)$ respectivamente en las ecuaciones anteriores.

Debe señalarse además, que como la estimación es recursiva, las señales u y y no son filtradas por los mismos parámetros para obtener u^* , y^* .

Además la estimación de las perturbaciones se hace con polinomios $\hat{A}(q^{-1})$, $\hat{B}(q^{-1})$ variables en el tiempo. Es necesario entonces que a lo

largo de las recurrencias, los últimos datos pesen más que las anteriores en las ecuaciones (7.14) y (7.15).

Para eso se puede por ejemplo, utilizar el esquema visto en el capítulo 5, de olvido exponencial, con algún coeficiente de olvido próximo a 1.

El método queda como se ha representado en la figura (7.1), como dos sistemas no lineales, acoplados a través de dos filtrados variables en el tiempo.

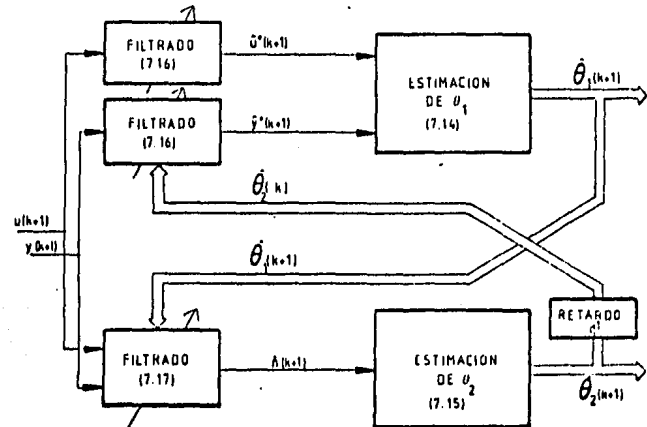


Fig 7.1

Algoritmo 3

Reescribamos explícitamente el algoritmo, con todas las modificaciones de las ecuaciones anteriores:

$$\hat{u}^a(k+1) = u(k+1) + \sum_{j=1}^m \hat{f}_j(k) u(k+1-j) \quad (7.18)$$

$$\hat{y}^a(k+1) = y(k+1) + \sum_{j=1}^m \hat{f}_j(k) y(k+1-j) \quad (7.19)$$

$$\hat{O}_1(k+1) = \hat{O}_1(k) - \frac{P_k x_{k+1}^a (\sum_{j=1}^m \hat{O}_1(k) - y^a(k+1))}{\alpha + x_{k+1}^a P_k x_{k+1}^a} \quad (7.20)$$

$$f(k+1) = \frac{1}{\alpha} \left[P_k - \frac{P_k x_{k+1}^a \sum_{j=1}^m P_k}{\alpha + x_{k+1}^a P_k x_{k+1}^a} \right] \quad (7.21)$$

$$\hat{n}(k+1) = y(k+1) - x_{k+1}^T \hat{O}_1(k+1) \quad (7.22)$$

dado

$$\begin{cases} x_{k+1}^T = (-v(k), \dots, y(k-1), \dots, y(k-n+1), u(k+1), \dots, u(k+1-m)) \\ x_{k+1}^a = (-\hat{y}^a(k), \dots, -\hat{y}^a(k-n+1), \hat{u}^a(k+1), \dots, \hat{u}^a(k+1-m)) \end{cases} \quad (7.23)$$

$$\hat{O}_1(k+1) = \hat{O}_1(k) - \frac{O_k c_{k+1} (\sum_{j=1}^m \hat{O}_2(k) - \hat{n}(k+1))}{B + c_{k+1}^T O_k c_{k+1}} \quad (7.24)$$

$$O_{k+1} = \frac{1}{B} \left[O_k - \frac{O_k c_{k+1} c_{k+1}^T O_k}{B + c_{k+1}^T O_k c_{k+1}} \right] \quad (7.25)$$

$$c_{k+1}^T = (-\hat{n}(k), \dots, -\hat{n}(k-r)) \quad (7.26)$$

en la recursión para $t = k+1$:

- 1) Calcular $\hat{u}^a(k+1)$, $\hat{y}^a(k+1)$ con (7.18) y (7.19)
- 2) Actualizar con $\hat{u}^a(k+1)$, $\hat{y}^a(k)$ el vector x_k^{aT} para obtener x_{k+1}^{aT} como (7.23)
- 3) Calcular $\hat{O}_1(k+1)$ con (7.20) y actualizar P_k para obtener P_{k+1} con (7.21)
- 4) Actualizar el vector x_k^T para obtener x_{k+1}^T como en (7.23)
- 5) Calcular $\hat{n}(k+1)$ con (7.22)
- 6) Actualizar el vector c_k para tener c_{k+1} como en (7.26)
- 7) Calcular $\hat{O}_2(k+1)$ con (7.24) y actualizar O_k para obtener O_{k+1} con (7.25)

inicializar con:

$$\hat{O}_1(0) = 0; \hat{O}_2(0) = 0; x_0 = 0; x_0^a = 0; \varepsilon_0 = 0; P_0 = KI; O_0 = KI$$

con $k \gg 1$. En las simulaciones hemos usado distintos valores de k con $100 \leq k \leq 10000$.

Observaciones sobre α y β

Como se ha discutido en el Capítulo 5, los valores de α y β ($0 < \alpha \leq 1$, $0 < \beta \leq 1$) influyen en la capacidad de las estimaciones de O_1 y O_2 , de seguir las variaciones de los parámetros en el tiempo.

Si se va a trabajar con sistemas variables en el tiempo (lentamente variables o cuasiestáticos) conviene elegir α y β apropiados para las variaciones de θ_1 y θ_2 , o mejor dicho apropiados para las velocidades de variación de θ_1 y θ_2 . Ver para esto las secciones 5.8.1 y 5.12.7.

De todos modos hay que fijar un valor máximo para α y β de acuerdo al transitorio inicial en las estimaciones $\hat{\theta}_1$ y $\hat{\theta}_2$ que se pueden admitir. Una forma más eficiente de reducir el transitorio es utilizar el esquema de factor de olvido variable, ya discutido en la sección 5.8.1 y que consiste en sustituir α en (7.21) y (7.22) y β en (7.24), (7.25) por

$$\begin{cases} \alpha_{k+1} = \alpha_{(0)} \alpha_k + (1 - \alpha_{(0)}) \\ \beta_{k+1} = \beta_{(0)} \beta_k + (1 - \beta_{(0)}) \end{cases} \quad (7.27)$$

con valores iniciales α_0 , β_0 y velocidades $\alpha_{(0)}$, $\beta_{(0)}$ elegidas apropiadamente.

Está claro que pueden haber también esquemas intermedios entre ambas soluciones, que disminuyan el transitorio y que permitan el trabajo con sistemas lentamente variables en el tiempo, modificando (7.27) de manera que

$$\begin{cases} \alpha_{k+1} = \alpha' \neq 1 \\ \beta_{k+1} = \beta' \neq 1 \end{cases} \quad \text{para } k \rightarrow \infty$$

por ejemplo:

$$\begin{cases} \alpha'_{k+1} = \alpha' \alpha_{k+1} \\ \beta'_{k+1} = \beta' \beta_{k+1} \end{cases} \quad (7.28)$$

con α_{k+1} , β_{k+1} dados por (7.27)

Para las simulaciones que veremos más adelante se ha utilizado $\alpha = \beta = 0.99$

Algoritmo 4

Consideremos el algoritmo 3, y observemos el segundo paso. Se actualiza el vector x_k^{dT} para obtener el vector x_{k+1}^{dT} :

$$x_{k+1}^{dT} = (-y^{\wedge}(k), \dots, -y^{\wedge}(k-n+1), u^{\wedge}(k+1), \dots, u^{\wedge}(k+1-m))$$

los elementos con índice i , han sido calculados de (7.18) y (7.19) con los coeficientes $f_j(i-1)$, $j=1, \dots, r$, es decir con la estimación $\hat{\theta}_1(i-1)$.

Pero si suponemos que $\hat{\theta}_2(k)$ es una mejor estimación de θ_2 que $\hat{\theta}_2(k-1)$, por qué no calcular el vector x_{k+1}^{dT} con:

$$\begin{aligned} u^{\wedge}(i) &= \hat{u}(i) + \sum_{j=1}^r \hat{f}_j(k) u(i-j), \quad i = k+1-n, \dots, k+1 \\ y^{\wedge}(i) &= \hat{y}(i) + \sum_{j=1}^r \hat{f}_j(k) u(i-j), \quad i = k+1-n, \dots, k \end{aligned} \quad (7.29)$$

es decir recalculando (ó reestimando) todos los elementos de x_{k+1}^{dT} en base a la última estimación de θ_2 : $\hat{\theta}_2(k)$.

Se propone entonces el mismo algoritmo, excepto en el segundo paso, donde se reestima todo el vector x_{k+1}^* en lugar de agregarle sólo dos nuevos estimados ($\hat{y}^*(k)$, $\hat{u}^*(k+1)$).

Las propiedades de convergencia no se alteran, pero puede verse, que en muchos casos se reduce enormemente el transitorio inicial (ver p. ej: (47)).

La desventaja es que en cada recursión agregamos más filtrados, lo que implica mayor tiempo de procesamiento. Como siempre la elección de uno u otro algoritmo implica un compromiso entre sus ventajas y desventajas.

La convergencia de ambos algoritmos la discutiremos más adelante en el mismo capítulo.

Número de operaciones

Considerando que la inicialización es arbitraria, contabilicemos el nº de sumas y productos necesarios para llevar a cabo una recursión de los algoritmos 3 y 4:

Algoritmo 3

Cálculo	# de sumas	# de productos
1) Filtrado u,y	2R	2R
2) Estimación $\hat{\lambda}$, $\hat{\beta}$	$\frac{1}{2} [3p^2 + 7p]$	$2p^2 + 5p$
3) Estimación \hat{n}	p	p
4) Estimación $\hat{\Gamma}$	$\frac{1}{2} [3R^2 + 7R]$	$2R^2 + 5R$
Total (por recursión)	$1.5p^2 + 4.5p + 1.5R^2 + 5.5p$	$2p^2 + 6p + 2R^2 + 7R$

donde hemos puesto $p = M + N + 1$.

Algoritmo 4

Cálculo	# de sumas	# de productos
1) Filtrado u,y	pR	pR
2) Estimación $\hat{\lambda}$, $\hat{\beta}$	$\frac{1}{2} [3p^2 + 7p]$	$2p^2 + 5p$
3) Estimación \hat{n}	p	p
4) Estimación $\hat{\Gamma}$	$\frac{1}{2} [3R^2 + 7R]$	$2R^2 + 5R$
Total (por recursión)	$1.5p^2 + 4.5p + 1.5R^2 + 3.5R + pR$	$2p^2 + 6p + 2R^2 + 5R + pR$

La diferencia entre ambos es, como se ha mencionado, el filtrado para obtener u^h , y^h . En el algoritmo 3 esto implica 2R sumas y productos por recursión, y en el algoritmo 4 implica pR sumas y productos por recursión.

7.2 Implantación de los algoritmos 3 y 4

Los algoritmos 3 y 4, descritos en las páginas anteriores han sido implantados en la computadora B6800 del C.S.C. de la U.N.A.M. a través de los programas MICO/GEM3 y MICO/GEM4 respectivamente.

Estos programas fueron escritos en Pascal y su listado puede hallarse en el Apéndice D.

El diagrama de flujo básico de ambos programas es el de la figura 7.2. La única diferencia entre ambos es la sección donde se realiza el filtrado para producir u^h y y^h como se ha descrito en las páginas anteriores.

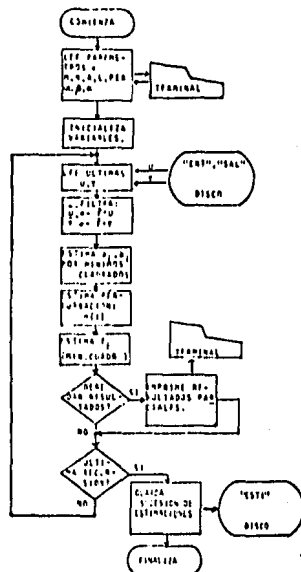


DIAGRAMA DE FLUJO DE MICU/GENE1, MICU/GENE2 ..

Fig 7.2

A diferencia de MICU/GENE1 y MICU/GENE2 en estos dos programas las entradas y salidas se leen de una pareja en cada recursión, directamente de los archivos ent y sal de disco.

La sucesión de estimaciones de a_j , $j=1..M$; b_j , $j=0..M$; f_k , $k=1..R$ para $i=1..L$ se guardan en archivo "ESTI" en disco, al final de la ejecución de ambos programas.

7.3 Identificación del sistema E

En esta sección se describen algunos resultados obtenidos con los algoritmos 3 y 4 sobre las señales del sistema E.

En el capítulo anterior se describió la simulación de un sistema de primer orden al que se denominó E.

Las señales obtenidas que se detallan en la sección 6.4.1, simulan relaciones señal a ruido $S/N = 10, 4, 1, 0.1$. Con dichas señales se procedió a identificar este sistema, utilizando para tal fin distintos valores del orden R. En todos los casos se utilizaron los parámetros reales del sistema: retraso $k=1$, orden $N=1$, orden $M=0$.

Influencia del orden R

Hemos encontrado en las simulaciones una mejora en el resultado de la estimación, al aumentar el orden R, para relaciones señal a ruido pequeñas. Sin embargo hemos notado una influencia fuertemente negativa del incremento de R sobre el transitorio inicial de los algoritmos. Vamos a estos hechos con algunos ejemplos.

En las figuras (7.3) y (7.4) se muestra el resultado de la estimación de a_j y b_j del sistema E para una misma relación señal a ruido $S/N = 1$ con el algoritmo 4 y $R = 1$ y $R = 3$ respectivamente.

Comparemos ambas figuras con la figura (5.65) que representa la identificación del mismo sistema, con el procedimiento de mínimos cuadrados

do; (lo que equivale a usar mínimos cuadrados generalizados con $R=0$).

Puede observarse que hacer $R=1$ (figura (7.3)) disminuye sensiblemente el sesgo de \hat{a}_1 sin empeorar mucho la estimación \hat{b}_0 . Debe tenerse en cuenta que estamos usando un factor de olvido $\alpha=\beta=0.99$, lo que junto con la estimación variable de $r(q^{-1})$ explican las oscilaciones de

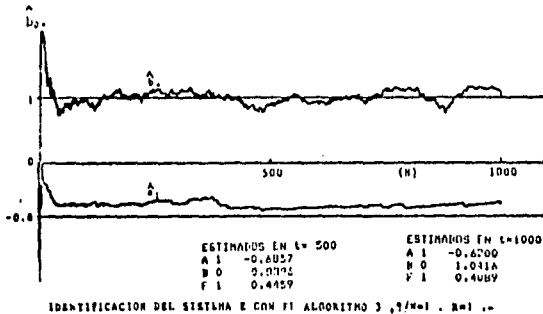


Fig 7.3

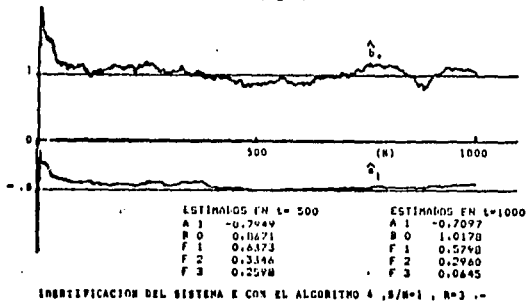


Fig 7.4

Aumentar R de 1 a 3 mejora ligeramente la estimación \hat{b}_0 y reduce bastante más el sesgo de la estimación \hat{a}_1 .

Esto parece deberse a que una mejor estimación del proceso de ruido, mejora la estimación del parámetro a_1 .

En la figura 7.5 se presentan las estimaciones de a_1 y b_0 contra el orden de R , obtenidas por los algoritmos 3 y 4. El número de muestras considerado fué 1000. El valor de la relación señal a ruido se mantuvo constante e igual a 10.

En la gráfica, el superíndice $j=3,4$ indica que se trata de la estimación obtenida con el algoritmo 3 o 4 respectivamente.

El nivel cero representa el valor real del parámetro ($a_1 = -0.8$, $b_0 = 1$), por lo que se desprende que en realidad se graficaron los alcanceamientos de los estimados respecto su valor real.

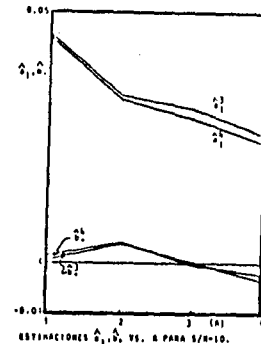


Fig 7.5

Puede verse que la estimación de b_0 está bastante más próxima al valor real que la estimación de a_1 . Ambos algoritmos dan resultados muy parecidos, especialmente para b_0 . La estimación \hat{a}_1 con el algoritmo 4 es ligeramente mejor que la que da el algoritmo 3.

