

472 *ejemplar*



Universidad Nacional Autónoma de México

FACULTAD DE CIENCIAS

UNA VISION DEL ANALISIS SECUENCIAL

T E S I S

Que para obtener el título de:

A C T U A R I O

P r e s e n t a :

GLORIA MARIA BELEM TREJO VALDIVIA

México, D. F.

1981



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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 esta 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.

UNA VISION DEL ANALISIS SECUENCIAL

INDICE.

INTRODUCCION	1
CAPITULO I. ESTADISTICA CLASICA	3
<i>Estimación puntual</i>	4
<i>Estimación por intervalos</i>	10
<i>Pruebas de Hipótesis</i>	11
CAPITULO II. ESTIMACION SECUENCIAL	15
<i>Formulación del problema de estimación - secuencial</i>	16
<i>Cota Inferior de Crámer y Rao</i>	24
<i>Suficiencia y completéz en el caso se- cuencial</i>	29
<i>Consistencia de ciertos estimadores se- cuenciales</i>	39
<i>Estimación por doble muestreo</i>	41
<i>Estimación secuencial del tamaño de una población finita</i>	47
<i>Intervalos de confianza secuenciales</i>	59
<i>Introducción</i>	59
<i>Intervalos de confianza para la me- dia de una población normal</i>	59
<i>Intervalos de confianza para la di- ferencia de medias</i>	73
CAPITULO III. PRUEBAS DE HIPOTESIS SECUENCIALES	80
<i>Método General</i>	81
<i>Función Característica Operante</i>	81
<i>Tamaño Esperado de Muestra</i>	82
<i>Prueba de Wald para hipótesis simples</i>	83

<i>Pruebas secuenciales para hipótesis compuestas</i>	86
<i>Prueba de Wald</i>	87
<i>Prueba de Cox</i>	92
<i>Prueba de Bartlett</i>	95
<i>Prueba de Nanak Chand</i>	97
<i>Pruebas secuenciales entre tres hipótesis</i>	100
<i>Método de Sobel-Wald</i>	100
<i>Método de Armitage</i>	102
CAPITULO IV. OTRAS ALTERNATIVAS	103
<i>Región de confianza secuencial para la <u>me</u> <u>dia</u> de una población normal multivariada</i>	104
<i>Región de confianza para los parámetros - de una regresión lineal</i>	105
<i>Intervalo secuencial para la mediana</i>	107
<i>Prueba del rango ordenado</i>	108
<i>Planteamiento de un problema secuencial - que puede ser resuelto por métodos baye- sianos</i>	109
<i>Bibliografía de información general</i>	110
CONCLUSIONES	142
APENDICE A	143
BIBLIOGRAFIA	144

INTRODUCCION.

La palabra "secuencial" fue utilizada primero por Abraham -- Wald y sus colegas, aunque más o menos al mismo tiempo también la usó Barnard en Inglaterra, para describir procedimientos para probar hipótesis, en los cuales el número de observaciones no era fijo de antemano pero dependía, de acuerdo a alguna regla definida, de las observaciones mismas. El objeto de tomar observaciones recae en el problema de elegir entre un número pequeño (usualmente dos) de posibles hipótesis o decisiones. Algunas veces, al terminar dicho procedimiento de -- muestreo el observador deseaba estimar los valores de uno o más parámetros desconocidos, y surgió el problema de la "estimación secuencial". Los artículos acerca de este problema empezaron a aparecer publicados alrededor de 1946. Al principio, los autores tácitamente suponían que el procedimiento de muestreo secuencial había sido diseñado básicamente para probar hipótesis, y que la estimación era una idea tardía, pero Wald en su libro *Sequential Analysis* (1947) dedicó un pequeña sección para discutir procedimientos secuenciales cuyo objeto primario era proveer (con tan pocas observaciones como fuese posible) de un intervalo para un parámetro desconocido habiéndose prefijado el coeficiente de confianza y la longitud, o que satisficiera alguna otra condición similar. Ejemplificó en particular, el problema de estimar la media de una población normal, en la cual la varianza era desconocida, por un intervalo de confianza con coeficiente de confianza y amplitud dados. Sin embargo, no intentó obtener soluciones inmediatamente útiles a tales problemas.

Posteriormente, un gran número de autores publicaron artículos referentes a diferentes tópicos del Análisis Secuencial; sin embargo y debido a que el material de este tema es muy extenso, han sido pocas las publicaciones que presenten al Análisis Secuencial en conjunto, esto es, casi no existen textos que contengan, desde el punto de vista

secuencial, todo lo relacionado con Inferencia Estadística. Es por esto que se pensó en escribir un trabajo que comprendiera los temas de: Estimación puntual, estimación por intervalos y pruebas de hipótesis; enfatizando las analogías y diferencias que existen de estos tópicos con la Estadística Clásica.

La presentación del material en el presente trabajo es como sigue:

En el capítulo I se hace una exposición de los conceptos de Estadística Clásica que tienen relación con los conceptos de Estadística desde el punto de vista secuencial.

En el capítulo II se expone el material acerca del Análisis Secuencial exclusivamente el que se refiere a Estimación.

En el capítulo III se expone la teoría existente acerca de pruebas de hipótesis secuenciales, enfatizando en la parte de hipótesis compuestas contra compuestas.

Y en el capítulo IV se dan algunas ideas de las posibles extensiones del Análisis Secuencial a otros tópicos estadísticos, así como una bibliografía de los artículos que se han escrito acerca de todo el Análisis Secuencial.

CAPITULO I

ESTADISTICA CLASICA

En este capítulo, se dará un resumen de los resultados que son necesarios para los capítulos siguientes con relación a la parte de estimación tanto puntual como por intervalos así como lo referente a las pruebas de hipótesis.

Tanto el proceso de estimación como el de pruebas de hipótesis forman parte de lo que conoce como Inferencia Estadística, que es un conjunto de procedimientos por los que se establecen conclusiones respecto de una población, en base a los resultados que se obtienen a partir de una muestra de la misma.

Estimación puntual.

Supongamos que se tiene una muestra aleatoria de tamaño n fijo, X_1, X_2, \dots, X_n que se extrajo de una población X con función de densidad $f(x/\theta)$ que depende de un valor desconocido, pero fijo θ llamado parámetro. El proceso de estimación conlleva calcular una función de la muestra que no depende del parámetro y cuya imagen sea el espacio parametral Θ (esta función es llamada un estimador de θ), con la cual se obtenga una aproximación de la mejor manera posible del valor del parámetro θ .

Para asegurar hasta cierto punto que la aproximación es "buena", se le pedirá al estimador propuesto que tenga ciertas propiedades, las cuales se mencionarán a continuación:

Insesgamiento: Dado un estimador $\hat{\theta}$, se dirá que es insesgado si $E(\hat{\theta}) = \theta \quad \forall \theta \in \Theta$. Esta definición puede extenderse a cualquier $h(\theta)$ función del parámetro θ , es decir, dado $h(\hat{\theta})$ estimador de $h(\theta)$ tal que $E(h(\hat{\theta})) = h(\theta) \quad \forall \theta \in \Theta$, se dirá que dicho estimador es insesgado para $h(\theta)$.

Además, si $\hat{\theta}'$ es otro estimador de θ tal que $E(\hat{\theta}') = \theta + b(\theta)$ entonces $b(\theta)$ es llamado el sesgo de $\hat{\theta}'$.

Varianza mínima: Supongamos que $\hat{\theta}$ es un estimador con segundo momento central finito, es decir, $\text{Var}(\hat{\theta}) < \infty$. Sería deseable que dicho estimador no tuviera una varianza muy grande; una manera de analizar dicha varianza es a través de la Cota Inferior de Crámer y Rao (CICR), es decir:

Dado un estimador $\hat{\theta}$ de θ tal que $E(\hat{\theta}) = \theta + b(\theta)$, bajo ciertas condiciones de regularidad* y suponiendo que $b'(\theta)$, la derivada del sesgo de $\hat{\theta}$ existe se tiene que:

$$\text{Var}(\hat{\theta}) \geq \frac{(1 + b'(\theta))^2}{n E \left[\frac{\partial}{\partial \theta} \ln f(x|\theta) \right]^2}$$

Esta desigualdad puede generalizarse para cuando θ es un parámetro multidimensional.

Eficiencia: Una medida que considera la CICR y la varianza del estimador propuesto $\hat{\theta}$ es la eficiencia definida como:

$$\text{ef}(\hat{\theta}) = \frac{\text{CICR}}{\text{Var}(\hat{\theta})}$$

Dadas las características de CICR, se tiene que $0 \leq \text{ef}(\hat{\theta}) \leq 1$ para cualquier estimador de θ . Si $\hat{\theta}$ es un estimador cuya eficiencia es igual a 1 (es decir $\text{Var}(\hat{\theta}) = \text{CICR}$), se dice que $\hat{\theta}$ es un estimador eficiente de θ por lo cual podremos asegurar que tiene la varianza más pequeña entre todos los estimadores de la clase a que pertenezca.

*Ver Apéndice A.

Suficiencia: El obtener la muestra de la población siempre trae consigo un cierto costo (trabajo, dinero, tiempo, etc.) por lo cual no se desea desperdiciar nada de la información que da la muestra, es decir, otra de las propiedades que se le pedirá a los estimadores es que guarden toda la información de la muestra, dichos estimadores son los llamados suficientes. Esta suficiencia se da en el sentido de que basta conocer la información que proporciona para estimar el parámetro de interés. Una definición más formal de suficiencia es la siguiente:

Definición 1.1.

Sea X_1, X_2, \dots, X_n una muestra aleatoria (m.a.) de una población X con función de densidad $f(x/\theta)$. Una estadística $S=S(X_1, X_2, \dots, X_n)$ se dice que es una estadística suficiente si y solo si la distribución condicional de X_1, X_2, \dots, X_n dado $S=s$ no depende de θ para cualquier valor s de S .

El trabajar con esta definición para verificar la suficiencia de una estadística dada puede ser muy engorroso o complicarse demasiado. Una manera más fácil es utilizar el siguiente teorema debido a Fisher y Neyman.

Teorema 1.2.

Sea X_1, X_2, \dots, X_n una m.a. de una población X , con función de densidad $f(x/\theta)$, $\theta \in \Theta$ (espacio parametral). Sea $Y=V(X_1, \dots, X_n)$ una estadística con función de densidad $g(y/\theta)$. Entonces Y es una estadística suficiente para θ si y solo si

$$f(x_1, x_2, \dots, x_n/\theta) = g(y(x_1, \dots, x_n)/\theta) H(x_1, \dots, x_n)$$

donde H es una función exclusivamente de la muestra.

El teorema anterior como puede verse sirve para verificar la suficiencia pero si no se conoce una estadística que pudiera ser suficiente no es posible aplicar dicho teorema. Una manera de construir una estadística suficiente está dada a partir del llamado Teorema de Factorización que dice:

Teorema 1.3.-

Sea X_1, X_2, \dots, X_n una m.a. de una población con función de densidad $f(x/\theta)$. Sea $S=S(X_1, X_2, \dots, X_n)$ una estadística, entonces

S es suficiente para θ si y solo si

$$f(X_1, \dots, X_n / \theta) = G(S, \theta) H(X_1, \dots, X_n)$$

donde H es una función solamente de la muestra.

Como puede darse cuenta, dado un parámetro θ de una población no existe una única estadística suficiente para él, por lo que sería deseable tener una estadística que siendo suficiente sea lo más simple posible en cuanto a manejo, dicha estadística queda definida por:

Definición 1.4.-

Una estadística se dice que es suficiente minimal si es suficiente y si cualquier reducción de la partición del espacio muestral definida por ella, ya no es suficiente.

El siguiente teorema reúne las anteriores características, es decir, insesgamiento, varianza mínima y suficiencia:

Teorema 1.5 (Rao-Blackwell).-

Sea X_1, X_2, \dots, X_n una m.a. de una población con función de densidad $f(x/\Theta)$. Sea $Y_1 = Y_1(X_1, \dots, X_n)$ una estadística suficiente para Θ , y sea $Y_2 = Y_2(X_1, \dots, X_n)$ un estimador insesgado de Θ . Entonces,

$$E(Y_2 / y_1) = \varphi(y_1)$$

define un estimador insesgado para Θ cuya varianza es menor o igual que la de Y_2 .

Completez: Algunas veces los estimadores que se construyen en base al teorema anterior resultan ser los estimadores insesgados de mínima varianza uniformemente para el parámetro Θ , para identificar estos estimadores es necesario el concepto de completez de una familia de densidades, dicho concepto se establece como:

Definición 1.6.-

Sea X_1, \dots, X_n una m.a. de una población con función de densidad $f(x/\Theta)$. Sea $T = T(X_1, \dots, X_n)$ una estadística. La familia de densidades de T se dice que es completa si y solo si dada $z(T)$ función de la estadística tal que:

$$E(z(T)) = 0 \quad \text{implica} \quad z(T) = 0 \quad \text{casi dondequiera}$$

Además; una estadística es completa si y solo si la familia de densidades es completa.

Para identificar los estimadores insesgados de mínima varianza uniformemente, se utiliza el siguiente teorema:

Teorema 1.7.

Sea X_1, \dots, X_n una m.a. de una densidad $f(x/\theta)$. Si $S=S(X_1, \dots, X_n)$ una estadística suficiente y completa y si $T^*=T^*(S)$ es un estimador insesgado de $t(\theta)$ alguna función de θ , entonces,

T^* es un estimador insesgado de mínima varianza uniformemente.

Hasta aquí se ha hablado de las propiedades deseables de los estimadores, pero no hemos dicho la forma de como se como se construyen. Hay diferentes métodos de estimación, como son método de momentos, método de máxima verosimilitud, método de mínimos cuadrados, etc. Aquí solo se dará la forma del método de máxima verosimilitud pues es el que se mencionará más adelante en este trabajo.

Sea X una población con función de densidad $f(x/\theta)$ donde es desconocido, se desea obtener un estimador de θ a partir de una muestra X_1, \dots, X_n de dicha población.

$$\text{Sea } \underline{f}(\underline{x}/\theta) = \prod_{i=1}^n f(x_i/\theta)$$

la función de densidad conjunta de la muestra evaluada en los valores obtenidos. En este caso $\underline{f}(\underline{x}/\theta)$ numéricamente coincide con $L(\theta)$, la función de verosimilitud. El método consiste en encontrar $\hat{\theta}$ tal que:

$$L(\hat{\theta}) = \sup_{\theta} L(\theta)$$

Algunas veces la función $L(\theta)$ tiene una forma complicada la cual hace difícil encontrar el valor $\hat{\theta}$, para librarse de dicho problema y a partir de las propiedades de la función logaritmo (función monótona, estrictamente creciente), se va a encontrar aquel valor que maximice: $\ln L(\theta)$.