Para b_1 , al considerar R creciendo de 1 a 4 no se obtiene una mejora, el valor de b_1 se mantiene dentro de un entorno de b_0 .

En cambio las estimaciones \hat{a}_1^3 y \hat{a}_1^4 del parámetro a_1 mejoran para valores crecientes de R .

Para valores más altos de la potencia de ruido hemos comprobado que el aumento de R no mejora la estimación.

Para ilustrar esto, en la figura (7.6) se presentan las estimaciones de a_1 y b_0 en condiciones análogas a las de la figura (7.5) pero para una relación señal a ruido $S/N = 1$. El número de muestras sigue siendo 1000.

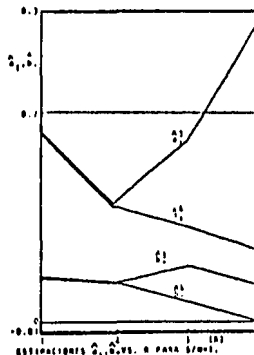


Fig. 7.6

En la figura (7.6) puede verse que aumentar R de 1 a 2 rondada en una mejora de las estimaciones \hat{a}_1 y \hat{b}_0 obtenidas con ambos algoritmos; para $R = 1$ y $R = 2$ las estimaciones que dan los algoritmos 3 y 4 son casi exactamente iguales.

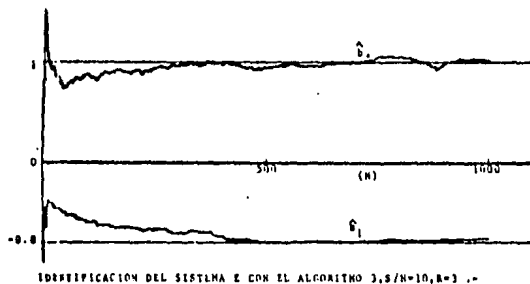
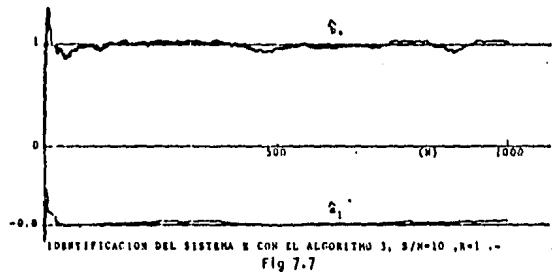
Pero para mayores valores de R (3 y 4), mientras que se mejoran sensiblemente las estimaciones \hat{a}_1 y \hat{b}_0 del algoritmo 4, las del algoritmo 3 empeoran.

Observando las gráficas correspondientes a 40 estimaciones simuladas, se notó la tendencia a aumentar el transitorio inicial del algoritmo, al aumentar el orden R .

El algoritmo 4 se mostró menos sensible a este transitorio. Incluso en simulaciones correspondientes a la figura (7.5), (donde se comprueba una mejora de la identificación para mayores valores de R), se notó un aumento del transitorio inicial para mayores valores de R .

En las figuras (7.7) y (7.8) se muestra un ejemplo: la identificación recursiva del sistema E , con una relación señal a ruido $S/N = 10$, con el algoritmo 3. En la figura (7.7) se presenta la identificación con $R=1$, y en la (7.8) se presenta la identificación con $R=3$.

La estimación final es mejor para $R=3$ que para $R=1$. En general para un número de muestras $N > 500$ la estimación con $R=3$ es mejor que con $R=1$. Pero para $N < 500$ el transitorio inicial es enorme con $R=3$, como puede apreciarse en la figura (7.8).



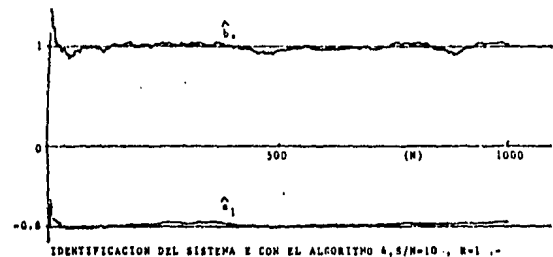
En las mismas condiciones, para el algoritmo 4, la identificación se ve apenas ligeramente deteriorada para $N < 100$ con $R = 3$, como apenas puede distinguirse en la comparación de las figura (7.9) y (7.10).

T. Söderström et al. (1978) (48), habían notado este fenómeno, indicando respecto a las diferencias entre el algoritmo 3 y el algoritmo 4, que éstas son menores y no tienen influencia en el comportamiento asintótico ($N \rightarrow \infty$), y en el análisis de convergencia de ambos algoritmos. Es decir que puede analizarse su convergencia por igual. "Sin

embargo" indican - (estas menores diferencias)... "pueden tener una influencia importante en el comportamiento transitorio del algoritmo".

Más adelante, refiriéndose a sus resultados obtenidos mediante simulaciones, los mismos autores señalan que para altos órdenes R han encontrado una tendencia creciente de la función de costo; indicando que nada puede ser ganado aumentando disminuido el orden R .

Para disminuir el efecto del transitorio ellos utilizan el factor de olvido recursivo que ya hemos discutido, y sugieren que las primeras recursiones (ellos lo hicieron en las primeras 50), se hagan imponiendo $\Gamma(q^{-1}) = 1$, es decir sin filtrar entradas y salidas.



En resumen, el aumento de k trae en general un beneficio en la estimación de los parámetros para $N \rightarrow \infty$, pero para un número finito de muestras deben tomarse precauciones para evitar el transitorio inicial, que se traduce en la convergencia más suave hacia un mejor estimador de los parámetros (para mayores valores de R), pero con una menor velocidad de convergencia.

Este fenómeno se ha encontrado en forma más acentuada para menores valores de la relación señal a ruido, y encuentra su explicación en la forma de las trayectorias en el plano de fase de una ecuación diferencial ordinaria asociada al algoritmo, que se discutirá en este capítulo junto con la convergencia de otros algoritmos recursivos.

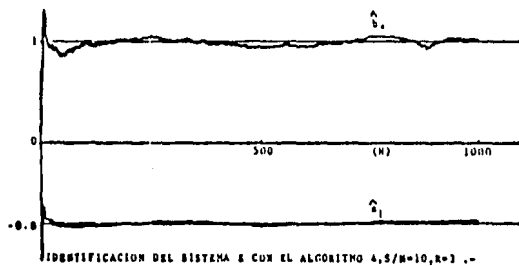


Fig 7.10

Influencia de la potencia de las perturbaciones

Como era previsible se ha encontrado que se obtienen mejores estimadores de los parámetros con menores niveles de potencia de las perturbaciones.

Por ejemplo en las figuras (7.11) y (7.12) se representa la evolución de la identificación hecha con el algoritmo 4, con $R = 4$ para las relaciones señal a ruido $S/N = 10$ y 1 respectivamente.

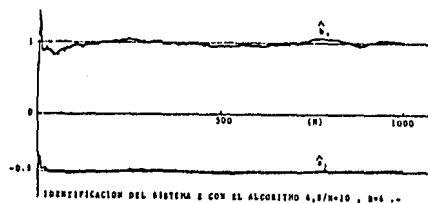


Fig 7.11

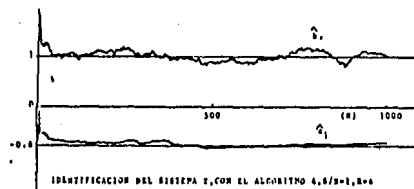


Fig 7.12

En las figuras (7.13) y (7.14) se presentan las estimaciones de a_1 y b_0 con los algoritmos 3 y 4 contra cuatro valores de la relación señal a ruido ($S/N = 0.1, 1, 4, 10$) para $R = 1$ y $R = 3$.

El número de muestras considerado fue $N = 1000$.

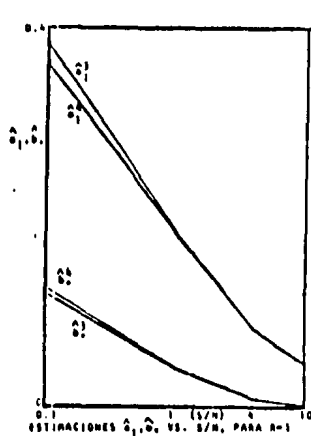


Fig 7.13

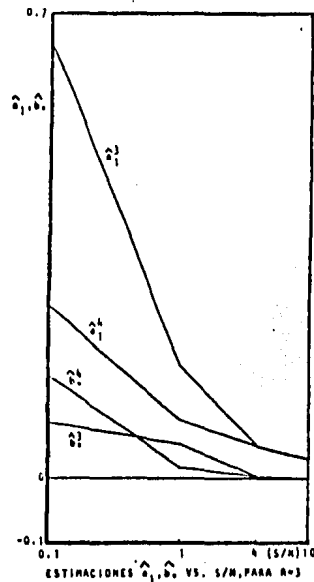


Fig 7.14

En todos los casos se observa que la estimación mejora al disminuir la potencia de las perturbaciones (S/N crecientes).

Las estimaciones \hat{b}_0 están siempre más próximas a b_0 que lo que están las estimaciones \hat{a}_1 a un valor real a_1 . El nivel de referencia cero corresponde al valor real de los parámetros, y lo que se indica en ordenadas es la distancia entre el estimador y el valor real.

Nótese que para los más altos valores de S/N casi no hay diferencia entre las estimaciones obtenidas con los dos algoritmos.

Para $S/N = 1$, tanto con $R = 1$ como con $R = 3$ se obtiene una mejor estimación con el algoritmo 4, y esta diferencia es más notable con más ruido, especialmente para $R = 3$.

Para valores grandes de la potencia del ruido, se encuentran estimadores fuertemente sesgados de a_1 .

Por ejemplo en las figuras (7.15) y (7.16) se presentan las estimaciones de los parámetros a_1 y b_0 para $S/N = 0.1$ con los algoritmos 3 y 4 respectivamente, para $R = 3$.

Puede verse que la estimación de b_0 es bastante mala, con fuertes oscilaciones, y la de a_1 parece sesgada.

En ambos casos (\hat{b}_0 y \hat{a}_1), la estimación obtenida con el algoritmo 4 es mejor. Con \hat{b}_0 las oscilaciones del estimador del algoritmo 4 tiene una amplitud algo menor, y el estimador de \hat{a}_1 obtenido con el mismo al

goritmo presenta menos sesgo que el obtenido con el algoritmo 3.



Fig 7.15

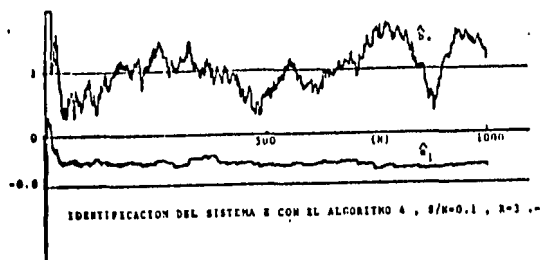


Fig 7.16

Como veremos más adelante, para bajos valores de S/N , la estimación puede en general ser mala, sesgada, y haber más de un punto 0 de convergencia, de la misma manera que para el algoritmo de mínimos cuadrados generalizados no recursivos visto en el capítulo anterior.

La estimación del proceso de las perturbaciones, aproximado como $1/\Gamma(\eta^{-1})$, en general ha tenido peor comportamiento que la estimación de los parámetros a_1 y b_2 de las simulaciones con el sistema E.

Si bien se aproximan o se mueven alrededor de los valores asintóticos (que fueron dados en la tabla 6.IX), tienen fuertes oscilaciones alrededor de estos valores y dependen fuertemente del ruido, aún para los valores más altos de S/N que fueron simulados.

En las figuras (7.17) y (7.18) se grafican las estimaciones de f_i , $i = 1, 2, 3$ a lo largo del tiempo, obtenidas en la identificación del sistema E con el algoritmo 4, para $R = 3$ y relaciones señal a ruido $S/N = 10$ y 1 respectivamente.

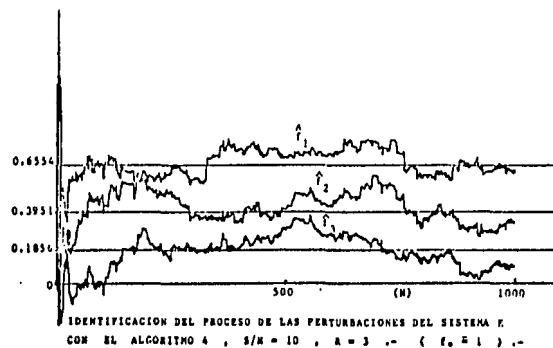


Fig 7.17

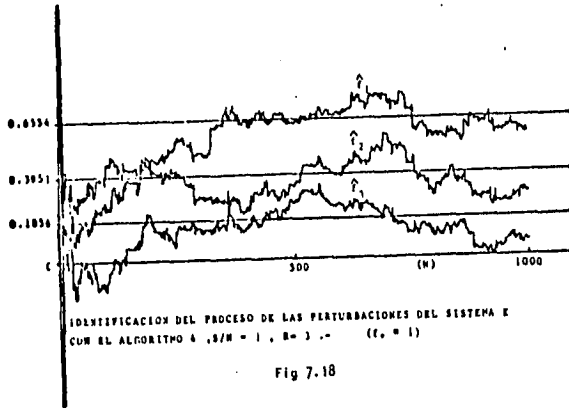


Fig 7.18

Se han graficado también los valores de f_i dados en la tabla 6.IX. Las estimaciones de las figuras (7.17) y (7.18) corresponden a las identificaciones de las figuras (7.10) y (7.4) respectivamente.

Comóranse dichas figuras para notar la diferencia cualitativa entre la estimación de los parámetros a_1 y b_0 con la de los parámetros f_i . En las figuras (7.17) y (7.18) puede observarse también que es mejor la estimación del proceso de las perturbaciones para $S/N = 10$ que para $S/N = 1$, especialmente durante las primeras 200 observaciones.

7.4 Convergencia de M.C.G.R. y otros algoritmos recursivos

7.4.1 Introducción

En los años 60 y principios de los 70, se desarrollaron varios algoritmos recursivos de identificación.

Algunos criterios para determinar los parámetros de un sistema dinámico lineal (por ejemplo basados en estimación estadística paramétrica) fueron aproximados en forma recursiva por distintos algoritmos.

La necesidad de aplicar los resultados de la identificación a espumas de trabajo en línea (por ejemplo en reguladores auto-sintonizables), condujeron a crear algoritmos rápidos que basaron su estimación de los parámetros del sistema a identificar, en una actualización o corrección anterior en base a la llegada de nuevos datos, más que en procesar todos los cálculos otra vez con todo el conjunto pasado de observaciones.

Es significativa la estructura similar de un grupo importante de algoritmos recursivos, y esto es lo que permitió un tratamiento conjunto al estudiar sus propiedades de convergencia.

Basados en los trabajos de Ljung en torno a la convergencia de algoritmos estocásticos recursivos, es que el propio L. Ljung, T. Söderström y J. Gustavson resolvieron (alrededor de 1974) el problema general de las condiciones de convergencia de los algoritmos de mínimos cuadrados generalizados recursivos, variables instrumentales, máxima verosimilitud recursiva (1), máxima verosimilitud recursiva (2), aproximaciones estocásticas.

En esta sección describiremos algunos de estos algoritmos, alguna idea básica de como se abordó el problema de su convergencia, y los resultados principales obtenidos.

7.4.2 Máxima verosimilitud recursiva (1) y (2)

Sea el sistema lineal discreto, descrito por:

$$A(q^{-1}) y(t) = B(q^{-1}) u(t) + C(q^{-1}) e(t) \quad (7.30)$$

dando

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \dots + a_n q^{-n} \\ B(q^{-1}) &= b_0 + b_1 q^{-1} + \dots + b_m q^{-m} \\ C(q^{-1}) &= 1 + c_1 q^{-1} + \dots + c_p q^{-p} \end{aligned} \quad (7.31)$$

$A(q^{-1})$ y $C(q^{-1})$ tienen todos sus ceros fuera del círculo unitario, y no hay factores comunes en los tres polinomios.

$e(t)$ es una señal de ruido blanco de distribución gaussiana: $N(0, \sigma^2)$ con σ^2 constante.

En esta situación, K.J. Aström y T. Bohlin (1965) (6) desarrollan el método de máxima verosimilitud para la identificación del sistema. Básicamente se hace lo siguiente:

Supongamos que se tienen observaciones de las entradas y salidas del sistema: $\{u(t), y(t)\} \quad t = 1, \dots, N$.