Usualmente se utiliza la derivada de $\ln L(\theta)$ para encontrar dicho máximo.

Estimación por intervalos.

Este tipo de estimación surge porque el estimador puntual solo da un valor numérico, pero no se cuenta con una medida de la aproximación de este punto al valor verdadero de θ , por lo cual ahora se dará un rango (intervalo) en el cual con cierta confianza se espera -- que se encuentre θ .

Esta forma de estimación se basa en la construcción de intervalos aleatorios (intervalos en donde por lo menos un extremo es una variable aleatoria). Hay casos donde se puede construir dicho intervalo por el método pivotal, pero también existe un método más general -- que puede ser aplicado en otros casos.

El método pivotal a grandes rasgos puede describirse como sigue:

Sea $Q = Q(X_1, \dots, X_n / \theta)$ una estadística cuya distribución no depende de θ , entonces para cualquier γ tal que $0 < \gamma < 1$, existen q_1 y q_2 tales que

$$P(q_1 < Q < q_2) = \gamma$$

si además, para cualquier muestra posible se tiene que

$$q_1 < Q(\underline{x} / \theta) < q_2 \quad \text{se cumple si y solo si}$$

$$T_1(\underline{x}) < \theta < T_2(\underline{x})$$

donde T_1 y T_2 son dos estadísticas que no dependen de θ , entonces,

$$(T_1, T_2) \text{ es un intervalo de confianza para } \theta \text{ al } 100 \cdot \gamma \%$$

Pruebas de hipótesis.

Una hipótesis estadística es una afirmación acerca de la población estadística en estudio. Las hipótesis estadísticas a menudo se expresan por medio de una afirmación de los valores del (o los) parámetro(s) del modelo probabilístico ajustado a dicha población.

Desde el punto de vista estadístico, existen 2 tipos de hipótesis:

- i) Hipótesis simple. Es aquella en donde se especifica por completo la distribución de la población de interés.
- ii) Hipótesis compuesta. Es aquella donde no queda determinada de manera única la distribución de la población.

Para verificar la validez de una hipótesis se obtendrá una muestra aleatoria de la población que mediante un cierto análisis proporcionará cierta información con la cual podremos rechazar la hipótesis (cuando los resultados no vayan de acuerdo a lo esperado) o no rechazarla (cuando dichos resultados se aproximen a los resultados esperados).

Al conjunto de valores con los cuales se rechaza la hipótesis de interés (que por lo general se conoce como hipótesis nula) se le conoce como región crítica o zona de rechazo.

Al tomar una decisión acerca de la hipótesis, se tiene la posibilidad de cometer 2 tipos de error:

- Rechazar la hipótesis cuando esta es cierta. A este error se le conoce como error tipo 1.

- No rechazar la hipótesis cuando esta es falsa. Esto constituye el error tipo II.

Cada uno de estos errores tiene una probabilidad de ocurrir (desde que están en función de la decisión la cual a su vez depende de la muestra), dichas probabilidades se conocen como α y β .

Desgraciadamente no es posible controlar las dos probabilidades de error, ya que cuando una decrece la otra aumenta y viceversa, -- por lo que dado un criterio (para rechazar o no rechazar la hipótesis) de decisión se podrá solamente dar el valor que se desea para una de estas probabilidades (por lo general el valor α es el que se da).

Para cada criterio de decisión es posible construir una región de rechazo, sin embargo es posible construir una región tal que dado α minimice el valor de β , a esta región se le conoce como una mejor región crítica de tamaño α , su obtención está dada a partir -- del siguiente teorema debido a Neyman y Pearson:

Teorema 1.8.-

Sean X_1, \dots, X_n una m.a. de una población con función de densidad $f(x/\theta)$ y

$$L(x/\theta) = \prod_{i=1}^n f(x_i/\theta)$$

Si θ_0 y θ_1 son dos valores conocidos, considere la prueba:

$$H_0: \theta = \theta_0 \quad \text{vs.} \quad H_1: \theta = \theta_1 \quad \dots \quad (1.9)$$

y $k > 0$. Sea $B \subset \Omega$ tal que:

$$i) \frac{L(x/\theta_0)}{L(x/\theta_1)} \leq k \quad \forall x \in \mathcal{P}$$

$$ii) \frac{L(x/\theta_0)}{L(x/\theta_1)} \geq k \quad \forall x \in \mathcal{P}^c$$

$$iii) P\{(X_1, \dots, X_n) \in \mathcal{P} / H_0\} = \alpha$$

entonces \mathcal{P} es una mejor región de tamaño α para (1.9)

Una caso más general que el anterior (en el cual las 2 hipótesis son simples), es considerar H_0 una hipótesis simple y H_1 una hipótesis compuesta. Cuando la hipótesis compuesta tiene cierta forma tal que existe un conjunto \mathcal{P} que es una mejor región crítica para la prueba H_0 contra cualquier hipótesis simple derivada de H_1 , entonces \mathcal{P} es llamada una región crítica uniformemente más potente.

Cuando la hipótesis compuesta no cumple con lo anterior y/o cuando la hipótesis nula es compuesta se utiliza un método más general de prueba llamado Principio de Razón de Verosimilitud que puede expresarse como sigue:

Considere la prueba

$$H_0: \theta \in \omega_0^* \quad \text{vs.} \quad H_1: \text{todas las alternativas}$$

Sea X_1, \dots, X_n una m.a. de la población con función de densidad $f(x/\theta)$ y

$$L(x/\theta) = \prod_{i=1}^n f(x_i/\theta)$$

La razón de verosimilitud se define como:

* ω_0^* es un subconjunto del espacio parametral (2)

$$\lambda = \frac{L(\underline{x} / \hat{\omega})}{L(\underline{x} / \hat{\theta})}$$

donde $L(\underline{x} / \hat{\omega}) = \sup_{\omega \in \omega_0} L(\underline{x} / \omega)$

$L(\underline{x} / \hat{\theta}) = \sup_{\theta \in \Theta} L(\underline{x} / \theta)$

Sea $\lambda_0 > 0$. El principio de razón de verosimilitud establece que H_0 se rechaza si y solo si

$$\lambda \leq \lambda_0$$

donde el nivel de significancia α está dado por:

$$\alpha = P(\lambda \leq \lambda_0 / H_0)$$

CAPITULO II

ESTIMACION SECUENCIAL

Los procedimientos secuenciales difieren de los procedimientos estadísticos usuales en el hecho de que el tamaño de muestra no es fijo desde un principio. El experimentador tiene la libertad de hacer observaciones una por una (o un número fijo) al mismo tiempo que decide si termina de muestrear y toma una decisión o continúa el muestreo y toma una decisión más tarde.

La característica principal de un procedimiento secuencial es que el número de observaciones que se requiere para terminar el experimento es una variable aleatoria pues depende de los resultados de las observaciones. A menudo uno puede estar interesado en procedimientos secuenciales porque suelen ser económicos en el sentido de que se puede tomar una decisión antes que si se usara un procedimiento de tamaño de muestra fijo. En la experimentación secuencial se necesitan:

- i) El tamaño de muestra inicial (por lo general se toma un tamaño de 1).
- ii) Una regla de terminación del experimento.
- iii) El número de observaciones que se tomarán en el caso de continuar con el experimento (aquí también por lo general este número es 1).
- iv) Una regla de decisión terminal.

Formulación del problema de estimación secuencial.

El procedimiento de estimación secuencial por conjuntos puede ser descrito como sigue:

Sea Ω el espacio muestral, se obtendrán secuencialmente -- conjuntos E_1, E_2, \dots con las siguientes características:

E_1 un conjunto de muestras de tamaño 1 e. d. $E_1 \subset \Omega$

$E_2 \subset E_1^c \times \Omega$ un conjunto de muestras de tamaño 2.

$E_3 \subset (E_1^c \times \Omega \times \Omega) \cap (E_2^c \times \Omega)$ un conjunto de muestras de tamaño 3.

...

y así sucesivamente.

Con cada muestra \underline{X}_m de tamaño m de E_m está asociado un subconjunto $W(\underline{X}_m)$ del espacio paramétrico. Las observaciones se siguen obteniendo hasta alcanzar un valor n tal que \underline{X}_n pertenezca a E_n . En esta etapa suspendemos el proceso y confiamos que $W(\underline{X}_n)$ contiene al verdadero valor del parámetro, es decir, $W(\underline{X}_n)$ es el conjunto de confianza resultante del procedimiento de estimación secuencial.

Por lo que, el procedimiento de estimación secuencial queda determinado por los conjuntos E_1, E_2, \dots y la función $W(\underline{X})$ definida para toda \underline{X} en E_1, E_2, \dots . Y el problema principal es, entonces, elegir apropiadamente los conjuntos E_1, E_2, \dots , y $W(\underline{X})$ para lo cual se deben de cumplir las dos siguientes condiciones:

Condición 1.- El conjunto de confianza $W(\underline{X}_m)$ que resulta del procedimiento de estimación secuencial, debe satisfacer ciertos requerimientos de tipo geométrico.

Condición 2.- El conjunto de confianza $W(\underline{X}_m)$ debe satisfacer que:

$$P(\theta \in W(\underline{X}_m) / \theta) > 1 - \alpha$$

La condición 1 no es un problema estadístico pero depende directamente de las consideraciones prácticas del problema bajo estudio, por ejemplo, si el espacio parametral es 1-dimensional, se puede pensar en $W(\underline{X}_m)$ como un intervalo de longitud fija, o que su longitud sea una función del punto medio del intervalo, como se hace para el caso de la estimación de la media de una distribución binomial. Esta idea se generaliza en términos de volumen o de diámetro del conjunto (definido como la máxima distancia posible entre dos puntos del conjunto).

Dada la definición de los conjuntos E_1, E_2, \dots el conjunto de todas las muestras posibles está dada por $E_1 \cup E_2 \cup E_3 \cup \dots$. Sea N la variable aleatoria que denota el número de observaciones que son tomadas para finalizar un procedimiento secuencial.

Se define

$$g(X_1, X_2, \dots, X_N) = g_i(x_1, x_2, \dots, x_i) \quad \text{si } (x_1, \dots, x_i) \in E_i \quad i=1, 2, \dots$$

es decir

$$g(X_1, X_2, \dots, X_N) = \begin{cases} g_1(x_1) & \text{si } x_1 \in E_1 \\ g_2(x_1, x_2) & \text{si } (x_1, x_2) \in E_2 \\ \vdots & \\ g_n(x_1, \dots, x_n) & \text{si } (x_1, x_2, \dots, x_n) \in E_n \\ \vdots & \\ \vdots & \end{cases}$$

Se supondrá de aquí en adelante que las variables X 's tienen función de densidad $f(x/\theta)$.

Por lo tanto $g(X_1, X_2, \dots, X_N)$ es una variable aleatoria discreta con --
función de probabilidad dada por:

$$P(g(X_1, \dots, X_N) = g_i(x_1, \dots, x_i)) = P((x_1, x_2, \dots, x_i) \in E_i)$$

$$= \int_{E_i} \prod_{j=1}^i f(x_j | \theta) dx_j$$

y cuya esperanza está dada por

$$E(g(X_1, \dots, X_N)) = \sum_{i=1}^{\infty} \int_{E_i} g_i(x_1, \dots, x_i) \prod_{j=1}^i f(x_j | \theta) dx_j$$

Definición 2.1.-

Sea $T_N = g(X_1, \dots, X_N)$, T_N es un estimador insesgado para θ
si

$$E(T_N) = \theta$$

Definición 2.2.-

Si $T_N = g(X_1, \dots, X_N)$ es tal que $E(T_N) = \theta + b(\theta)$ entonces
 $b(\theta)$ es llamado el sesgo del estimador.

La idea que se tiene en el caso tradicional acerca de un es-
timador con mínima varianza dentro de cierta clase de estimadores, es
utilizada para hacer una extensión al caso secuencial. Para poder ha-
cer dicha extensión probaremos algunos lemas que necesitaremos.

Lema 2.3.-

Sea $S_N = X_1 + X_2 + \dots + X_N$, donde X_i son variables aleatorias independientes idénticamente distribuidas (v.a.i.i.d.) entonces:

$$i) \text{ Si } E(N) < \infty \text{ y } E(|X|) < \infty \quad \Rightarrow$$

$$E(S_N) = E(N)E(X)$$

$$ii) \text{ Si } E(X) = 0, E(X^2) < \infty \text{ y } E(N) < \infty \quad \Rightarrow$$

$$E(S_N^2) = E(N)E(X^2)$$

$$iii) \text{ Si } E(g(X)) = E(h(X)) = 0, E(|g(X)h(X)|) < \infty \text{ y } E(N^2) < \infty \quad \Rightarrow$$

$$E\left(\sum_{i=1}^N g(X_i)\right)\left(\sum_{i=1}^N h(X_i)\right) = E(N)E(g(X)h(X))$$

Demostración:

$$i) \text{ Sea } Y_i = \begin{cases} 1 & \text{si } N \geq i \\ 0 & \text{si } N < i \end{cases} \quad i = 1, 2, \dots$$

que no depende de X_i

$$\begin{aligned} E(S_N) &= E(X_1 \cdot 1 + X_2 \cdot 1 + \dots + X_N \cdot 1 + X_{N+1} \cdot 0 + \dots) \\ &= E\left(\sum_{i=1}^{\infty} Y_i X_i\right) = \sum_{i=1}^{\infty} E(Y_i X_i) = \sum_{i=1}^{\infty} E(Y_i)E(X_i) \end{aligned}$$

pues Y_i y X_i son independientes para toda i

$$= \sum_{i=1}^{\infty} E(Y_i)E(X) = E(X) \sum_{i=1}^{\infty} E(Y_i)$$

$$E(Y_{\lambda}) = \sum_{y=0}^{\infty} y_j P(Y_{\lambda} = y_j) = P(Y_{\lambda} = 1) = P(N \geq \lambda)$$

$$E(S_N) = E(X) \sum_{\lambda=1}^{\infty} P(N \geq \lambda) = E(X) \sum_{\lambda=1}^{\infty} \sum_{j=\lambda}^{\infty} P(N=j)$$