Las entradas $u(t)$ y los parámetros del sistema:

$$\theta^T = (a_1, \dots, a_n, b_0, \dots, b_m, c_1, \dots, c_p, \sigma^2)$$

puede considerarse la salida $y(t)$, $t=1, \dots, N$ como una sucesión de variables aleatorias con una densidad de distribución conjunta:

$$p(\{y(t)\} | \{u(t)\}, a_1, \dots, a_n, b_0, \dots, b_m, c_1, \dots, c_p, \sigma^2) = p(\{y(t)\} | \{u(t)\}, \theta) \quad \dots \quad (7.32)$$

construyamos ahora el estimador $\hat{\theta}$ de máxima verosimilitud como lo hicimos en el capítulo 2: elegimos $\hat{\theta}$ como aquél que maximiza la función de máxima verosimilitud, esto es la densidad de distribución conjunta evaluada en las observaciones $\{y(t)\} \quad t=1, \dots, N$.

Para obtener la expresión analítica de la función de verosimilitud notemos que las cantidades $e(t)$, $t=1, \dots, N$ definida por

$$C(q^{-1}) e(t) = A(q^{-1}) y(t) - B(q^{-1}) u(t) \quad (7.33)$$

son por hipótesis variables aleatorias normales $(0, \sigma^2)$, e independientes. Entonces el logaritmo de la densidad de distribución conjunta de $\{e(t)\}$ es:

$$L = -\frac{1}{2\sigma^2} \sum_{t=1}^N e^2(t) - N \log \sigma - \frac{N}{2} \log 2\pi \quad (7.34)$$

(esto ya fue deducido en (2.15) del ejemplo 2 del capítulo 2).

Para conocer ahora la función de verosimilitud de las observaciones $\{y(t)\} \quad t = 1, \dots, N$ aplicamos el teorema 2.1.

Las dos hipótesis de este teorema se cumplen: 1) la transformación $G: \{e(t)\} \rightarrow \{y(t)\}$ es uno a uno, y 2) Existe el jacobiano de la trans

7.4.2 Máxima verosimilitud recursiva (1) y (2)

Sea el sistema lineal discreto, descrito por:

$$A(q^{-1}) y(t) = B(q^{-1}) u(t) + C(q^{-1}) e(t) \quad (7.30)$$

dando

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \dots + a_n q^{-n} \\ B(q^{-1}) &= b_0 + b_1 q^{-1} + \dots + b_m q^{-m} \\ C(q^{-1}) &= 1 + c_1 q^{-1} + \dots + c_r q^{-r} \end{aligned} \quad (7.31)$$

$A(q^{-1})$ y $C(q^{-1})$ tienen todos sus ceros fuera del círculo unitario, y no hay factores comunes en los tres polinomios.

$e(t)$ es una señal de ruido blanco de distribución gaussiana: $N(0, \sigma^2)$ con σ^2 constante.

En esta situación, K.J. Aström y T. Bohlin (1965) (6) desarrollan el método de máxima verosimilitud para la identificación del sistema. Básicamente se hace lo siguiente:

Supongamos que se tienen observaciones de las entradas y salidas del sistema: $\{u(t), y(t)\} \quad t = 1, \dots, N$.

Sean las entradas $u(t)$ y los parámetros del sistema:

$$\theta^T = (a_1, \dots, a_n, b_0, \dots, b_m, c_1, \dots, c_r, \sigma^2)$$

puede considerarse la salida $y(t)$, $t=1, \dots, N$ como una sucesión de variables aleatorias con una densidad de distribución conjunta:

$$p(\{y(t)\} | \{u(t)\}, a_1, \dots, a_n, b_0, \dots, b_m, c_1, \dots, c_r, \sigma^2) = p(\{y(t)\} | \{u(t)\}, \theta) \quad \dots \quad (7.3)$$

construyamos ahora el estimador $\hat{\theta}$ de máxima verosimilitud como lo hicimos en el capítulo 2: elegimos $\hat{\theta}$ como aquél que maximiza la función de máxima verosimilitud, esto es la densidad de distribución conjunta evaluada en las observaciones $\{y(t)\} \quad t=1, \dots, N$.

Para obtener la expresión analítica de la función de verosimilitud notemos que las cantidades $e(t)$, $t=1, \dots, N$ definida por

$$C(q^{-1}) e(t) = A(q^{-1}) y(t) - B(q^{-1}) u(t) \quad (7.33)$$

son por hipótesis variables aleatorias normales $(0, \sigma^2)$, e independientes. Entonces el logaritmo de la densidad de distribución conjunta de $\{e(t)\}$ es:

$$L = -\frac{1}{2\sigma^2} \sum_{t=1}^N e^2(t) - N \log \sigma - \frac{N}{2} \log 2\pi \quad (7.34)$$

(esto ya fue deducido en (2.15) del ejemplo 2 del capítulo 2).

Para conocer ahora la función de verosimilitud de las observaciones $\{y(t)\} \quad t = 1, \dots, N$ aplicamos el teorema 2.1.

Las dos hipótesis de este teorema se cumplen: 1) la transformación $G: \{e(t)\} \rightarrow \{y(t)\}$ es uno a uno, y 2) Existe el jacobiano de la trans

formación. Además el jacobiano es $|J| = 1$, por lo que (7.34) es también el logaritmo de la función de verosimilitud de las observaciones $\{y(t)\}$, $t=1 \dots N$.

En resumen, el estimador de máxima verosimilitud del sistema se obtiene maximizando (7.34), es decir minimizando la función

$$(-L) = \frac{1}{2\sigma^2} \sum_{t=1}^N \varepsilon^2(t) + \frac{N}{2} \log \sigma^2 \quad (7.35)$$

Puesto que el último sumando de (7.34) es constante. La minimización debe ser hecha entre $\theta^T = (a_1, \dots, a_n, b_0, \dots, b_m, c_1, \dots, c_p, \sigma^2)$. Haciendo

$$\begin{aligned} \frac{\partial L}{\partial \sigma^2} &= 0 \\ \sigma^2 &= \hat{\sigma}^2 \\ \hat{\sigma}^2 &= \frac{\sum_{t=1}^N \varepsilon(t)^2}{N} \end{aligned} \quad (7.36)$$

Luego, la minimización de (7.35) puede ser hecha en dos partes:

Definido

$$V(\theta) = \frac{1}{2} \sum_{t=1}^N \varepsilon(t)^2 \quad (7.37)$$

donde $\theta^T = (a_1, \dots, a_n, b_0, \dots, b_m, c_1, \dots, c_p)$, se obtiene $\hat{\theta}$ minimizando (7.37), y σ^2 como:

$$\hat{\sigma}^2 = \frac{2}{N} V(\hat{\theta}) \quad (7.38)$$

que resulta de (7.36).

Para minimizar (7.37) Åström y Bohlin usan un esquema iterativo de Newton-Raphson. Desde su trabajo muchas formas de minimizar (7.34) han sido empleadas.

El estimador de máxima verosimilitud así obtenido tiene propiedades asintóticas que lo hacen muy atractivo.

Debe tenerse en cuenta, que igual a lo sucedido en la estimación por mínimos cuadrados, este estimador de máxima verosimilitud no tiene inmediatamente las propiedades óptimas estudiadas en el Capítulo 2.

Esto se debe precisamente a que las observaciones $y(1) \dots y(N)$ no son independientes, y no constituyen por lo tanto una muestra aleatoria de una misma distribución. Esta era la hipótesis esencial de los teoremas 2.8, 2.9 y 2.10.

Åström, Bohlin y Wensmark (1965) (7), demuestran que bajo ciertas restricciones a la clase de entradas (básicamente que la entrada sea persistentemente excitante de un orden suficientemente grande, y que existan los límites $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t)$, $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t) u(t+T) \neq T$ finito), entonces asintóticamente se cumplen las propiedades óptimas del estimador de máxima verosimilitud ($N \rightarrow \infty$):

- 1) El estimador es consistente: $\hat{\theta}_N \xrightarrow{\text{prob } 1} \theta_0$
- 2) El estimador es asintóticamente normal: $\sqrt{N} (\hat{\theta}_N - \theta_0) \xrightarrow{\text{en ley}} N(0, I)$

3) El estimador es asintóticamente eficiente (la varianza de la distribución asintótica alcanza la cota de Crammer-Rao).

Un resumen de estos resultados se encuentran en (6). El esquema de optimización planteado por Åström y Bohlin (6), está pensado para una implantación fuera de línea, y si bien es estadísticamente óptimo, es computacionalmente largo.

Panuska y Young en 1968, por separado, sugieren un esquema recursivo, para la aproximación del estimador de máxima verosimilitud, al que de acuerdo con (47) hemos denominado:

Algoritmo recursivo de máxima verosimilitud (1)

Sea el sistema (7.30) que genera las observaciones $\{y(t)\}$ $t=1, \dots, N$, a partir de entradas $\{u(t)\}$ y ruido blanco gaussiano $\{e(t)\}$.

Si la sucesión del ruido fuera medible, el problema de identificación podría ser considerado como un problema de mínimos cuadrados ordinarios de la siguiente manera:

$$\begin{aligned} \text{Llamando } \theta^T &= (a_1, \dots, a_n, b_0, \dots, b_m, c_1, \dots, c_p) \\ x_t^T &= (-y(t-1), \dots, -y(t-n), u(t), \dots, u(t-m), e(t-1), \dots, e(t-r)) \\ &\dots \quad (7.39) \end{aligned}$$

entonces

$$y(t) = x_t^T \theta + e(t) \quad (7.40)$$

para $t = 1, \dots, N$, con errores $e(t)$ normales, independientes e igualmente distribuidos. Aplicando el esquema de mínimos cuadrados recursivos de Plackett, tendríamos una estimación asintóticamente insesgada de θ

$$\begin{aligned} \hat{\theta}(t+1) &= \hat{\theta}(t) + k(t+1) e(t+1) \\ k(t+1) &= \frac{P(t) x_{t+1}}{1 + x_{t+1}^T P(t) x_{t+1}} \\ P(t+1) &= P(t) - \frac{P(t) x_{t+1} x_{t+1}^T P(t)}{1 + x_{t+1}^T P(t) x_{t+1}} \quad (7.41) \\ e(t+1) &= y(t+1) - x_{t+1}^T \hat{\theta}(t) \end{aligned}$$

El algoritmo (7.41), así como está, tiene un problema, en el vector x_{t+1} hay elementos $e(t-i)$, $i=0, \dots, r-1$, que en realidad se desconocen. Lo que proponen Panuska y Young es aproximar x_t por \hat{x}_t :

$$\hat{x}_t = (-y(t-1), \dots, -y(t-n), u(t), \dots, u(t-m), e(t-1), \dots, e(t-r)) \quad \dots \quad (7.42)$$

es decir estimar el ruido $e(t)$ con la última ecuación (7.41).

El algoritmo (7.41) con \hat{x}_{t+1} en lugar de x_{t+1} , dado por (7.42) es el algoritmo recursivo de máxima verosimilitud (1).

Algoritmo recursivo de máxima verosimilitud (2)

La minimización de $V(0)$ de (7.37), que es el problema de máxima verosimilitud, da lugar a un problema lineal en los parámetros a_j , b_j , $j = 1, \dots, n$, $j = 0, \dots, m$, y no lineal en los parámetros c_k , $k = 1, \dots, r$.

Esto es debido a la estructura de los residuos $e(t)$, $V(0) = \frac{1}{2} \sum_{t=1}^N e^2(t)$

$$e(t) = \frac{A(q^{-1})}{C(q^{-1})} y(t) - \frac{B(q^{-1})}{C(q^{-1})} u(t) \quad (7.43)$$

En 1973, B. Furth [22] presenta una técnica alternativa para la minimización de $V(0)$. Es un método iterativo basado en la obtención en la iteración $i+1$ de una aproximación $e^{i+1}(t)$ al proceso $e(t)$ a partir de la estimación de los parámetros en la iteración anterior i , que conlleva al siguiente algoritmo:

Calculando:

$$\begin{cases} \hat{y}^i(t) = \frac{v(t)}{c^i(q^{-1})} \\ \hat{u}^i(t) = \frac{u(t)}{c^i(q^{-1})} \\ \hat{e}^i(t) = \frac{e(t)}{c^i(q^{-1})} \end{cases} \quad (7.44)$$

$$\begin{cases} \hat{x}_t^i = \{-\hat{y}^i(t-1), \dots, -\hat{y}^i(t-n), \hat{u}^i(t), \dots, \hat{u}^i(t-m), \hat{e}^i(t-1), \dots, \hat{e}^i(t-r)\} \\ \hat{w}^i(t) = \hat{y}^i(t) - \hat{e}^i(t) + c^i(t) \\ \hat{c}^{i+1}(t) = \hat{w}^i(t) - \hat{x}_t^i \hat{\theta}^i(i+1) \end{cases} \quad (7.45)$$

donde:

- $\hat{\theta}^i(i)$ es la estimación de $\theta = (a_1 \dots a_n, b_0 \dots b_m, c_1 \dots c_r)$ en la iteración i
- $\hat{c}^i(q^{-1})$ es la estimación de $c(q^{-1})$ con los parámetros correspondientes de $\hat{\theta}^i(i)$
- las señales \hat{y}^i , \hat{u}^i , \hat{e}^i son respectivamente las que resultaban de filtrar $y(t)$, $u(t)$, $e^i(t)$ por el filtro $1/\hat{c}^i(q^{-1})$

entonces:

$$\hat{\theta}^{i+1} = \left[\begin{array}{c} N \\ \sum_{t=1} \end{array} \begin{array}{cc} x_t^i & x_t^i \end{array} \right]^{-1} \left[\begin{array}{c} N \\ \sum_{t=1} \end{array} \begin{array}{cc} x_t^i & w^i(t) \end{array} \right] \quad (7.46)$$

El algoritmo queda como sigue:

- 1) Encontrar valores iniciales para la estimación de los parámetros
- 2) En la iteración i ($i=1, 2, \dots$):
 - a. Calcular los errores con 7.43 utilizando los parámetros estimados en $i-1$
 - b. Filtrar las observaciones de entrada, salida, y errores como (7.44)
 - c. Actualizar la estimación de los parámetros con (7.46)
 - d. Si converge parar, si no volver a a en la siguiente iteración

Furth (22) sugiere comenzar el algoritmo con una iteración idéntica a mínimos cuadrados generalizados, para obtener buena inicialización para los parámetros c_i , $i=1, \dots, r$.

Versión recursiva

Utilizando el "lema de inversión de matrices", Furth (21), llega a una expresión recursiva para el algoritmo anterior, a la que se le ha denominado algoritmo recursivo de máxima verosimilitud (2).

Por separado, Söderström al mismo tiempo propone un algoritmo casi idéntico, pero algo más preciso. Es el que presentaremos:

$$\hat{\theta}(t+1) = \hat{\theta}(t) + k(t+1) e(t+1)$$

$$k(t+1) = \frac{P(t) \hat{x}_{t+1}^T}{1 + \hat{x}_{t+1}^T P(t) \hat{x}_{t+1}} \quad (7.47)$$

$$P(t+1) = P(t) - \frac{F(t) \hat{x}_{t+1}^T \hat{x}_{t+1}^T P(t)}{1 + \hat{x}_{t+1}^T P(t) \hat{x}_{t+1}}$$

dando:

$$e(t) = y(t) - (-y(t-1), \dots, -y(t-n), u(t), \dots, u(t-m), c(t-1), \dots, c(t-r)) \hat{\theta}(t-1)$$

$$\hat{x}_t^T = \left(-\frac{1}{c(q^{-1})} y(t-1), \dots, -\frac{1}{c(q^{-1})} y(t-n), \frac{1}{c(q^{-1})} u(t), \dots, \right.$$

$$\left. \frac{1}{c(q^{-1})} u(t-m), \frac{1}{c(q^{-1})} e(t-1), \dots, \frac{1}{c(q^{-1})} e(t-r) \right)$$

... (7.48)

el vector \hat{x}_t puede ser actualizado con:

$$\hat{x}_{t+1}^T = \begin{bmatrix} -\hat{c}_1 \dots -\hat{c}_r & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & \vdots & \vdots \\ -\hat{c}_1 \dots -\hat{c}_r & 0 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & -\hat{c}_1 \dots -\hat{c}_r \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \end{bmatrix} \hat{x}_t^T + \begin{bmatrix} y(t) \\ \vdots \\ \vdots \\ -u(t) \\ \vdots \\ \vdots \\ -c(t) \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$$

... (7.49)

en (7.49) los coeficientes \hat{c}_i , $i=1, \dots, r$ son los que corresponden a $\hat{\theta}(t)$.

7.4.3 Forma general del algoritmo recursivo. Ecuación diferencial ordinaria asociada

Habíamos dicho que el problema de la convergencia fue abordado por L. Ljung y otros autores, en esta sección se definen algunas de las ideas básicas del procedimiento utilizado, en una forma muy general.