$$= E(X) (P(N=1) + P(N=2) + P(N=3) + \dots)$$

$$P(N=2) + P(N=3) + \dots)$$

$$+ P(N=3) + \dots)$$

$$= E(X) (P(N=1) + 2P(N=2) + 3P(N=3) + \dots)$$

$$= E(X) \sum_{j=1}^{\infty} j P(N=j) = E(X) E(N)$$

$$\text{ii) } E(S_N^2) = E\left(\sum_{i=1}^N X_i\right)^2 = E\left(\sum_{i=1}^N Y_i X_i\right)^2 = E\left(\sum_{i=1}^N Y_i^2 X_i^2 + \right.$$

$$\left. + 2 \sum_{i < j} Y_i Y_j X_i X_j\right)$$

$$= E\left(\sum_{i=1}^N Y_i^2 X_i^2\right) + 2E\left(\sum_{i < j} Y_i Y_j X_i X_j\right)$$

$$= \sum_{i=1}^{\infty} E(Y_i^2 X_i^2) + 2 \sum_{i < j} E(Y_i Y_j X_i X_j)$$

$$= \sum_{i=1}^{\infty} E(X_i^2) E(Y_i^2) + 2 \sum_{i < j} E(Y_i Y_j X_i X_j)$$

$$= \sum_{i=1}^{\infty} E(X_i^2) E(Y_i^2) + 2 \sum_{i < j} E(Y_i Y_j X_i X_j)$$

$$\sum_{i < j} E(Y_i Y_j X_i X_j) = \sum_{i < j} E(Y_i Y_j) E(X_i X_j)$$

$$= \sum_{i < j} E(Y_i Y_j) E(X_i) E(X_j)$$

$$= E(X) E(X) \sum_{i < j} E(Y_i Y_j) = 0$$

pues $E(X) = 0$

por lo tanto

$$\begin{aligned} E(S_N^2) &= \sum_{i=1}^{\infty} E(X^2) E(Y_i^2) = E(X^2) \sum_{i=1}^{\infty} P(Y_i = 1) \\ &= E(X^2) \sum_{i=1}^{\infty} P(N \geq i) = E(X^2) E(N) \end{aligned}$$

$$\begin{aligned} \text{iii) } E\left(\sum_{i=1}^N g(X_i)\right) \left(\sum_{i=1}^N h(X_i)\right) &= \\ E\left(\sum_{i=1}^{\infty} Y_i g(X_i)\right) \left(\sum_{j=1}^{\infty} Y_j h(X_j)\right) &= \\ = E\left(\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} Y_i Y_j g(X_i) h(X_j)\right) &= \\ = E\left(\sum_{i=1}^{\infty} Y_i^2 g(X_i) h(X_i) + \sum_{i \neq j} Y_i Y_j g(X_i) h(X_j)\right) &= \\ = \sum_{i=1}^{\infty} E(Y_i^2 g(X_i) h(X_i)) + \sum_{i \neq j} E(Y_i Y_j g(X_i) h(X_j)) &= \\ = \sum_{i=1}^{\infty} E(Y_i^2) E(g(X_i) h(X_i)) + \sum_{i \neq j} E(Y_i Y_j) E(g(X_i) h(X_j)) \end{aligned}$$

pues Y_i es independiente de X_j para toda i, j

$$= \sum_{i=1}^{\infty} E(Y_i^2) E(g(X) h(X)) + \sum_{i \neq j} E(Y_i Y_j) E(g(X_i)) E(h(X_j))$$

pues X_i, X_j son independientes lo que implica que $g(X_i), h(X_j)$ también lo son.

$$\begin{aligned} &= E(g(X) h(X)) \sum_{i=1}^{\infty} E(Y_i^2) + \sum_{i \neq j} E(Y_i Y_j) E(g(X)) E(h(X)) \\ &= E(g(X) h(X)) \sum_{i=1}^{\infty} E(Y_i^2) = E(N) E(g(X) h(X)) \end{aligned}$$

el cambio de la suma infinita y la esperanza se justifica pues

$$E\left(\left|g(X) h(X)\right|\right) < \infty$$

$$\begin{aligned}
 \text{además } \sum_{i \neq j} E(Y_i Y_j) &= \sum_{i \neq j} P(Y_i = 1, Y_j = 1) \\
 &= 2 \sum_{i < j} P(Y_i = 1, Y_j = 1) \\
 &= 2 \sum_{i < j} P(Y_i = 1 / Y_j = 1) P(Y_j = 1)
 \end{aligned}$$

Si $Y_j = 1$ entonces $N \geq j > i$ por lo tanto $Y_i = 1$, es decir

$$P(Y_i = 1 / Y_j = 1) = 1$$

por lo tanto

$$\begin{aligned}
 \sum_{i \neq j} E(Y_i Y_j) &= 2 \sum_{i < j} P(Y_j = 1) \\
 &= 2 \sum_{i < j} P(N \geq j) = 2 \sum_{j=2}^{\infty} \sum_{i=1}^{j-1} P(N \geq j) \\
 &= 2 \sum_{j=2}^{\infty} (j-1) P(N \geq j) = 2 \sum_{j=1}^{\infty} (j-1) P(N \geq j)
 \end{aligned}$$

$$\forall j \geq 1 \quad j-1 < 2(j-1) < 2(j-1) + 1 = 2j-1$$

por lo tanto

$$\begin{aligned}
 \sum_{i \neq j} E(Y_i Y_j) &< 2 \sum_{j=1}^{\infty} (2j-1) P(N \geq j) \\
 &= 2(P(N \geq 1) + 3P(N \geq 2) + 5P(N \geq 3) + \dots) \\
 &= 2(P(N = 1) + P(N = 2) + P(N = 3) + \dots \\
 &\quad P(N = 2) + P(N = 3) + \dots \\
 &\quad P(N = 2) + P(N = 3) + \dots \\
 &\quad P(N = 2) + P(N = 3) + \dots \\
 &\quad P(N = 3) + \dots)
 \end{aligned}$$

$$\begin{aligned}
 &= 2(P(N = 1) + 4P(N = 2) + 9P(N = 3) + \dots) \\
 &= 2 \sum_{j=1}^{\infty} j^2 P(N = j) = 2E(N^2) < \infty
 \end{aligned}$$

El lema anterior puede ser generalizado para el caso en que las variables X_i 's sean dependientes de la siguiente forma:

Lema 2.4.-

Sea $v_i = E(X_i / N \geq i)$ el cual existe si $P(N \geq i) \neq 0$
 y $v_i' = E(|X_i - v_i| / N \geq i)$ para $i = 1, 2, 3, \dots$
 supongamos que la serie

$$\sum_{i=1}^{\infty} \sum_{j=1}^i v_j' P(N=i) \quad \text{converge, entonces}$$

$$E(S_N - \sum_{i=1}^N v_i) = 0$$

La demostración es similar a la hecha en el lema anterior por lo cual se omitirá.

La extensión mencionada anteriormente queda expresada en el siguiente teorema:

Teorema 2.5 (Cota Inferior de Crámer y Rao).-

Sea $T_N = g(X_1, X_2, \dots, X_N)$ un estimador de θ , tal que $E(T_N) = \theta + b(\theta)$. Suponiendo que la diferenciación puede intercambiarse con los signos de suma e integración (esto constituye lo que se conoce como condiciones de regularidad) y que $b'(\theta)$ existe, entonces:

$$\sigma_{TN}^2 > \frac{(1 + b'(\theta))^2}{E(N) E\left(\frac{\partial}{\partial \theta} \ln f(x|\theta)\right)^2}$$

Demostración:

Dado que el conjunto de todas las muestras posibles es -----
 $E_1 \cup E_2 \cup E_3 \cup \dots$, entonces:

$$\sum_{i=1}^{\infty} \int_{E_i} \prod_{j=1}^i f(x_j|\theta) d\mathbf{x}_j = \int_{\cup E_i} \prod_{j=1}^i f(x_j|\theta) d\mathbf{x}_j = 1$$

derivando ambos miembros de la última igualdad se tiene

$$\frac{\partial}{\partial \theta} (1) = \frac{\partial}{\partial \theta} \sum_{i=1}^{\infty} \int_{E_i} \prod_{j=1}^i f(x_j|\theta) d\mathbf{x}_j$$

$$0 = \sum_{i=1}^{\infty} \int_{E_i} \frac{\partial}{\partial \theta} \prod_{j=1}^i f(x_j|\theta) d\mathbf{x}_j$$

$$= \sum_{i=1}^{\infty} \int_{E_i} \frac{\prod_{j=1}^i f(x_j|\theta)}{\prod_{j=1}^i f(x_j|\theta)} \frac{\partial}{\partial \theta} \prod_{j=1}^i f(x_j|\theta) d\mathbf{x}_j$$

$$= \sum_{i=1}^{\infty} \int_{E_i} \prod_{j=1}^i f(x_j|\theta) \frac{\partial}{\partial \theta} \ln \prod_{j=1}^i f(x_j|\theta) d\mathbf{x}_j$$

$$= \sum_{i=1}^{\infty} \int_{E_i} \prod_{j=1}^i f(x_j|\theta) \frac{\partial}{\partial \theta} \left(\sum_{j=1}^i \ln f(x_j|\theta) \right) d\mathbf{x}_j$$

$$= \sum_{i=1}^{\infty} \int_{E_i} \prod_{j=1}^i f(x_j|\theta) \left[\sum_{j=1}^i \frac{\partial}{\partial \theta} \ln f(x_j|\theta) \right] d\mathbf{x}_j$$

$$\text{si } g_i(x_1, \dots, x_i) = \sum_{j=1}^i \frac{\partial}{\partial \theta} \ln f(x_j | \theta) \quad \Rightarrow$$

$$E\left(\sum_{j=1}^N \frac{\partial}{\partial \theta} \ln f(x_j | \theta)\right) = 0$$

$$\text{Sean } Y_j = \frac{\partial}{\partial \theta} \ln f(x_j | \theta) \quad j=1, 2, \dots$$

$$\text{y } S_N = \sum_{j=1}^N Y_j = \sum_{j=1}^N \frac{\partial}{\partial \theta} \ln f(x_j | \theta)$$

$$\therefore E(S_N) = 0$$

$$\text{si } E(|Y_j|) = E\left(\left|\frac{\partial}{\partial \theta} \ln f(x_j | \theta)\right|\right) < \infty$$

$$E(S_N) = E(N)E(Y) = 0$$

utilizando el inciso i) del Lema 2.3

$$\therefore E(Y) = 0 \quad \text{es decir} \quad E\left(\frac{\partial}{\partial \theta} \ln f(x | \theta)\right) = 0$$

$$\theta + b'(\theta) = E(T_N) = \sum_{i=1}^{\infty} \int_{E_i} T_i \prod_{j=1}^i f(x_j | \theta) dx_j$$

derivando tenemos que:

$$1 + b'(\theta) = \sum_{i=1}^{\infty} \int_{E_i} T_i \left(\frac{\partial}{\partial \theta} \prod_{j=1}^i f(x_j | \theta)\right) dx_j$$

$$= \sum_{i=1}^{\infty} \int_{E_i} T_i \left(\prod_{j=1}^i f(x_j | \theta)\right) \left(\frac{\partial}{\partial \theta} \ln \prod_{j=1}^i f(x_j | \theta)\right) dx_j$$

$$\begin{aligned}
&= \sum_{i=1}^{\infty} \int_{E_i} \left(T_i \frac{\partial}{\partial \theta} \sum_{j=1}^i \ln f(x_j | \theta) \right) \prod_{j=1}^i f(x_j | \theta) d\mathbf{x}_j \\
&= \sum_{i=1}^{\infty} \int_{E_i} \left(T_i \sum_{j=1}^i \frac{\partial}{\partial \theta} \ln f(x_j | \theta) \right) \prod_{j=1}^i f(x_j | \theta) d\mathbf{x}_j \\
&= \sum_{i=1}^{\infty} \int_{E_i} \left(T_i \sum_{j=1}^i Y_j \right) \prod_{j=1}^i f(x_j | \theta) d\mathbf{x}_j
\end{aligned}$$

$$\therefore 1 + b'(\theta) = E \left(T_N \sum_{j=1}^N Y_j \right) = E(T_N S_N)$$

Además se tiene que para cualesquiera 2 variables aleatorias

W_1, W_2 :

$$\text{Cov}(W_1, W_2) \leq \sqrt{\text{Var}(W_1)\text{Var}(W_2)}$$

en particular si $W_1 = T_N$ y $W_2 = S_N$ entonces

$$\text{Cov}(T_N, S_N) \leq \sqrt{\text{Var}(T_N)\text{Var}(S_N)}$$

pero $\text{Cov}(T_N, S_N) = E(T_N S_N) - E(T_N)E(S_N)$

$$= E(T_N S_N)$$

por lo tanto

$$E(T_N S_N) \leq \sqrt{\text{Var}(T_N)\text{Var}(S_N)}$$

$$1 + b'(\theta) \leq \sqrt{\text{Var}(T_N)\text{Var}(S_N)}$$

$$(1 + b'(\theta))^2 \leq \text{Var}(T_N)\text{Var}(S_N)$$

ahora

$$\begin{aligned}\text{Var}(S_N) &= E(S_N^2) - E^2(S_N) \\ &= E(S_N^2)\end{aligned}$$

si $E(Y^2) < \infty$ se puede utilizar el inciso ii) del Lema 2.3, por lo que:

$$\text{Var}(S_N) = E(N)E(Y^2)$$

por lo tanto

$$(1 + b'(\theta))^2 \leq \text{Var}(T_N)E(N)E(Y^2)$$

$$(1 + b'(\theta))^2 \leq \text{Var}(T_N)E(N)E\left(\frac{\partial}{\partial \theta} \ln f(x|\theta)\right)^2$$

\therefore

$$\text{Var}(T_N) \geq \frac{(1 + b'(\theta))^2}{E(N)E\left(\frac{\partial}{\partial \theta} \ln f(x|\theta)\right)^2}$$

Corolario 2.6.-

Si T_N es un estimador insesgado de θ entonces

$$\text{Var}(T_N) \geq \frac{1}{E(N)E\left(\frac{\partial}{\partial \theta} \ln f(x|\theta)\right)^2}$$

A continuación se darán dos resultados importantes pero no se demostrarán debido a que las demostraciones son similares a la del teorema 2.5.