La discusión detallada puede ser leída en las referencias (36)y(47). Consideremos el sistema discreto lineal:

$$y(t) = G(q^{-1}) u(t) + H(q^{-1}) e(t) \quad (7.50)$$

donde G y H son funciones racionales en q^{-1} , $H(z)$ tiene todos sus polos fuera del círculo unitario, y $e(t)$ es ruido blanco.

Consideraremos la forma particular de (7.50), distinta para cada algoritmo:

a) Mínimos cuadrados: (M.C.R.)

$$\begin{aligned} G(q^{-1}) &= \frac{B(q^{-1})}{\lambda(q^{-1})} \\ H(q^{-1}) &= \frac{1}{\lambda(q^{-1})} \end{aligned} \quad (7.51)$$

b) Máxima verosimilitud recursiva (1) y (2): (M.V.R. (1) y (2))

$$\begin{aligned} G(q^{-1}) &= \frac{B(q^{-1})}{\lambda(q^{-1})} \\ H(q^{-1}) &= \frac{B(q^{-1})}{\lambda(q^{-1})} \end{aligned} \quad (7.52)$$

c) Mínimos cuadrados generalizados recursivos (M.C.G.R.)

$$\begin{aligned} G(q^{-1}) &= \frac{B(q^{-1})}{A(q^{-1})} \\ H(q^{-1}) &= \begin{cases} \text{caso I: } 1/A(q^{-1}) C(q^{-1}) \\ \text{caso II: } C(q^{-1})/A(q^{-1}) \end{cases} \end{aligned} \quad (7.53)$$

donde

$$\begin{cases} A(q^{-1}) = 1 + a_1 q^{-1} + \dots + a_n q^{-n} \\ B(q^{-1}) = b_0 + b_1 q^{-1} + \dots + b_m q^{-m} \\ C(q^{-1}) = 1 + c_1 q^{-1} + \dots + c_p q^{-p} \end{cases} \quad (7.54)$$

La hipótesis hecha implica que $A(z)$ tiene sus ceros fuera del círculo unitario. Supongamos además que $C(z)$ tiene sus ceros fuera

del círculo unitario y no hay factores comunes a los tres polinómios.

En cada caso se supone que el sistema que genera los datos tiene la forma (7.50), con (7.51), (7.52) ó (7.53) dependiendo de qué algoritmo se trate. Llamemos θ , vector de parámetros a:

$$\begin{aligned} \text{a) } \theta^T &= (a_1, \dots, a_n, b_0, \dots, b_m) \text{ para M.C.R.} \\ \theta^T &= (a_1, \dots, a_n, b_0, \dots, b_m, c_1, \dots, c_p) \text{ para M.C.G.R. M.V.R.} \\ &\quad (1) \text{ y M.V.R. (2)} \end{aligned}$$

Llamemos θ_0 el vector de parámetros cuyos coeficientes corresponden al sistema real que genera los datos, y $\hat{\theta}$ al vector estimado, cuyos elementos son los estimadores de cada elemento de θ_0 (su valor real).

Notase que todos los algoritmos tienen la forma recursiva:

$$\begin{cases} \hat{\theta}(t+1) = \hat{\theta}(t) + K(t+1) e(t+1) \\ k(t+1) = \frac{P(t) z(t+1)}{\lambda(t+1) + \psi(t+1)^T P(t) z(t+1)} \\ P(t+1) = \left[P(t) - \frac{P(t) z(t+1) \psi(t+1)^T P(t)}{\lambda(t+1) + \psi(t+1)^T P(t) z(t+1)} \right] / \lambda(t+1); \\ \lambda(t+1) = \lambda_0 \lambda(t) + (1-\lambda_0) \end{cases} \quad (7.55)$$

Para comprobarlo basta con comparar (7.55) con (5.64), (7.20) a (7.25), (7.41) y (7.47).

Lo único que cambia es el significado de $\hat{\theta}(t)$, $e(t)$, $\psi(t)$ y $z(t)$, que se dan en la tabla 7.I para cada algoritmo.

En los cuatro algoritmos que hemos visto (M.C.R., M.C.G.R., M.V.R(1), M.V.R.(2) se cumple $\psi(t) = z(t)$, la única razón para mantener la variable $z(t)$ es para ser coherentes con la notación de Söderström et al. (3), donde se introduce la posibilidad $z(t) \neq \psi(t)$ para poder incluir en la forma general (7.55) el algoritmo de variables instrumentales re cursivo.

La sucesión $\lambda(t)$ se interpreta asociada a los pesos relativos de la función de costo del algoritmo, que para todos los casos vistos, es una función cuadrática de los errores del modelo de predicción.

Para obtenerse fácilmente una forma alternativa para la expresión general (7.55):

$$\begin{cases} \hat{O}(t+1) = \hat{O}(t) + \gamma(t+1) \tilde{K}(t+1) e(t+1) \\ \tilde{Y}(t) = P(t) z(t+1) / [1 + \gamma(t+1) \psi(t+1)^T P(t) z(t+1) - 1] \\ \tilde{P}(t+1)^{-1} = \tilde{P}(t)^{-1} + \gamma(t+1) [z(t+1) \psi(t+1)^T - \tilde{Y}(t)^{-1}] \end{cases} \quad (7.55')$$

Para ello basta con hacer:

$$\begin{cases} \gamma(t+1) = \frac{\gamma(t)}{\lambda(t+1) + \gamma(t)}; \quad \gamma(0) = 1 \\ \tilde{P}(t) = \frac{1}{\gamma(t)} P(t) \\ \tilde{K}(t) = \frac{1}{\gamma(t)} K(t) \end{cases}$$

Sobre este algoritmo general es que se analiza la convergencia. Surgen g raves que la entrada al sistema, en el caso más general es:

$$u(t) = u_R(t) + \Gamma_1(q^{-1}) y(t) + \Gamma_2(\hat{O}(t), q^{-1}) y(t) \quad (7.56)$$

con $u_R(t)$ estacionaria e independiente de $e(t)$. Llamaremos ahora

$$\begin{cases} F(0) = E \{ \bar{z}(t; 0) \bar{e}(t; 0) \} \\ G(0) = E \{ \bar{z}(t; 0) \bar{\psi}(t; 0)^T \} \end{cases} \quad (7.57)$$

donde las esperanzas son tomadas sobre $\{e(t)\}$, y sumando

- 1) $u(t)$ y $e(t)$ son estacionarias
- 2) \bar{z} , \bar{e} , $\bar{\psi}$ son calculadas suponiendo que se ha alcanzado una situación estacionaria, con un estimador $\hat{O} = \hat{O}$ invariante.

Definimos:

$$\begin{cases} D_u = \{0\} \text{ el sistema (7.50) realimentado con (7.56) es asintóticamente estable es decir el conjunto de parámetros } \theta \text{ tales que las soluciones de:} \\ I - G(z) F_1(z) - G(z) F_2(\hat{O}, z) = 0 \text{ están fuera del círculo (7.58)} \\ \text{unitario.} \\ D_p = \{0\} \text{ } C(z) \text{ tiene sus ceros fuera del círculo unitario} \\ D_1 = \{0\} \text{ } G(0) \text{ es invertible} \end{cases}$$

consideremos la ecuación diferencial ordinaria:

$$\begin{cases} \dot{U} = R^{-1} F(0) \\ \dot{R} = G(0) - R \end{cases} \quad (7.59)$$

Teorema 7.1

Sea el sistema (7.50) con entrada (7.56). Si existe el tercer sumando en (7.56), es $F_2(\hat{O}(t), q^{-1})$ lipshitziana en 0

Si se cumple que:

- i) $\sum_{t=1}^{\infty} \gamma(t) < \infty$
- ii) $\sum_{t=1}^{\infty} \gamma(t)^p < \infty$ para algún $p > 1$
- iii) $\gamma(t+1) > \gamma(t) (1 - \gamma(t+1))$
- iv) $\{\gamma(t)\}$ es decreciente
- v) $E |e(t)|^p < \infty \quad p > 0$

entonces se cumple que:

a) $\hat{O}(t) \rightarrow O^* \in D_a \cap D_p \cap D_f$ con probabilidad distinta de cero solo si:

si:

- i) $f(O^*) = 0$
- ii) los autovalores de $G(O^*)^{-1} \frac{d}{dO} f(O)|_{O=O^*}$

tienen partes reales no positivas.

b) Si $\hat{O}(t)$ pertenece a un subconjunto compacto contenido en $D_a \cap D_p$, infinitamente seguido con probabilidad 1, y $(O, R) = (O^*, G(O^*))$ con $O^* \in D_f$ es punto estacionario de la ecuación diferencial (7.59), globalmente asintóticamente estable, entonces

$$\hat{O}(t) \xrightarrow{\text{Prob. 1}} O^* \text{ para } t \rightarrow \infty$$

La idea central del análisis de convergencia de los algoritmos recursivos de la forma (7.55) es utilizar los resultados de este teorema.

Los posibles puntos de convergencia del algoritmo deben ser tales que cumplan la condición (7.60).

Lo ideal sería que hubiera un solo valor de 0 (O^*) que fuera un posible punto de convergencia del algoritmo. No es así siempre, y en la verificación de las condiciones (7.60) se puede establecer en que condiciones cada algoritmo puede converger (con probabilidad mayor que cero) a uno o más puntos.

La parte b del teorema establece que si hay un punto O^*, R^* que es estacionario ($\dot{O}|_{O=O^*} = 0, \dot{R}|_{R=R^*} = 0$) y es globalmente asintóticamente estable (para la ecuación $\frac{d}{dt} \begin{matrix} O \\ R \end{matrix} = \begin{matrix} f(O) \\ g(O, R) \end{matrix}$ diferencial ordinaria asociada al algoritmo), entonces $\hat{O}(t) \xrightarrow{\text{Prob. 1}} O^*$ para $t \rightarrow \infty$.

Lo deseable, obviamente es que $O^* = O_0$.

El estudio de la convergencia de los algoritmos recursivos se divide entonces en dos partes: 1) Estudio de unicidad ó multiplicidad de los posibles puntos de convergencia; 2) Estudio de la estabilidad de los posibles puntos de convergencia.

En los siguientes párrafos presentaremos los resultados obtenidos para los algoritmos recursivos que hemos visto. En el párrafo 7.4.4. se discute el caso de mínimos cuadrados generalizados recursivos que es el discutido en este capítulo, y en 7.4.5 se enuncian en forma comparativa los resultados para los algoritmos M.G.R., M.C.G.R., M.V.R. (2)

Tabla 611.1. Similitud del método $\bar{c}(t)$ de $\bar{c}(t)$ del algoritmo recursivo optimizado de arriba a algoritmos anteriores (Forma de los algoritmos, véase el Capítulo 1.).

Algoritmo	Método	Parámetro optimizado	$\bar{c}(t)$	$\bar{c}(t)$
1	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$
2	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$
3	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$
4	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$	$\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t) = \bar{c}(t)$

* La forma algoritmo que se aplica, es $\bar{c}(t) = \bar{c}(t)$. (En M.C.G.R. $\bar{c}(t) = \bar{c}(t) + \bar{c}(t) - \bar{c}(t)$)

y M.V.R. (2).

7.4.4 Convergencia de M.C.G.R.

Para este método, que consiste en dos algoritmos recursivos acoplados (ver tabla 7.1), la condición (7.59) para un punto estacionario es:

$$E\{\bar{c}_1(t, 0^h) \bar{\psi}_1(t, 0^h)\} = 0 \quad (7.61)$$

$$E\{\bar{c}_2(t, 0^h) \bar{\psi}_2(t, 0^h)\} = 0$$

se puede demostrar que (7.61) se pueden escribir en forma equivalente (ver (47)), como:

$$V^*(0^h) = 0 \quad (7.62)$$

donde

$$V^*(0^h) = \frac{1}{2} E\{e^2(0^h, t)\} \quad (7.63)$$

con

$$e(0^h, t) = A^h(q^{-1}) C^h(q^{-1}) y(t) - B^h(q^{-1}) C^h(q^{-1}) u(t) \quad (7.64)$$

donde A^h, B^h, C^h son los polinomios A, B, C en función con los parámetros en $0=0^h$.

(Nota: se ha asumido que el modelo que genera los datos corresponde al caso 1 de (7.53)).

Esto significa que el método minimiza la misma función de costo $V(t)$ dada por (6.23) que el algoritmo 2 visto en el capítulo anterior (con

$$\hat{F} = \hat{G}.$$

Vale entonces la discusión hecha por Söderström en torno a los posibles puntos estacionarios y mínimos locales de dicha función.

De particular se concluye que si:

- 1) El sistema es controlable
- 2) La señal de entrada es de excitación persistente de orden $n+m$, entonces:

Existen dos constantes S_0 y S_1 ; $0 < S_1 < S_0 < \infty$, tales que:

- a) Si la relación señal a ruido S/N es tal que $S_0 < S/N$ entonces $O^0 = O_0$ es la única solución de (7.62) (y por tanto de (7.61))
- b) Si la relación señal a ruido S/N es tal que $0 < S/N < S_1$, entonces (7.62) tiene múltiples soluciones.

El análisis de la estabilidad de los posibles puntos de convergencia del algoritmo es bastante complicado.

Puede demostrarse (ver (47) por ejemplo), que cuando $O^0 = O_0$ es un punto estacionario de la ecuación diferencial ordinaria asociada (al menos para S/N grandes, caso (a), lo es), este es siempre globalmente asintóticamente estable. Para el caso (b), con S/N pequeño, puede verse que los puntos estacionarios son estables para valores suficientemente pequeños de S/N .

Hemos supuesto que los datos eran generados con el modelo (7.53) caso i), pero si son generados por (7.53) caso ii) valen los mismos resultados con \hat{O}_0 en lugar de O_0 , donde \hat{O}_0 es un parámetro próximo a O_0 en el sentido de (6.24), es decir:

$$\hat{O}_0 = (\hat{a}_1, \dots, \hat{a}_n, \hat{b}_0, \dots, \hat{b}_m, \hat{c}_1, \dots, \hat{c}_p)^T$$

con

$$\begin{aligned} \hat{a}_i &= a_i + O(1/s) \quad s \rightarrow \infty, \quad i = 1 \dots n \\ \hat{b}_i &= b_i + O(1/s) \quad s \rightarrow \infty, \quad i = 0 \dots m \\ \hat{c}_i &= c_i + O(1/s) \quad s \rightarrow \infty, \quad i = 1 \dots p \end{aligned}$$

donde \hat{c}_i son tales que

$$C(q^{-1}) = 1 + \hat{c}_1 q^{-1} + \dots + \hat{c}_p q^{-p}$$

mínimiza

$$E\{\hat{C}(q^{-1}) \cdot C(q^{-1}) e(t)\}^2$$

7.4.5 Convergencia de los algoritmos M.C.R., M.C.G.R., M.V.R. (1), M.V.R. (2).

Se describen en las tablas (7.2) y (7.3) los resultados obtenidos en (47) para estos algoritmos. Las hipótesis adicionales en operación en lazo abierto (excepto para el caso $S/N > S_0$ en M.C.G.R.), la entrada $u(t)$ persistentemente excitante de orden $n+m$.

La tabla (7.3) señala las condiciones que garantizan un solo punto posible de convergencia, que además es $O^0 = O_0$, y las condiciones que garantizan varios puntos posibles de convergencia.

La tabla (7.3) señala las condiciones suficientes para la estabilidad local del punto de convergencia θ_0 .

Tabla 7.ii. Condiciones suficientes para una solución única o múltiples soluciones a (7.60), para cuatro algoritmos recursivos.

Algoritmo	Condiciones suficientes para una solución única	Condiciones suficientes para varias soluciones
M.C.R.	No se necesita	Nunca hay varias soluciones
M.C.G.R.	$S/N > S_0$ (cte)	$S/N < S$ (cte)
M.V.R.(1)	Cualquiera de las siguientes: i) el proceso es AFMA($B(q^{-1}) = 0$) ii) $S/N < S_0$ (cte) iii) $n_1 \neq n_2$, y $u(t)$ ruido blanco	No se examina
M.V.R.(2)	Cualquiera de las siguientes: i) el proceso es AFMA ($B(q^{-1}) = 0$) ii) $S/N < S_0$ (cte) iii) $S/N > S_0$ (cte) y $S_0 > S_1$	No se examina

Tabla 7.iii Condiciones suficientes para la estabilidad local de θ_0 para cuatro algoritmos recursivos.

Algoritmo	Condiciones suficientes para estabilidad local de θ_0
M.C.R.	Siempre estable
M.C.G.R.	Siempre estable
M.V.R.(1)	Estable para procesos MA y procesos AFMA de primer orden
M.V.R.(2)	Siempre estable

Para el estudio de la estabilidad pueden levantarse en general restricciones sobre la operación en lazo abierto y permitir realimentación incluso con dependencia de las estimaciones $\hat{\theta}(t)$.