Teorema 2.7.-

Si T_N es un estimador insesgado de $h(\theta)$ se tiene que:

$$\text{Var}(T_N) \geq \frac{(h'(\theta))^2}{E(N) E\left(\frac{\partial}{\partial \theta} \ln f(x|\theta)\right)^2}$$

Teorema 2.8.-

Si $\theta = (\theta_1, \dots, \theta_k)'$ y además es de interés la varian-
za de un estimador insesgado de una componente de θ digamos θ_i , en
tonces:

$$\text{Var}(T_{N,i}) \geq \frac{1}{E(N) E\left(\frac{\partial}{\partial \theta} \ln f(x|\theta)\right)^2}$$

donde $T_{N,i}$ denota un estimador insesgado de θ_i .

Suficiencia y completez en el caso secuencial.

La elección de una buena regla de decisión (en el sentido --
que minimice el tamaño de muestra esperado) nos podría ayudar a obte-
ner un buen estimador del parámetro θ . Los métodos que a continua-
ción serán descritos a pesar de que no puedan ser usados para encontrar
una regla de decisión óptima, la desigualdad de Crámer y Rao puede ser
usada.

En esta sección se supondrá que las probabilidades condicionales usadas existen. Supongamos que para cada n fija, la función de densidad conjunta de X_1, X_2, \dots, X_n admite una estadística suficiente. Sea $T_n = T(X_1, X_2, \dots, X_n)$ una estadística suficiente minimal para la densidad conjunta de X_1, X_2, \dots, X_n .

Definición 2.9.-

La sucesión $\{T_1, T_2, \dots\}$ es llamada una sucesión suficiente para el modelo secuencial.

Teorema 2.10.-

Si para cada n , $T_n = T(X_1, X_2, \dots, X_n)$ es una estadística suficiente para Θ en la muestra fija X_1, X_2, \dots, X_n entonces

(T_N, N) es una estadística suficiente para Θ en el caso secuencial.

Demostración:

(Este teorema fue probado por Blackwell bajo la suposición de que la regla de decisión depende solo de las T 's).

Sea $\bar{E} \subset \bigcup_{i=1}^{\infty} E_i$ un conjunto medible

P.D. $P(E / N = n, T_n = t)$ no depende de Θ

$$P(E/N=n, T_n=t) = \frac{P(E, N=n, T_n=t)}{P(N=n, T_n=t)}$$

$$\begin{aligned}
 &= \frac{P(E, N=n, T_n=t)}{P(T_n=t)} \cdot \frac{P(T_n=t)}{P(N=n, T_n=t)} \\
 &= P(E, N=n/T_n=t) \cdot \frac{1}{P(N=n/T_n=t)}
 \end{aligned}$$

Si $N = n$ entonces $X_1, X_2, \dots, X_n \in E_n$ y viceversa, por lo que

$$\begin{aligned}
 P(E/N=n, T_n=t) &= P(E, E_n/T_n=t) \cdot \frac{1}{P(E_n/T_n=t)} \\
 &= \frac{P(E \cap E_n/T_n=t)}{P(E_n/T_n=t)}
 \end{aligned}$$

tanto el numerador como el denominador no dependen de θ , pues T_n es suficiente para θ para cada n , por lo tanto

$P(E/N=n, T_n=t)$ no depende de θ

es decir (T_N, N) es suficiente para θ en el caso secuencial.

Sea $\{T_N\}$ una sucesión suficiente en un modelo secuencial, entonces dado que T_n es suficiente para toda n y por lo tanto encierra toda la información de la muestra, es razonable pensar en que la decisión que se tome debe depender de T_n si el experimento se termina en el n -ésimo paso. Del hecho anterior se desprende la siguiente definición.

Definición 2.11.

La sucesión de estadísticas suficientes $\{T_n\}$ se dice que

es una sucesión transitiva si para toda $n \geq 1$ y para $\theta \in \Theta$, la distribución condicional de T_{n+1} dado que $(X_1, X_2, \dots, X_n) = (x_1, x_2, \dots, x_n)$ es idéntica a la distribución condicional de T_{n+1} dado que -----
 $T_n(X_1, X_2, \dots, X_n) = t_n(x_1, x_2, \dots, x_n)$.

Es decir que toda la información acerca de T_{n+1} que está contenida en $X_n = (X_1, \dots, X_n)$ está concentrada en la función T_n . Se demostró (ver Bahadur) que al tener una sucesión suficiente y transitiva -- $\{T_n\}$, cualquier procedimiento secuencial basado en $\{X_n\}$ es equivalente a cualquier procedimiento que en el n -ésimo paso esté basado solamente en T_n . Cuando se tiene una familia de v.a.i.i.d., $\{T_n\}$ es -- una sucesión transitiva si:

$$T_{n+1}(X_{n+1}) = \Psi_n(T_n(X_n), X_{n+1}) \quad \text{para toda } n \geq 1$$

De la misma manera como se consideró la suficiencia en un -- caso secuencial, la completéz de una familia (T_N, N) estará basada en -- la completéz de T_m para m fija como sigue:

Definición 2.12.-

La familia de distribuciones de (T_N, N) se dice que es completa si

$$E(g(T_N, N)) = 0 \text{ implica que } g(T_N, N) = 0 \text{ casi dondequiera.}$$

Definición 2.13.-

La familia de distribuciones de (T_N, N) se dice que es acotadamente completa si para toda n , $g(T_n, n)$ es acotada y (T_N, N) es completa.

Lehman y Stein (1950) obtuvieron que en el caso normal, con media θ y varianza 1, (T_N, N) es completa si $N=m$ y $T_m = \sum_1^m X_i$; en el caso de una distribución Poisson también se tiene que si $N=m$ y $T_m = \sum_1^m X_i$ (T_N, N) será completa.

Sea S_m el conjunto de valores de T_m con los cuales el procedimiento se termina en el m -ésimo paso, una condición necesaria y suficiente para que (T_N, N) sea completa es que las S_m 's sean intervalos ajenos tales que $\forall x \in S_m$ y $\forall y \in S_{m+1}$ se tiene que $x < y$ (en el caso en que T_m sea una estadística k -dimensional, la condición anterior debe cumplirse para cada coordenada). Por ejemplo si la regla de término es que continúe hasta que $T_m > c$ (un valor dado) entonces (T_N, N) no es completa.

Ejemplo (Caso binomial)

$$f(x/\theta) = \begin{cases} \theta^x (1-\theta)^{1-x} & x=0,1 \\ 0 & \text{e.o.l.} \end{cases}$$

Sea n un entero fijo y X_1, X_2, \dots, X_n la muestra correspondiente, entonces,

$$\begin{aligned} L(\underline{x}/\theta) &= \prod_{i=1}^n f(x_i/\theta) \\ &= \theta^{\sum_1^n x_i} (1-\theta)^{n - \sum_1^n x_i} \\ &= (\theta(1-\theta)^{-1})^{\sum_1^n x_i} (1-\theta)^n \end{aligned}$$

entonces $T_n = \sum_1^n X_i$ es suficiente para θ .

Si tenemos una regla de decisión la cual solo depende de T_n . Cada punto (T_m, m) con $m=1, 2, \dots$, puede ser de 3 tipos:

- i) Punto de decisión.
- ii) Punto de continuación.
- iii) Punto imposible.