Para M.C.R., M.C.G.R., y M.V.R. (2) se puede probar estabilidad global.

$$A(x + \Delta x) = b + \Delta b \quad (A.2)$$

$$\therefore A \Delta x = \Delta b$$

$$\Delta x = A^{-1} \Delta b$$

$$\|\Delta x\| = \|A^{-1} \Delta b\| < \|A^{-1}\| \|\Delta b\|$$

dividiendo entre $\|x\|$:

$$\frac{\|\Delta x\|}{\|x\|} \leq \|A^{-1}\| \frac{\|\Delta b\|}{\|b\|} \quad (A.3)$$

de (A.1) se deduce que

$$\|A\| \|x\| \geq \|b\| \quad (A.4)$$

$$\therefore \|x\| \geq \frac{\|b\|}{\|A\|} \quad (A.5)$$

Sustituyendo (A.5) en (A.3)

$$\frac{\|\Delta x\|}{\|x\|} \leq \|A\| \|A^{-1}\| \frac{\|\Delta b\|}{\|b\|} \quad (A.6)$$

Definimos el número de condición de A como:

$$\text{cond}(A) = \|A\| \|A^{-1}\|$$

$$\therefore \frac{\|\Delta x\|}{\|x\|} \leq \text{cond}(A) \frac{\|\Delta b\|}{\|b\|}$$

El número de condición es una cota para el factor de amplificación de los errores relativos. Para resolver (A.1) lo que se hace es multiplicar por transformaciones Q_1 en forma que $Q_1 \dots Q_n A = RA$.

APENDICE A

A.1 Algunas problemas numéricos de $Ax=b$

Sean A ($n \times n$), x ($n \times 1$) (incógnita), b ($n \times 1$)

se quiere encontrar x tal que

$$Ax = b \quad (A.1)$$

se supone A no singular. Hay muchos métodos numéricos para resolver el problema (A.1). En la práctica interesa conocer qué tan buena es la solución, en el sentido de si depende fuertemente de errores de aproximación, redondeo, etc.

Supongamos errores en el vector b ¿qué tanto variaría la solución x ?
Si b varía en Δb , x varía en Δx :

En forma que el problema $Rx = Rb$ sea más sencillo de resolver (p. ej. triangular). Una buena selección de las matrices Q_i es elegir las de forma que $\|Q_i\| = \|Q_i^{-1}\| = 1$ de modo que no se imponga el número de condición del problema (A.1).

A.2 Transformaciones de Householder

Se define una matriz de transformación de Householder Q como:

$$Q = I - 2u u^T \quad (A.7)$$

$$\|u\| = 1 \text{ (norma euclidiana)} \quad (A.8)$$

Se pueden demostrar las siguientes propiedades:

- a) Q es simétrica: $Q^T = Q$
- b) Q es ortogonal: $Q^T Q = Q Q^T = I$
- c) El producto de transformaciones de Householder es una transformación ortogonal:

Si $\psi = Q_p \dots Q_1$ (Q_i de Householder $i = 1, \dots, p$)

Entonces $\psi^T \psi = I$

- d) Dado cualquier vector x de dimensión finita, entonces existe una transformación Q de Householder tal que:

$$Qx = (\lambda, 0, 0, \dots, 0)^T, \text{ con } \lambda = \sqrt{(x^T x)}$$

Hasta para ello elegir

$$Q = I - 2uu^T \text{ de forma que:}$$

$$\text{si } x = (x_1, \dots, x_p)^T, \quad u = (u_1, \dots, u_p)^T \text{ entonces:}$$

$$\begin{aligned} u_1 &= \sqrt{\frac{1}{2} (1 - \frac{x_1}{\lambda})} \\ u_i &= -\frac{x_i}{2u_1 \lambda} \\ \lambda &= \sqrt{x^T x} \end{aligned} \quad (A.9)$$

La matriz Q queda:

$$Q = \begin{bmatrix} 1-2u_1^2 & -2u_1u_2 & \dots & -2u_1u_p \\ -2u_1u_2 & 1-2u_2^2 & \dots & -2u_2u_p \\ \vdots & \vdots & \ddots & \vdots \\ -2u_1u_p & -2u_2u_p & \dots & 1-2u_p^2 \end{bmatrix} \quad (A.10)$$

A.3 Triangularización con Matrices de Householder

Sea una matriz $A(N \times p)$, $N > p$, de rango p . Se busca una transformación ortogonal $Q(N \times N)$ tal que

$$Q^T A = \begin{bmatrix} R \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} p \\ N-p \\ \downarrow \\ +p \end{matrix}$$

donde R es triangular superior. Elegimos $Q^T = Q'_{p-1} \dots Q'_1$, donde

las matrices Q'_i tienen la forma:

$$Q'_i = \begin{bmatrix} I_i & & & 0 \\ & \dots & & \\ & & \dots & \\ 0 & & & H_i \end{bmatrix} \begin{matrix} \uparrow \\ \vdots \\ \downarrow \\ H_i^{\downarrow} I \end{matrix} \quad (A.11)$$

Con I_i : matriz identidad ($i \times i$), $H_i(N-i \times N-i)$ de Householder. Es fácil ver que la matriz Q'_i también es de Householder $Q'_i = I_N - 2u^i u^{i T}$ donde los primeros i elementos de u^i son nulos. Obsérvese que para cualquier matriz $B(N \times p)$ particionada de la forma:

$$B = \left[\begin{array}{c|c} B_1 & B_2 \\ \hline B_3 & B_4 \end{array} \right]^T \begin{matrix} i \\ i \end{matrix}$$

$\leftarrow j \rightarrow$

se cumple que $Q'_j B = \left[\begin{array}{c|c} B_1 & B_2 \\ \hline H_j B_3 & H_j B_4 \end{array} \right]$

Tomemos H_0 tal que

$$Q'_0 A = H_0 A = \left[\begin{array}{c|c} \lambda_1 & x \dots x \\ \hline 0 & x \dots x \\ \vdots & \vdots \\ 0 & x \dots x \end{array} \right] \quad (\text{los elementos indicados con } x \text{ pueden tener cualquier valor}).$$

Luego H_1 tal que:

$$Q'_1 C'_0 A = Q'_1 \left[\begin{array}{c|c} \lambda_1 & x \dots x \\ \hline 0 & x \dots x \\ \vdots & \vdots \\ 0 & x \dots x \end{array} \right] = \left[\begin{array}{c|c} \lambda_1 & x \dots x \\ \hline 0 & \lambda_2 x \dots x \\ \vdots & \vdots \\ 0 & x \dots x \end{array} \right]$$

La H_1 es tal que

$$Q'_1 Q'_{1-1} \dots Q'_0 A = Q'_1 \left[\begin{array}{c|c} \lambda_1 & x \dots x \\ \hline \lambda_2 & x \dots x \\ \vdots & \vdots \\ 0 & x \dots x \\ \hline 0 & x_1 \dots x \\ \vdots & \vdots \\ 0 & x_{N-1} \dots x \end{array} \right] = \left[\begin{array}{c|c} \lambda_1 & x \dots x \\ \hline \lambda_2 & x \dots x \\ \vdots & \vdots \\ 0 & x \dots x \\ \hline 0 & \lambda_1 x_1 \dots x \\ \vdots & \vdots \\ 0 & x \dots x \end{array} \right]$$

para ello basta considerar H_1 de Householder tal que

$$H_1 (x_1 \dots x_{N-1})^T = (\lambda_1, 0 \dots 0)^T \quad (\lambda_1^2 = \sum_{j=1}^{N-1} x_j^2)$$

la matriz H_1 se construye con las ecuaciones (A.9). Sustituyendo los valores de $u_1 \dots u_{N-1}$ de las ecuaciones (A.9) se obtiene que

$$H = \{h_{ij}\} \quad i, j = 1, 2 \dots N-1 \text{ con:}$$

$$h_{ij} = \begin{cases} \frac{x_j}{\lambda} & \text{si } i=1, j=1, \dots, N-1 \\ 1 - \frac{x_i^2}{\lambda^2 - x_i^2} & \text{si } i=j, i=2, \dots, N-1 \\ -\frac{x_j x_i}{\lambda^2 - x_i^2} & j>i, i=2, \dots, N-1 \end{cases} \quad (A.12)$$

$$\lambda^2 = x_1^2 + x_2^2 + \dots + x_{N-1}^2$$

El resto de los elementos son simétricos. Por estabilidad numérica conviene elegir el signo de λ tal que $-x_1 \lambda > 0$.

A.4 Aplicación a la solución de $Ax=b$. Descomposición QR

En el caso que $n = p$ apliquemos los resultados anteriores. Sean $A(p \times p)$ de rango completo, $x(p \times 1)$, $b(p \times 1)$ sea la ecuación a resolver el problema (A.1) conocidos A y b , encontrar x tal que:

$$Ax = b$$

En la sección anterior vimos como encontrar q tal que

$$Q^T A = R \text{ (con } R \text{ triangular superior)} \quad (\text{A.13})$$

$$\therefore A = QR \text{ (por ortogonalidad de } Q) \quad (\text{A.14})$$

Esta descomposición de la matriz A se conoce como descomposición QR.

El problema (A.1) se resuelve como sigue:

por (A.14) $QRx = b$

$$\therefore Rx = Q^T b \quad (\text{A.15})$$

$$\therefore x = R^{-1} Q^T b \quad (\text{A.16})$$

R es no singular pues A es no singular, entonces, por (A.14)

$$\det(A) = \det(Q) \det(R) = \det(R)$$

Oviamente la solución de (A.15) es elemental desde el momento en que R es triangular superior.

A.5 Cálculo de $x = R^{-1} Qb$

Llamemos $b' = Qb$; sea $R(p \times p)$ triangular superior, con elementos $R(i, j)$ $i, j = 1 \dots p$, $b'(p \times 1)$ de elementos $b'(i)$ $i = 1 \dots p$, queremos hallar $x(p \times 1)$ de elementos $x(i)$ $i = 1 \dots p$ tal que

$$Rx = b' \text{ (se supone } \det(R) \neq 0)$$

El sistema es:

$$\begin{bmatrix} R(1,1) & \dots & R(1,p) \\ & \ddots & \\ & & R(p,p) \end{bmatrix} \begin{bmatrix} x(1) \\ \vdots \\ x(p) \end{bmatrix} = \begin{bmatrix} b'(1) \\ \vdots \\ b'(p) \end{bmatrix}$$

El algoritmo:

1) Conocidos $x(p) \dots x(i+1)$ se calcula $x(i)$ como

$$x(i) = \frac{b'(i) - R(i, i+1)x(i+1) - \dots - R(i, p)x(p)}{R(i, i)} \quad (\text{A.17})$$

Si $i=1$ se para, si $i > 1$ se repite para $i-1$.

2) Inicialización: $x(p) = \frac{b'(p)}{R(p, p)}$ (A.18)

El código generado en Pascal para resolver el problema es:

```
FOR I:=P DOWNTO 1 DO
BEGIN X(I) := B(I);
FOR J:=P DOWNTO (I+1) DO X(I) := X(I) - R(I,J)*X(J);
X(I) := X(I)/R(I,I);
END;
```

la suposición $\det(R) \neq 0$ puede verificarse facilmente comprobando que los elementos diagonales $R(i, i) \neq 0$ $i = 1 \dots p$.

A.6 Inversión de una matriz triangular superior

Sea $R(p \times p)$ una matriz triangular superior no singular (los elementos diagonales no nulos). Queremos calcular $S = R^{-1}$

Si R es triangular superior S también lo es. Sean $R = (r_{ij})$ $i, j = 1..p$

$S = (s_{ij})$ $i, j = 1..p$, entonces:

$$RS = I = \begin{bmatrix} r_{11} & \dots & r_{1p} \\ & \ddots & \vdots \\ 0 & & r_{pp} \end{bmatrix} \begin{bmatrix} s_{11} & \dots & s_{1p} \\ & \ddots & \vdots \\ & & s_{pp} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$\therefore \sum_{k=i}^p r_{ik} s_{kj} = \delta_{ij} \quad (A.19)$$

pero

$r_{ik} = 0$ si $i > k$; $s_{kj} = 0$ si $k > j$, luego (A.19) queda:

$$\sum_{k=i}^j r_{ik} s_{kj} = \delta_{ij} \quad (A.20)$$

para $\begin{cases} i = j & \rightarrow s_{ij} = \frac{1}{r_{ii}} \quad i = 1..p \\ i \neq j & \rightarrow s_{ij} = -\frac{1}{r_{ii}} \sum_{k=i+1}^j r_{ik} s_{kj} \end{cases} \quad (A.21)$

$$\begin{cases} i = j & \rightarrow s_{ij} = \frac{1}{r_{ii}} \quad i = 1..p \\ i \neq j & \rightarrow s_{ij} = -\frac{1}{r_{ii}} \sum_{k=i+1}^j r_{ik} s_{kj} \end{cases} \quad (A.22)$$

Se propone calcular primero la diagonal de S con (A.21). Conocida la diagonal calcular los elementos de S fila por fila comenzando desde la última, hacia arriba y de izquierda a derecha con A.22.

$$\begin{array}{cccccccc} s_{11} & \rightarrow & s_{12} & \rightarrow & s_{13} & \rightarrow & \dots & \rightarrow & s_{1p-1} & \rightarrow & s_{1p} \\ & & \uparrow & & \uparrow & & & & \uparrow & & \uparrow \\ & & s_{22} & \rightarrow & s_{23} & \rightarrow & \dots & & s_{2p-1} & \rightarrow & s_{2p} \\ & & & & \uparrow & & & & \uparrow & & \uparrow \\ & & & & s_{33} & \rightarrow & \dots & & s_{3p-1} & \rightarrow & s_{3p} \\ & & & & & & \ddots & & & & \uparrow \\ & & & & & & & & & & s_{p-1p-1} & \rightarrow & s_{p-1p} \\ & & & & & & & & & & & & \uparrow \\ & & & & & & & & & & & & s_{pp} \end{array}$$

El código generado en PASCAL es:

```
FOR I:=1 TO P DO S(I,I) := 1/R(I,I);
FOR I:=P-1 DOWNTO 1 DO
FOR J:=i+1 TO P DO
BEGIN: S(I,J) := 0;
FOR K:=I+1 TO J DO S(I,J) := S(I,J) + R(I,K)* S(K,I);
S(I,J) := -S(I,J)/R(I,I)
END;
```

El tema discutido en este apéndice puede verse en Farah, J.L (18), y

Stoer J. y Burlirsh R. (49)

- 2) Si $X_n \xrightarrow{\text{dist}} X$, $Y_n \xrightarrow{P} C \Rightarrow X_n + Y_n \xrightarrow{\text{dist}} X + C$
- $$X_n Y_n \xrightarrow{\text{dist}} CX$$
- $$\frac{X_n}{Y_n} \xrightarrow{\text{dist}} \frac{X}{C}, \text{ si } C \neq 0$$
- 3) Si $X_n \xrightarrow{P} Y_n$, $X_n \xrightarrow{\text{dist}} X \Rightarrow Y_n \xrightarrow{\text{dist}} X$.

La prueba puede encontrarse en [43] p 122-123.

TEOREMA B.2

Sea $g(y)$ una función continua de la variable y . Sea X_n una sucesión de variables aleatorias, entonces:

- 1) Si $X_n \xrightarrow{\text{dist}} X \Rightarrow g(X_n) \xrightarrow{\text{dist}} g(X)$
- 2) Si $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$

La demostración puede encontrarse en [43] pág 124.

TEOREMA B.3 (Chebishev, Una ley débil de los grandes números)

Sea $\{X_i\} i = 1, 2, \dots$ una sucesión de variables aleatorias tales que:

APENDICE 8

TEOREMAS DE CONVERGENCIA

Supuestas las definiciones de convergencia en probabilidad, con probabilidad 1, en media cuadrática y en distribución, dadas al comienzo del capítulo 2, pueden probarse muchos resultados sobre convergencia. En este apéndice se reproducen algunos particularmente importantes y útiles.

TEOREMA B.1

Sea (X_n, Y_n) una sucesión de parejas de variables aleatorias, entonces:

- 1) Si $X_n \xrightarrow{\text{dist}} X$, $Y_n \xrightarrow{P} 0 \Rightarrow X_n Y_n \xrightarrow{P} 0$

$E\{X_i\} = \mu_i$; $\text{var}\{X_i\} = \sigma_i^2$; $\text{cov}\{X_i, X_j\} = 0$ si $i \neq j$, entonces:

$$\sum_{i=1}^n \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 = 0 \Rightarrow \bar{X}_n - \bar{\mu}_n \xrightarrow{P} 0$$

donde:
$$\begin{cases} \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \\ \bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n \mu_i \end{cases}$$

TEOREMA B.4 (Khinchina, una ley débil de grandes números)

Sea $\{X_i\}$ $i=1, 2, \dots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas. Si existe su media $\mu = E\{X_i\}$, entonces:

$$E\{X_i\} = \mu \Rightarrow \bar{X}_n \xrightarrow{P} \mu$$

TEOREMA B.5 (Kolmogorov, una ley fuerte de grandes números)

1) Sea $\{X_i\}$ $i=1, 2, \dots$ una sucesión de variables aleatorias independientes, tales que $E\{X_i\} = \mu_i$ y $\text{var}\{X_i\} = \sigma_i^2$, entonces:

$$\text{Si } \sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < \infty \Rightarrow \bar{X}_n - \bar{\mu}_n \xrightarrow{\text{prob. } 1} 0$$

2) En las mismas condiciones que 1) salvo que en lugar de independencia se exige $\text{cov}\{X_i, X_j\} = 0$ $i \neq j$, entonces

$$\text{Si } \sum_{i=1}^{\infty} \frac{\sigma_i^2 (\log i)^2}{i^2} < \infty \Rightarrow \bar{X}_n - \bar{\mu}_n \xrightarrow{\text{prob. } 1} 0$$

TEOREMA B.6 (Kolmogorov, una ley fuerte de grandes números)

Sea $\{X_i\}$ $i=1, 2, \dots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas entonces:

$$E\{X_i\} = \mu \Rightarrow \bar{X}_n \xrightarrow{\text{prob. } 1} \mu$$

La demostración de los teoremas B.3 al B.6 puede encontrarse en [43] pp 112-116.

TEOREMA B.7 (Un teorema central del límite)

Sea $\{X_i\}$ $i=1, 2, \dots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, con media finita μ y varianza σ^2 , entonces:

$$Z_n \xrightarrow{\text{dist}} Z \quad \text{para } n \rightarrow \infty$$

$$\text{donde } \begin{cases} Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \\ Z \sim N(0, \sigma^2) \end{cases}$$

La demostración para el caso general donde X_i es vectorial y Σ es la matriz de covarianzas, puede ser encontrada en [26], pág. 223.

TEOREMA CENTRAL DEL LIMITE PARA MARTINGALAS

Sea un espacio de probabilidad (Ω, \mathcal{A}, P) :

Ω : espacio de eventos, no vacío

\mathcal{A} : una σ -álgebra de subconjuntos de Ω

P : una medida de probabilidad definida sobre \mathcal{A} .

Definición:

Una sucesión $(\mathcal{A}_i)_{i=1,2,\dots}$ de σ -álgebras de subconjuntos de Ω se dice creciente si:

- i) $\mathcal{A}_i \subset \mathcal{A}$
- ii) $\mathcal{A}_i \subset \mathcal{A}_j$ para $i \leq j$

Definición:

Un proceso estocástico $(X_i)_{i=1,2,\dots}$ se dice que es una martingala respecto a una sucesión de σ -álgebras \mathcal{A}_i ,

$i=1,2,\dots$ si:

i) (\mathcal{A}_i) es creciente

ii) (X_i) está adaptada a \mathcal{A}_i , esto es: $E(X_i | \mathcal{A}_i) = X_i$

iii) $E(|X_i|) < \infty$

iv) $E(X_i | \mathcal{A}_j) = X_j$ para $j < i$

TEOREMA B.B (Brown, teorema central del límite para martingalas)

Sea (S_n) , $n=1,2,\dots$ una martingala respecto la sucesión de σ -álgebras $(\mathcal{A}_n)_{n=1,2,\dots}$ en el espacio de probabilidad (Ω, \mathcal{A}, P) , con $S_0 = 0$, $X_n = S_n - S_{n-1}$, $n=1,2,\dots$. \mathcal{A}_0 no debe ser el σ -campo trivial (ϕ, Ω)

Llamemos:

$$\begin{aligned} \sigma_n^2 &= E_{n-1}(X_n^2) & (E_{n-1} \text{ indica } E(\cdot | \mathcal{A}_{n-1})) \\ V_n^2 &= \sum_{j=1}^n \sigma_j^2 \\ \delta_n^2 &= E V_n^2 = E S_n^2 \end{aligned}$$

si:

$$i) \quad V_n^2 S_n^{-2} \xrightarrow{P} 1$$

$$ii) \quad \delta_n^{-2} \sum_{j=1}^n E[X_j^2 \cdot I(|X_j| \geq c \delta_n)] \xrightarrow{P} 0$$

para todo $c > 0$, donde $I(B)$ es la función indicador del conjunto B ;

$$\begin{cases} I(x) = 1 & \text{si } x \in B \\ I(x) = 0 & \text{si } x \notin B \end{cases}$$

entonces se cumple que

$$\frac{S_n}{n} \xrightarrow{\text{d.i.s.t.}} N(0, 1) \quad \text{para } n \rightarrow \infty$$

La demostración se encuentra en [12].

El teorema anterior debió ser incluido en este trabajo para demostrar el siguiente teorema que completa la prueba del teorema 3.1 necesario para justificar una prueba sobre las estimaciones de la autocorrelación de una sucesión.

Como no fue encontrada en la literatura una demostración que fundamente dicha prueba (de la sección 3.3.2.), hubo que hacerla.

Esta demostración basada en el teorema anterior (B.8) fue hecha en colaboración y con la orientación del Dr. José L. Farrah, sin cuya ayuda hubiera resultado una tarea imposible.

TEOREMA B.9

Sea $x(\ell)$, $\ell=1, 2, \dots$ una sucesión de v.a. idénticamente distribuidas, con media cero y varianza σ^2 .

Sea la sucesión $\hat{R}_{xx}^N(\tau) = \frac{1}{N} \sum_{\ell=1}^N x(\ell) x(\ell+\tau)$, entonces, si

$E\{x(\ell) x(j)\} = 0 \quad \forall \ell \neq j$, se cumple que

$$\hat{R}_{xx}^N(\tau) \xrightarrow{\text{L.C.V.}} N(0, \frac{\sigma^4}{N}) \quad \text{para } N \rightarrow \infty$$

Demostración:

Llamemos S_N a la sucesión $\hat{R}_{xx}^N(\tau)$, $N=1, 2, 3, \dots$ es decir

$$S_N = \sum_{\ell=1}^N [x(\ell) x(\ell+\tau)] \quad , \tau \neq 0$$

Sea \mathcal{A}_N la sucesión de σ -álgebras generadas por las realizaciones $x(1) \dots x(N) \quad x(1+\tau) \dots x(N+\tau)$:

$$\mathcal{A}_N = \sigma\{x(1), \dots, x(N), x(1+\tau), \dots, x(N+\tau)\}$$

Puede verificarse que S_N es una martingala respecto a la sucesión de σ -álgebras \mathcal{A}_N :

i) $\{\mathcal{A}_\ell\}$ es creciente

$$\text{ii) } E\{S_j | \mathcal{A}_j\} = E\left\{ \sum_{\ell=1}^j x(\ell) x(\ell+\tau) \mid x(1), \dots, x(j), x(1+\tau), \dots, x(j+\tau) \right\} = S_j$$

\therefore está adaptada a \mathcal{A}_ℓ

$$\text{iii) } E\{|S_\ell|\} < \infty \quad , \quad \text{pues } E\{|S_\ell|\} = \ell \sigma^2$$

iv) $E\{S_k | A_j\} = S_j$ para $j < k$ pues:

$$E\{S_k | A_j\} = E\left\{ \sum_{\ell=1}^k x(\ell)x(\ell+\tau) \mid x(1), \dots, x(j), x(j+\tau), \dots, x(j+\tau+\ell) \right\}$$

$$= E\left\{ S_j + \sum_{\ell=j+1}^k x(\ell)x(\ell+\tau) \mid A_j \right\} = S_j + \sum_{\ell=j+1}^k E\{x(\ell)x(\ell+\tau) | A_j\}$$

$$\text{pero } E\{x(\ell)x(\ell+\tau) | A_j\} = \begin{cases} 0 & \text{si } \ell \leq \tau \text{ pues } E\{x(\ell)x(\ell+\tau)\} = 0 \\ x(\ell)E\{x(\ell+\tau)\} = 0 & \text{si } \ell > \tau \text{ pues } E\{x(\ell+\tau)\} = 0. \end{cases}$$

$$\therefore E\{S_k | A_j\} = S_j$$

Ahora comprobaremos que está en las condiciones del teorema anterior.

$$\text{Llamemos } \begin{cases} X_N = S_N - S_{N-1} = x(N)x(N+\tau) \\ \sigma_N^2 = E_{N-1}\{X_N^2\} \\ V_N^2 = \sum_{j=1}^N \sigma_j^2 \\ \delta_N^2 = E V_N^2 \end{cases}$$

Calculemos σ_N^2 :

$$\sigma_N^2 = E\{X_N^2 | A_{N-1}\}$$

$$\therefore \sigma_N^2 = E\{x^2(N)x^2(N+\tau) | A_{N-1}\}$$

podemos distinguir dos casos:

$$1) \text{ Si } \tau \geq N$$

$$2) \text{ Si } \tau < N$$

$$1) \text{ Si } \tau \geq N \text{ entonces } x(N) \in a_{N-1}, x(N+\tau) \in a_{N-1}$$

$$\text{donde } a_{N-1} = \{x(1) \dots x(N-1), x(1+\tau) \dots, x(N+\tau-1)\}$$

$$\therefore \text{ es } \sigma_N^2 = \sigma^4$$

$$2) \text{ Si es } \tau < N \text{ entonces } x(N) \in a_{N-1}, x(N+\tau) \notin a_{N-1}$$

$$\text{luego se cumple que } \sigma_N^2 = x^2(N) \sigma^2$$

Calculemos ahora V_N^2 :

$$V_N^2 = \sum_{j=1}^N \sigma_j^2 = \begin{cases} N \sigma^4 & \text{si } N \leq \tau \\ \tau \sigma^4 + \sum_{\ell=\tau+1}^N x^2(\ell) \sigma^2 & \text{si es } N > \tau \end{cases}$$

Calculemos δ_N^2 :

$$\delta_N^2 = E\{V_N^2\} = E\left\{ \sum_{j=1}^N \sigma_j^2 \right\} = \begin{cases} = N \sigma^4 & \text{si } N \leq \tau \\ = \tau \sigma^4 + \sum_{\ell=1}^N E\{x(\ell)^2\} \sigma^2 = N \sigma^4 & \text{si } N > \tau \end{cases}$$

$\therefore S_N^2 = N\sigma^4$ no importando si $N \leq \tau$ ó $N > \tau$.

Verifiquemos ahora la condición (4) del teorema B.8:

$$V_N^2 S_N^{-2} = \frac{\tau\sigma^2 + \frac{N}{\sum_{i=\tau+1}^N x(i)^2 \sigma^2}}{N\sigma^4} = \frac{\tau + \frac{1}{\sigma^2} \frac{N}{\sum_{i=\tau+1}^N x(i)^2}}{N}$$

por el teorema B.4: $\frac{1}{N-\tau} \frac{N}{\sum_{i=\tau+1}^N x(i)^2} \xrightarrow{P} \sigma^2$, y por el teorema B.1:

$$V_N^2 S_N^{-2} \xrightarrow{P} 1, \text{ lo que verifica la condición (4)}$$

Comprobemos la condición (1):

debemos probar que $\frac{1}{N\sigma^4} \sum_{j=1}^N E(X_j^2 I(|X_j| \geq c\sigma)) \xrightarrow{P} 0 \quad \forall c > 0$

esto es que $\frac{1}{N\sigma^4} \sum_{j=1}^N E(X_j^2 I(|X_j| \geq c\sigma)) \xrightarrow{P} 0 \quad \forall c > 0$

pero $\frac{N}{\sum_{j=1}^N} E(X_j^2 I(|X_j| \geq c\sigma)) = N \int x^2 I(|x| > c\sigma) dP = J$

donde $x = x(i) \quad x(i+\tau) \quad x_i \quad (i=1, \dots, N)$ son idénticamente distribuidos con la misma densidad dP .

pero la cantidad J es:

$$J = N \int_{|x| > c\sigma} x^2 dP = N \left[\int_{|x| > c\sigma} x^2 dP - \int_{|x| < c\sigma} x^2 dP \right]$$

como $\int_{|x| < c\sigma} x^2 dP \xrightarrow{P} E(X^2) \quad \forall \epsilon > 0 \quad \text{y} \quad N \rightarrow \infty$

Entonces:

$$\frac{1}{N\sigma^4} \sum_{j=1}^N E(X_j^2 I(|X_j| \geq c\sigma)) = \frac{N}{N\sigma^4} \left[E(X^2) - \int_{|x| < c\sigma} x^2 dP \right] = \frac{1}{\sigma^4} \left[E(X^2) - \int_{|x| < c\sigma} x^2 dP \right] \xrightarrow{P} 0$$

para todo $\epsilon > 0$ y $N \rightarrow \infty$

esto verifica la segunda condición.

Entonces, en virtud del teorema B.8, $\frac{S_N}{N} \xrightarrow{P} \mathcal{N}(0, 1)$, si $N \rightarrow \infty$

$$\begin{aligned} \frac{\frac{N}{\sum_{i=1}^N} x(i) x(i+\tau)}{N\sigma^4} &\xrightarrow{P} \mathcal{N}(0, 1) \\ \therefore R_{xx}^N(\tau) = \frac{\frac{N}{\sum_{i=1}^N} x(i) x(i+\tau)}{N} &\xrightarrow{P} \mathcal{N}(0, \frac{\sigma^4}{N}) \quad \text{para} \quad N \rightarrow \infty \\ &\quad \tau \neq 0 \end{aligned}$$

que es lo que queríamos demostrar.

A continuación se muestran, señal por señal, gráficas de sus primeros elementos, semilla de generación (cuando corresponde), y los resultados a las pruebas de independencia y distribución.

RUIDO 1, generado con semilla = 25



PRIMEROS 150 PUNTOS DE LA SERAL RUIDO1.-

Pruebas de descorrelación

1) Cruces por cero. (Con un nivel de confianza de 95%).

NO. DE INTERVALOS : 444
 NO. DE CRUCES POR CERO: 331
 % DE CRUCES POR CERO: 0.514
 % ADMISIBLE : 0.461 - 0.539

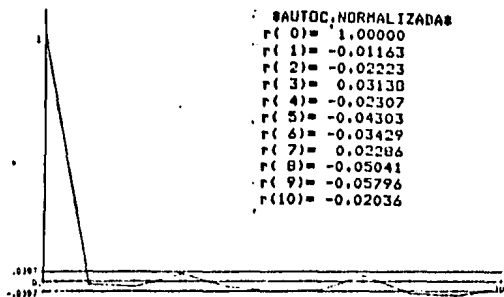
Se acepta.

APENDICE C. SEREALES SIMULADAS DE RUIDO BLANCO

Se simularon 10 señales con características que se detallan en la tabla

NOMBRE	TIPO DE DISTRIBUCION	LARGO	METODO DE GENERACION	PROGRAMA DE GEN.
RUIDO 1	Normal (0.1)	645	Box - Muller	ALEATORIO
RUIDO 2	Normal (0.1)	645	Box - Muller	ALEATORIO
RUIDO 3	Normal (0.1)	645	Box - Muller	ALEATORIO
RUIDO 4	Normal (0.1)	645	Box - Muller	ALEATORIO
RUIDO 5	Normal (0.1)	1127	Box - Muller	ALEATORIO
RUIDO 6	Normal (0.1)	310	Box - Muller	ALEATORIO
RUIDO 7	Uniforme en (-0.5,0.5)	645	Congruencial combinado	ALEATORIO
RND 1	Binaria, aleatoria a int. disc.	310	A partir de uniforme	RUIDO/BINARIO
PRBS 63	Sucesión PRBS largo máximo	325	Ecuación de diferencias mtd. 2	PRBS/63
PRBS 127	Sucesión PRBS largo máximo	1127	Ecuación de diferencias mtd. 2	PRBS/127

TABLA CI



ESTIMACION DE LA FUNCION DE AUTOCORRELACION DE LA SEÑAL RUIDO1.

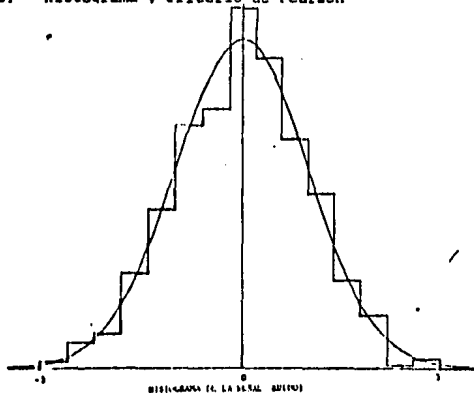
VAR. TEORICA DE LOS $r(1/N)$ 0.00157
 D.C.M. TEORICA DE LOS $r(1/N)$ 0.03968
 VAR. DE LA MUESTRA DE $r(1/N)$ 0.00120
 D.C.M. DE LA MUESTRA DE $r(1/N)$ 0.03466
 r(k) MAXIMO ADMISIBLE! 0.11111
 D.C.M. DE r(k) ADMISIBLE! 0.05600
 0.05369

Pruebas de distribución

1) Media y varianza:

** MEDIA! 0.00481
 RANGO ADMISIBLE! -0.07717 0.07717
 ** VARIND CORREG.!! 0.97453
 RANGO ADMISIBLE! 0.89086 1.10914
 ** DCMINO CORREG.!! 0.90710

2) Histograma y criterio de Pearson



Se consideraron 13 intervalos del eje real, contiguos y disjuntos, con límites: $-\infty, -2.2, -1.8, -1.4, -1, -0.6, -0.2, 0.2, 0.6, 1, 1.4, 1.8, 2.2, +\infty$

Con un nivel de confianza de 95%, para 13 intervalos se acepta un valor de $x \leq 21.026$

Se observó $x = 8.22$ se acepta

RUIDO 2

Generado con una semilla: 2097147

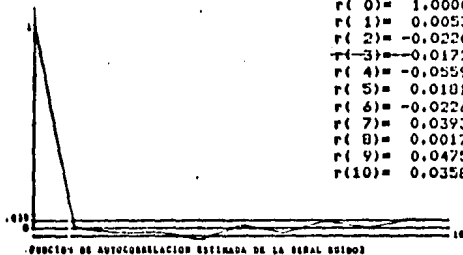
Pruebas de descorrelación

1) Cruces por cero: (con nivel de confianza de 95%)

NU. DE INTERVALOS 1 444
 NO. DE CRUCES POR CERO! 312
 % DE CRUCES POR CERO! 0.484
 % ADMISIBLE! 0.461 - 0.539

2) Autocorrelación

\$AUTOC.NORMALIZADA\$
 r(0) = 1.00000
 r(1) = 0.00537
 r(2) = -0.02208
 r(3) = -0.01725
 r(4) = -0.05595
 r(5) = 0.01019
 r(6) = -0.02269
 r(7) = 0.03937
 r(8) = 0.00179
 r(9) = 0.04754
 r(10) = 0.03580



\$CORRELACIONES NORMALIZADAS r\$

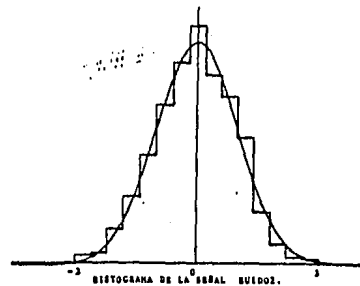
VAR. TEORICA DE LOS r(1/N) 0.00157
 D.C.M. TEORICA DE LOS r(SQRT(1/N)) = 0.03968
 VAR. DE LA MUESTRA DE 10 r = 0.00099
 D.C.M. DE LA MUESTRA DE R = 0.03144
 r(k) MAXIMO ADMISIBLE: 0.11111 (CONF=95%)
 D.C.M. DE r(k) ADMISIBLE: 0.05680 (CONF=97.5%)
 0.05369 (CONF=95%)

Pruebas de distribución:

1) Media y varianza

** MEDIA: -0.05278
 RANGO ADMISIBLE: -0.07717 0.07717
 ** VARIANZA CORREG. **) 0.99507
 RANGO ADMISIBLE: 0.89086 1.10914
 ** RANGO CORREG. **) 0.99793

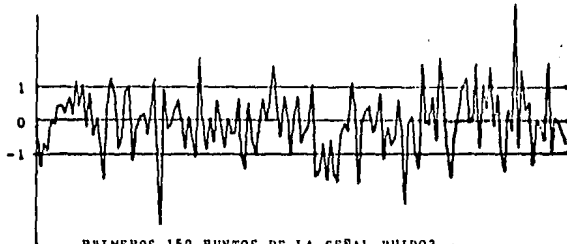
2) Histograma y prueba de Pearson.



Se consideraron 13 intervalos para la prueba de Pearson, con límites como en ruido 1. Con un nivel de confianza de 95% se acepta $x \leq 21.026$. Se observó $x = 8.4$. Se acepta.

RUIDO 3

Se generó con semilla = 524299.

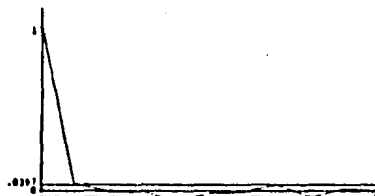


Pruebas de descorrelación

1) Cruces por cero. (con un nivel de confianza de 95%).

NO. DE INTERVALOS : 644
 NO. DE CRUCES POR CERO: 325
 % DE CRUCES POR CERO: 0.505
 % ADMISIBLE : 0.461 - 0.539

2) Autocorrelación



AUTOC.NORMALIZADAS

r(0) = 1.00000
 r(1) = 0.04980
 r(2) = 0.00197
 r(3) = 0.00482
 r(4) = -0.05744
 r(5) = -0.00363
 r(6) = 0.00382
 r(7) = 0.04574
 r(8) = -0.03176
 r(9) = 0.02353
 r(10) = 0.00399

VAR.TEORICA DE LOS r(i)/N 0.00157
 D.C.M.TEORICA DE LOS r(i)/N = 0.03960
 VAR. DE LA MUESTRA DE 10 r = 0.00095
 D.C.M. DE LA MUESTRA DE N = 0.03084
 r(k) MAXIMO ADMISIBLE: 0.11111 (CONF=95%)
 D.C.M. DE r(k) ADMISIBLE: 0.05680 (CONF=97.5%)
 0.05369 (CONF=95%)

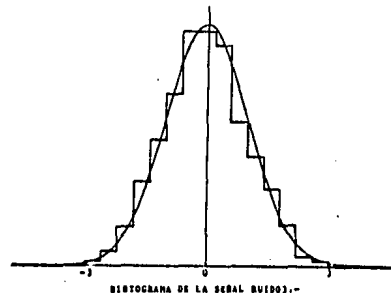
Pruebas de distribución

1) Media y varianza

Para un nivel de confianza 95%

** MEDIA: -0.04078
 RANGO ADMISIBLE: -0.07717 0.07717
 ** VAR(NO CORREG.): 1.06488
 RANGO ADMISIBLE: 0.89086 1.10914
 ** DCM(NO CORREG.): 1.03193

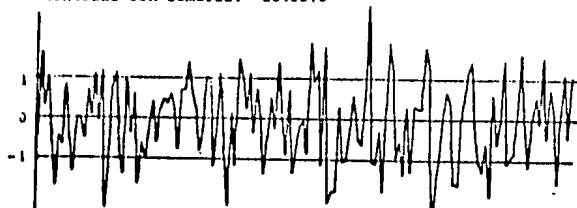
2) Histograma y prueba de Pearson.



Se consideraron 13 intervalos, como en los casos anteriores.
 Con un nivel de confianza de 95%, se acepta $x < 21.026$, se
 observa $x = 5.758$, se acepta.

RUIDO 4

Generado con semilla: 1048575



PRIMEROS 150 PUNTOS DE LA SEÑAL RUIDO4

Pruebas de descorrelación:

1) Cruces por cero: (con un nivel de confianza de 95%).

NO. DE INTERVALOS : 644
 NO. DE CRUCES POR CERO: 326
 λ DE CRUCES POR CERO: 0.506
 λ ADMISIBLE : 0.461 - 0.539

2) Autocorrelación

VAR. TEORICA DE LOS $r(1/N)$ 0.00157
 D.C.M. TEORICA DE LOS $r(1/N)$ 0.03960
 VAR. DE LA MUESTRA DE 10 r 0.00109
 D.C.M. DE LA MUESTRA DE r 0.03307
 $|r(k)|$ MAXIMO ADMISIBLE: 0.11111 (CONF=95%)
 D.C.M. DE $r(k)$ ADMISIBLE: 0.05300 (CONF=97.5%)
 0.05369 (CONF=95%)

#AUTOC. NORMALIZADAS#

$r(0) = 1.00000$
 $r(1) = -0.02272$
 $r(2) = -0.01634$
 $r(3) = -0.04469$
 $r(4) = -0.00637$
 $r(5) = -0.04233$
 $r(6) = 0.07049$
 $r(7) = 0.00903$
 $r(8) = 0.02150$
 $r(9) = -0.01822$
 $r(10) = 0.02184$

FUNCION DE AUTOCORRELACION ESTIMADA DE LA SEÑAL RUIDO4.

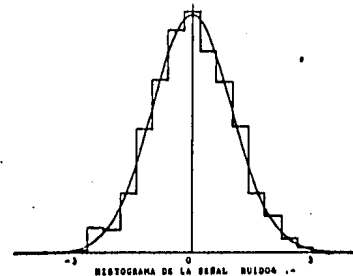
Pruebas de distribución:

1) Media y varianza

Con un nivel de confianza de 95%

** MEDIA: -0.01490
 RANGO ADMISIBLE: -0.07717 0.07717
 ** VAR(NO CORREG.): 1.01254
 RANGO ADMISIBLE: 0.89086 1.10914
 ** DCM(NO CORREG.): 1.00625

2) Histograma y prueba de Pearson:



Se consideraron 13 intervalos como en los casos anteriores, con un nivel de confianza de 95%, se admite $\alpha < 21.026$, se observa $\alpha = 6.381$, se acepta.

RUIDO 5

Generado con semilla = 2097153

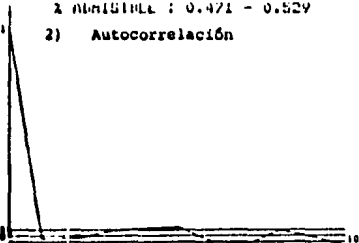
Pruebas de descorrelación:

}) Cruces por cero. (95% de confianza)

NO. DE INTERVALOS : 1126
 NO. DE CRUCES POR CERO: 572
 Z DE CRUCES POR CERO: 0.500
 Z ADMISIBLE : 0.471 - 0.529

2) Autocorrelación

AUTOCORRELACION NORMALIZADA
 r(0) = 1.00000
 r(1) = -0.01130
 r(2) = -0.00761
 r(3) = 0.02760
 r(4) = 0.03913
 r(5) = 0.04399
 r(6) = -0.02560
 r(7) = -0.04034
 r(8) = 0.02917
 r(9) = -0.00459
 r(10) = -0.04719



FUNCION DE AUTOCORRELACION ESTERADA DE LA SEÑAL RUIDO5.

VAR. TEORICA DE LOS r(1)/N 0.00070
 D.C.M. TEORICA DE LOS r(1)/N = 0.02992
 VAR. DE LA MUESTRA DE 10 r = 0.00090
 D.C.M. DE LA MUESTRA DE R = 0.03130
 |r(k)| MAXIMO ADMISIBLE: 0.08378 (CONF=95%)
 D.C.M. DE r(k) ADMISIBLE: 0.04282 (CONF=97.5%)
 0.04048 (CONF=95%)

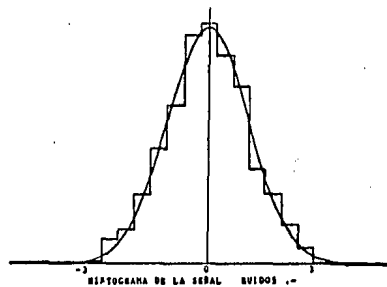
Pruebas de distribución:

1) Media y varianza

Con un nivel de confianza de 95%

** MEDIA: -0.01399
 RANGO ADMISIBLE: -0.05038 0.05838
 ** VAR(ND CORREG.): 1.02937
 RANGO ADMISIBLE: 0.91743 1.08257
 ** DCM(ND CORREG.): 1.01458

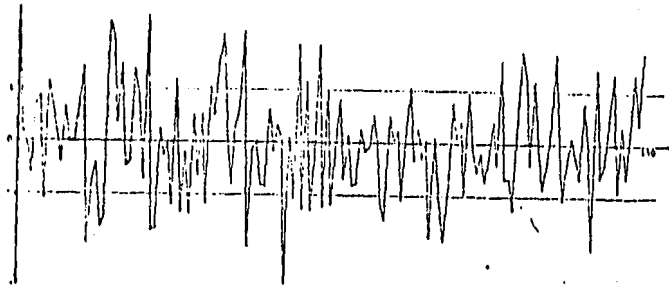
2) Histograma y prueba de Pearson:



Si consideraron los mismos 13 intervalos que en los casos anteriores con un nivel de confianza de 95% debe ser $x < 21.026$, se observa $x = 6.94$, se acepta.

RUIDO 6.

Generado con semilla: 524287

Pruebas de descorrelación

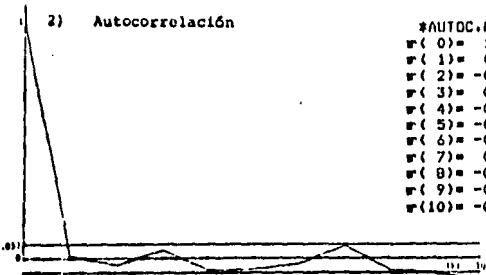
1) Cruces por cero (nivel de confianza = 95%)

NO. DE INTERVALOS : 309
 NO. DE CRUCES POR CERO: 161
 % DE CRUCES POR CERO: 0.521
 % ADMISIBLE : 0.412 - 0.550

2) Autocorrelación

AUTOCORRELACION NORMALIZADA

r(0)	=	1.00000
r(1)	=	0.00704
r(2)	=	-0.02890
r(3)	=	0.03315
r(4)	=	-0.04914
r(5)	=	-0.04321
r(6)	=	-0.02249
r(7)	=	0.05204
r(8)	=	-0.04657
r(9)	=	-0.06171
r(10)	=	-0.09440



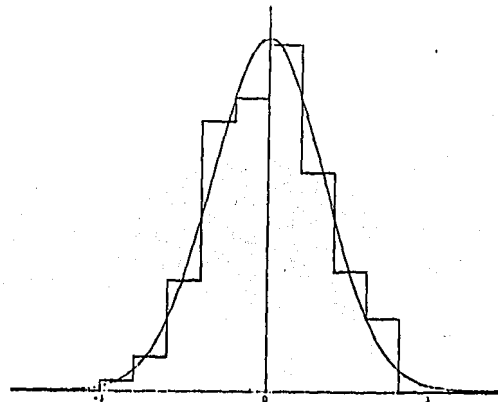
VAR. TEORICA DE LOS $r(1)/N$ 0.00333
 D.C.M. TEORICA DE LOS $r(1)/\sqrt{N}$ = 0.05774
 VAR. DE LA MUESTRA DE 10 $r=$ 0.00245
 D.C.M. DE LA MUESTRA DE $r=$ 0.04945
 $r(k)$ MAXIMO ADMISIBLE: 0.16166 (CONF=95%)
 D.C.M. DE $r(k)$ ADMISIBLE: 0.08263 (CONF=97.5%)
 0.07812 (CONF=95%)

Pruebas de distribución:

1) Media y varianza

** MEDIA: 0.04605
 RANGO ADMISIBLE: -0.11132 0.11132
 ** VAR(NO CORREG.): 0.99104
 RANGO ADMISIBLE: 0.84257 1.15743
 ** DCM(NO CORREG.): 0.99551

2) Histograma y prueba de Pearson:

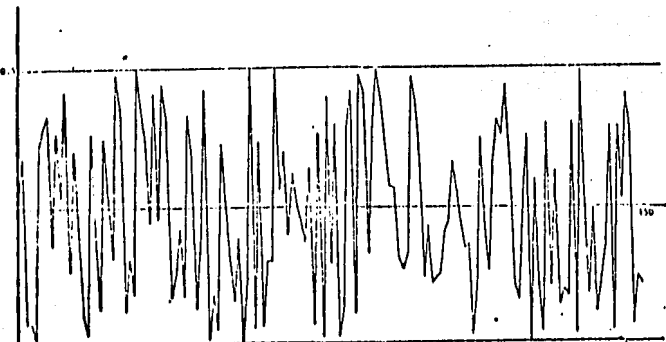


Se consideraron 8 intervalos, con límites: $-\infty, -2.4, -1.8, -1.2, -0.6, 0.6, 1.2, 1.8, 2.4, \infty$. Con un nivel de confianza de 95% debe ser $\alpha \leq 14.1$.

El cálculo es $\alpha = 4.108$, se acepta.

RUIDO 7

Generado con semilla: 1048223

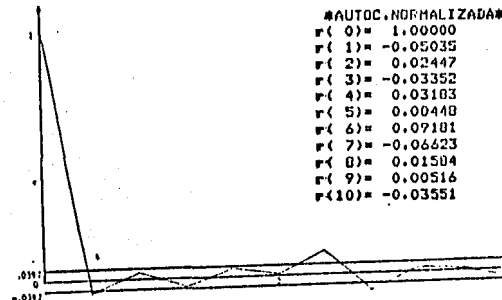


Pruebas de desc correlación:

1) Cruces por cero (con nivel de confianza: 95%).

NO. DE INTERVALOS : 644
 NO. DE CRUCES POR CERO: 334
 % DE CRUCES POR CERO: 0.519
 % ADMISIBLE: 0.461 - 0.539

2) Autocorrelación.



#AUTOCORRELACION#
 r(0)= 1.00000
 r(1)= -0.05035
 r(2)= 0.02447
 r(3)= -0.03352
 r(4)= 0.03183
 r(5)= 0.00448
 r(6)= 0.07181
 r(7)= -0.06623
 r(8)= 0.01504
 r(9)= 0.00516
 r(10)= -0.03551

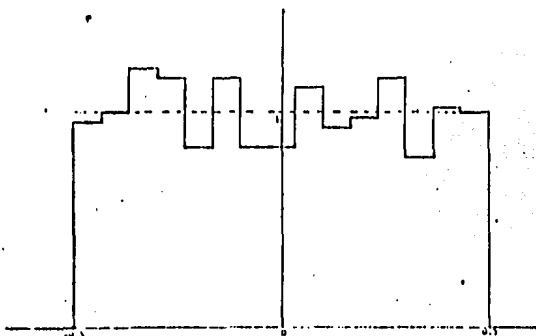
VAR. TEORICA DE LOS $r(1)/N$ 0.00157
 D.C.M. TEORICA DE LOS $r(1)/N$ = 0.03968
 VAR. DE LA MUESTRA DE 10 r = 0.00196
 D.C.M. DE LA MUESTRA DE k = 0.04432
 $|r(k)|$ MAXIMO ADMISIBLE: 0.1111 (CONF=95%)
 D.C.M. DE $r(k)$ ADMISIBLE: 0.05660 (CONF=97.5%)
 0.05369 (CONF=95%)

Pruebas de distribución:

1) Media y varianza

MEDIA: -0.00432
 RANGO ADMISIBLE: -0.02228 0.02228
 ## VAR(IND CORREG.): 0.08406
 RANGO ADMISIBLE: 0.07758 0.08909
 ## DCN(IND CORREG.): 0.28993

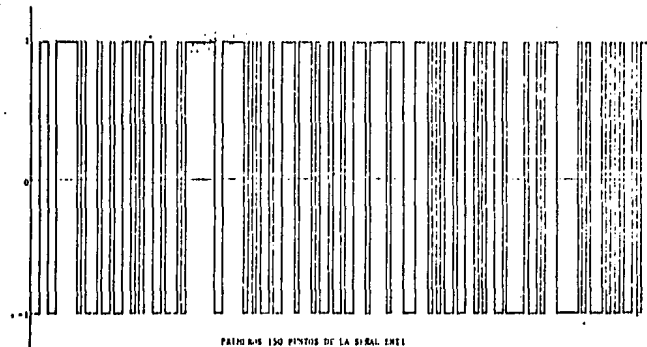
2) Histograma y prueba de Pearson.



Se consideraron 15 intervalos idénticos en (-0.5, 0.5). Para un nivel de confianza de 95% se admite $\alpha < 23.685$, el χ^2 cuyo arroja $\alpha = 11.535$, se acepta.

ENT 1

Generada a partir de uniforme con semilla = 12.

Pruebas de descorrelación:

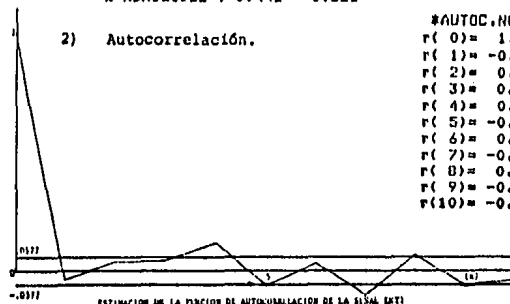
1) Cruces por cero (con nivel de confianza = 95%)

NO. DE INTERVALOS : 309
 NO. DE CRUCES POR CERO: 163
 % DE CRUCES POR CERO: 0.528
 % ADMISIBLE : 0.442 - 0.558

2) Autocorrelación.

#AUTOC. NORMALIZADAS

r(0) = 1.00000
 r(1) = -0.04000
 r(2) = 0.03133
 r(3) = 0.04000
 r(4) = 0.11333
 r(5) = -0.06667
 r(6) = 0.02667
 r(7) = -0.10667
 r(8) = 0.06000
 r(9) = -0.06667
 r(10) = -0.04000



VAR. TEORICA DE LOS $r(k)/N$ 0.00333
 D.C.M. TEORICA DE LOS $r(k)/N$ 0.05774
 VAR. DE LA MUESTRA DE 10 r 0.00433
 D.C.M. DE LA MUESTRA DE r 0.06503
 $r(k)$ MAXIMO ADMISIBLE! 0.16166 (CONF=95%)
 D.C.M. DE $r(k)$ ADMISIBLE! 0.08263 (CONF=97.5%)
 0.07812 (CONF=95%)

Pruebas de distribución:

1) Media y varianza

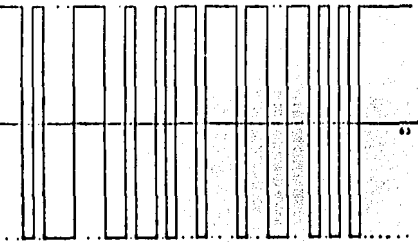
Para un nivel de confianza de 95% se exige $|X| \leq 0.11132$

(la prueba de varianza no hace sentido en este caso)

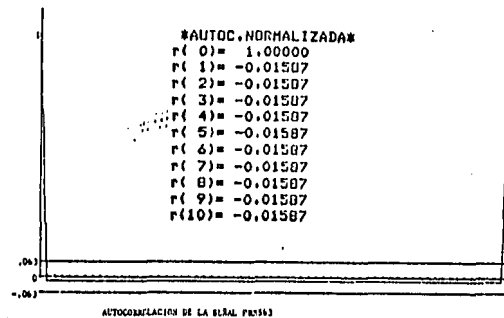
Se calcula, $X = -0.01935$, se acepta.

PRBS 63

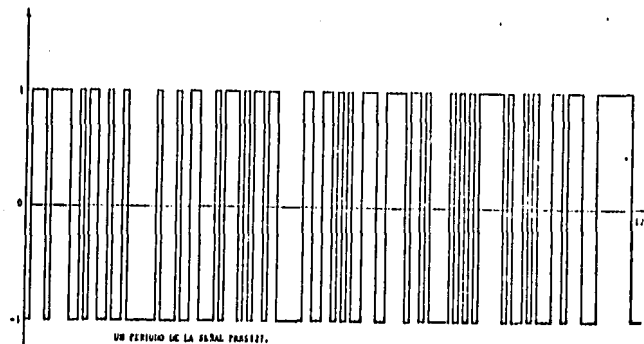
Tanto para esta señal como para PRBS 127 fueron estudiadas sus propiedades en la sección (3.2.4), por lo que no se practican las pruebas. Se presentan las gráficas de un período de la señal y su autocorrelación.



UN PERIODO DE LA SEÑAL PRBS63



PRBS 127



UN PERIODO DE LA SEÑAL PRBS127.

AUTOC-NORMALIZADA

r(0)= 1.00000
r(1)= -0.00707
r(2)= -0.00707
r(3)= -0.00707
r(4)= -0.00707
r(5)= -0.00707
r(6)= -0.00707
r(7)= -0.00707
r(8)= -0.00707
r(9)= -0.00707
r(10)= -0.00707

0417

AUTOCORRELACION DE LA SEÑAL P33127

APENDICE D

En este apéndice se presentan los listados de los programas más relevantes utilizados en la elaboración de este trabajo. En cada listado se incorpora una breve explicación a modo de "autodocumentación", de la función de cada programa.

Programa	Página
NICU/C	561
NICU/S	562
NICU/S/OLVEXP	563
NICU/S/OLVVAR	564
NICU/S/VERSION3	566
NICU/GENE1	567
NICU/GENE2	569
NICU/GENE3	572
NICU/GENE4	573
PRBS/63	574

PRBS/127	574
HERTO	575
SROOT/C	575
SROOT/S	577
SROOT/IMP/S	580
CRUCES/CERO	582
ESTINA/PISO	582
SIMULA	583
VARIANZA/CALIDA	584
ALLATORIO	584
ELIMPOIDE	585
CORRELACION	585
CORRELACION/CRUZADA	586
FUNCION PESO	586
RUIDOBINARIO	587

REFERENCIAS

1. ACKERMANN, J., "Abstatregelung". Springer Verlag, Berlín. Traducción al español en proceso, A. Estandía y J.L. Farah. IINAS-UNAM, México.
2. ASTROM, K.J., "On the achievable accuracy in identification problems". Proc. IFAC Symp. on Identification in Automatic Control Systems. Artículo 1.8, Praga 1967.
3. ASTROM, K.J., "Lectures on the identification problem—the least squares method". Reporte 6806, Septiembre 1968. Lund Institute of Tech. Division of Automatic Control. Lund, Suecia.
4. ASTROM, K.J. "Introduction to Stochastic Control Theory". Academic Press, 1970.
5. ASTROM, K.J., "Maximum likelihood and prediction error methods". Automatica Vol. 16, pp 571-574. Pergamon Press Ltd., Great Britain 1980.
6. ASTROM, K.J., y BOHLIN, T. "Numerical identification of linear dynamic systems from normal operating records". Proc. 2nd IFAC Symp. on the theory of self-adaptive control systems. National Physics Lab., Teddington England, 1965.
7. ASTROM, K.J., BOHLIN, T., WENSMARK, S., "Automatic construction of linear stochastic models for stationary industrial processes with random disturbances using operating records". IBM Nordik Lab, (reporte TP 18.150) Suecia 1965.
8. ASTROM, K.J., EYKHOFF, P., "System Identification—a survey". Automatica Vol. 7 pp 123-167. 1971.
9. ASTROM, K.J., WITTENMARK, B., "Problems of identification and control". Journal of Mathematical Analysis and Applications Vol. 34, pp 90-113. 1971.
10. BARTLETT, M.S., "On the theoretical specification and sampling properties of autocorrelated time-series". Journal Royal Statistical Soc. B8, 27. 1946.
11. DOX G., MULLER, M., "A note on the generation of random normal deviates". The Annals of Mathematical Statistics, 1958, No. 28 pp 610-611.
12. BROWN, B.M., "Martingale Central limit theorems". The Annals of Mathematical Statistics, 1971, Vol. 42-1 pp 59-66.

13. BURROUGHS "B6800 Mathematical Intrinsic".
14. CADZON, J.A., "Discrete time systems". Prentice-Hall New Jersey, 1970.
15. CLARCKE, D.W., "Generalized least squares estimation of parameters of a dynamic model". 1st IFAC Symposium on Identification in Automatic Control Systems. Praga 1967.
16. EYKHOFF, P., "System Identification". Wiley, Londres 1974.
17. FARAH, J.L., "Notas del Seminario de Control". IIMAS-UNAM. (sin publicar), México, 1980.
18. FARAH, J.L., "Notas del curso de Análisis Numérico". IIMAS-UNAM (sin publicar), México, 1981.
19. FASOL, K.H., JORGL, H.P., "Principles of model building and Identification". Automática Vol. 16, pp 505-518, Gran Bretaña, 1980.
20. FISHER, R., YATES, F., "Statistical tables for biological, agriculture and medical research". Oliver & Boyd, Londres, 1938.

21. FURHT, B.P., "New estimator for the identification of dynamic processes". Institut "BORIS KIDRIČ", Belgrado, 1973.
22. FURHT, B.P., "Maximum likelihood identification of Astrom model by quasilinearization". Proc. 3rd IFAC Symp. The Hague/Delft, Países Bajos, 1973.
23. GAUSS, K.F., "Theory of the motion of the heavenly bodies". ("Theoria motus corporum coelestium") Dover Pub. Inc., New York, 1969.
24. GMURMANN, V.E., "Teoría de las probabilidades y la estadística matemática". Editorial Mir, Moscú, 1974.
25. GODFREY, K.R., "Correlation Methods". Automatica Vol. 16, pp 527-534. Gran Bretaña, 1980.
26. GOODWIN, G.C., PAYNE, R.L., "Dynamic system Identification: Experimental design and data analysis". Academic Press, New York, 1977.
27. GRAUPE, D., "Identification of systems". Van Nostrand, New York, 1972.

28. HASTINGS-JAMES, R., SAGE, M.W., "Recursive generalized least squares procedure for on-line identification of process parameters". IEE Proceedings, Vol. 116, No. 12, pp 2057-2062, 1969.
29. HUZURBAZAR, V.S., "Sufficient statistics, selected contributions". Marcel Dekker Inc., New York, 1976.
30. I.M.E., "Notas de estadística". Instituto de Matemática y Estadística, Facultad de Ingeniería y Agrimensura, Montevideo, 1971.
31. ISERMANN, R., BAUR, V., DAMBERGER, W., KNEPPO, P., SIEBERT, H., "Comparison of six on-line identification and parameter estimation methods with three simulated processes". Proceedings of 3rd IFAC Symp. on Identification and System Parameter Estimation. The Hague/Delft, Países Bajos, 1973.
32. JENSEN, K., WIRTH, N., "Pascal User Manual". Springer-Verlag.
33. KENDALL, M.G., STUART, A., "The advanced theory of statistics". Vol. II. Charles Griffin & Co., Londres, 1961.

34. KUO, B., "Discrete Data Control Systems". Prentice Hall, New Jersey, 1973.
35. LJUNG, L., "On the consistency of prediction error identification methods". En "System Identification-Advances and Case Studies". MEHRA, R.K., LAINIOTIS, D. G., Editores. Academic Press, 1976.
36. LJUNG, L., "Analysis of recursive stochastic algorithms". IEE Trans. on Automatic Control, Vol. AC-22, pp 551-575. 1977.
37. MOOD, A., GRAYBILL, F., BOES, D., "Introduction to the theory of Statistics". Mc Graw-Hill Co., Tokio, 1974.
38. NEKOLNY, J., BENES, J., "Simultaneous Control of Stability and Quality of Adjustment-Applications of Statistical Dynamics". Proc. 1st IFAC Cong., Moscú, 1960.
39. NEWMAN, T.G., ODELL, P.L., "The generation of random variables". Charles Griffin Co., Londres 1971.
40. PAPOULIS, A., "Probability, random variables and stochastic processes". Mc Graw Hill - Kogakusha, Tokio, 1965.

41. PLACKETT, R.L., "Some theorems in least squares". Biometrika, Vol. 37, p 149. 1950.
42. PUGACHEV, P.S., "Introducción a la teoría de las probabilidades". Editorial Mir, Moscú. 1973.
43. RAO, C.R., "Linear statistical inference and its applications". John Wiley & Sons, New York. 1973.
44. SAGE, A.P., MELSA, J.L., "System identification". Academic Press. New York. 1971.
45. SODERSTROM, T., "On the convergence properties of the generalized least squares identification method". Reporte 7228. Division of Automatic Control, Lund Inst. of Tech., 1972.
46. SODERSTROM, T., "Convergence properties of the generalized least squares identification method". Proc. 3rd IFAC Symp., North Holland. 1973.
47. SODERSTROM, T., LJUNG, L., GUSTAVSSON, I., "A comparative study of recursive identification methods". Reporte 7427, Lund Inst. of Tech., 1974.
48. SODERSTROM, T., LJUNG, L., GUSTAVSSON, I., "A Theoretical analysis of recursive identification methods".

- Automática Vol. 14 pp 231-244, Gran Bretaña, 1978.
49. STOER, J., BULIRSH, R., "Introduction to numerical analysis". Springer-Verlag. New York, 1980.
50. STREJC, V., "Least squares in identification theory". Kybernetika Vol. 13, pp 83-105, Praga. 1977.
51. STREJC, V., "Least squares parameter estimation". Automatica Vol. 16, pp 535-558. 1980.
52. VAN DEN BOOM, A.J.W., VAN DEN ENDEN, A.W.M., "The determination of the order of process and noise dynamics". Proc. 3rd IFAC Symp. on Identification and System Parameter Estimation. The Hague/Delft. Países Bajos. 1973.
53. WELLSTEAD, P., "System identification, part C (C1): recursive estimation". UNIST, Manchester, 1975.
54. YOUNG, P., "Recursive approaches to time series analysis". Bulletin Institute of Maths. and Applications, 10 (1974).