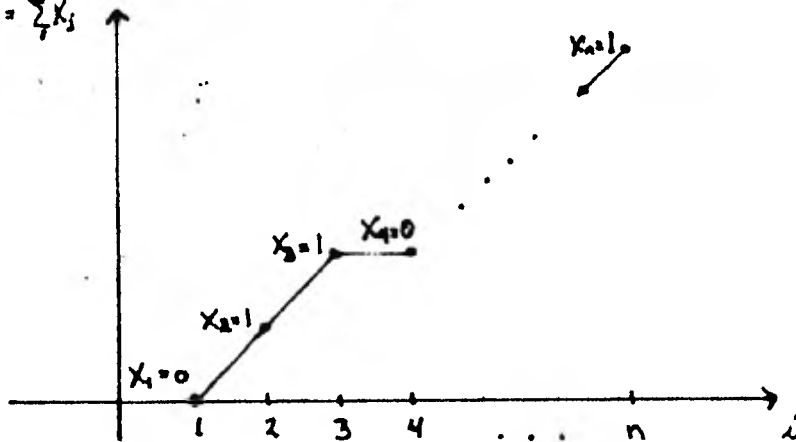
Por lo que la estadística (T_N, N) consta de todos los puntos del tipo i), pues los puntos de continuación no son valores posibles de (T_N, N) por la naturaleza de la variable N (regla cerrada de decisión). Se desea un estimador insesgado de θ , la primera observación X_1 es insesgada pues $E(X_1) = \theta$, además se tiene que (T_N, N) es suficiente entonces al aplicar el teorema de Rao-Blackwell

$$Y = E(X_1 / N, T_N) = P(X_1 = 1 / N, T_N)$$

es tal que $E(Y) = \theta$ y $\text{Var}(Y) \leq \text{Var}(X_1)$.

Si para cada (X_1, \dots, X_n) analizamos su trayectoria desde el origen hasta el punto (n, T_n) , se tendrá que en el i -ésimo paso el curso será a la derecha si $X_i = 0$ o hacia arriba si $X_i = 1$, y obviamente esta trayectoria no pasará por ningún punto de decisión (N, T_N) antes de llegar a (n, T_n) .

$$T_n = \sum_{j=1}^n X_j$$



La probabilidad de que se pase del origen al punto (m, t) es $\theta^t (1 - \theta)^{m-t}$. Sean $\#(m, t)$ el número de trayectorias o caminos -- que van del origen al punto (m, t) y $\#^*(m, t)$ el número de caminos que van del punto $(1, 1)$ a (m, t) , entonces

$$\begin{aligned} y(m, t) &= P(X_1 = 1 / N=m, T_N=t) \\ &= \frac{P(X_1 = 1, (N, T_N) = (m, t))}{P((N, T_N) = (m, t))} \\ &= \frac{\theta^t (1 - \theta)^{m-t} \#^*(m, t)}{\theta^t (1 - \theta)^{m-t} \#(m, t)} \\ &= \frac{\#^*(m, t)}{\#(m, t)} \end{aligned}$$

es decir $y = \frac{\#^*(N, T_N)}{\#(N, T_N)}$

Consideremos ahora algunas reglas de decisión:

i) Muestrear con un tamaño de muestra fijo m

$$\#(m, t) = \binom{m}{t} \quad y \quad \#^*(m, t) = \binom{m-1}{t-1}$$

pues se debe escoger los t puntos de los m que se tienen en donde la variable vale 1 y de manera similar para $\#^*$ escogiendo los $(t-1)$ necesarios de los $(m-1)$.

$$y = \frac{\#^*(m, t)}{\#(m, t)} = \frac{\binom{m-1}{t-1}}{\binom{m}{t}} = \frac{t}{m}$$

ii) Muestrear hasta obtener c éxitos y entonces parar el procedimiento.

Sea m fijo entonces

$$h(m, c) = \binom{m-1}{c-1} \quad \text{y} \quad h^*(m, c) = \binom{m-2}{c-2}$$

pues el m -ésimo paso debe ser tal que $X_m = 1$

$$y(m, c) = \frac{\binom{m-1}{c-1}}{\binom{m-2}{c-2}} = \frac{c-1}{m-1}$$

es decir
$$y = \frac{c-1}{N-1}$$

iii) Considerar al $(1,1)$ como punto de decisión entonces

$$y(m, t) \begin{cases} 1 & \text{si } (m, t) = (1, 1) \\ 0 & \text{e.o.l.} \end{cases}$$

es decir $y = X_1$. Como puede verse este es un caso particular de ii) cuando $c=1$.

Antes de dar las condiciones suficientes y necesarias para la existencia de un único estimador insesgado de θ , veremos algunos teoremas y ciertas definiciones de términos que usaremos.

Definición 2.14.-

- a) Para todo punto (n, t_n) , n es llamado el índice del punto.
- b) El índice de una región es el supremo de los índices de sus puntos.

- c) Una región se dice que es finita si existe un $m \in \mathbb{N}$ tal que los índices de los posibles puntos son menores que m .
- d) Una región se dice que es cerrada si la suma de las probabilidades de los puntos de decisión en la región es uno.

Teorema 2.15.-

Una condición suficiente para que R sea una región cerrada es que

$$\liminf_{n \rightarrow \infty} \frac{D(n)}{n^{1/2}} = 0$$

donde $D(n)$ es el número de posibles puntos de índice n en R .

Corolario 2.16.-

Si el número de posibles puntos de índice n en una región R es acotado entonces la región es cerrada.

Definición 2.17.-

Una región R se dice que es simple, si para cualquiera dos puntos posibles en R , cualquier otro punto sobre la recta que los une es posible. En otras palabras, R es una región simple si el conjunto de puntos posibles de índice n forma un intervalo.

Teorema 2.18.-

Una condición necesaria y suficiente para que V sea el único estimador insesgado de θ es que para toda región R finita y cerrada, R sea simple.

Como se vio anteriormente la Cota de Crámer y Rao está definida como:

$$CIV = \frac{1}{E(N) E \left(\frac{\partial}{\partial \theta} \ln f(x|\theta) \right)^2}$$

con la propiedad de que $\text{Var}(T) \geq CIV$ para cualquier estimador T del parámetro θ que sea insesgado.

Como esta CIV depende de la $E(N)$, la cual a su vez depende del criterio de decisión que se adopte, entonces a distintos planes de decisión corresponderán diferentes valores de CIV; de aquí la necesidad de tener la "óptima" CIV que quedará determinada utilizando la siguiente definición.

Definición 2.19.-

Un plan de muestreo S^* (con su correspondiente regla de decisión) y un estimador T , se dice que son óptimos en $\theta = \theta^*$, si entre todos los procedimientos con tamaño de muestra esperado en θ^* no mayor que el de S , no existe un estimador insesgado con varianza menor en θ^* que T .

Si un estimador T y un plan de muestreo S son óptimos para todo $\theta \in \Theta$ entonces se dirá que dicho estimador es eficiente.

Consistencia de ciertos estimadores secuenciales.

Dado un estimador insesgado y la estadística suficiente para un cierto parámetro de interés, se puede usar el procedimiento de Rao-Blackwell para obtener otro estimador como la esperanza del estimador insesgado dada la estadística suficiente. Wolfowitz consideró la consistencia de tales estimadores para el parámetro p de una distribución binomial en el caso secuencial, Loynes y Berk* estudiaron la consistencia de los estimadores secuenciales del tipo de Rao-Blackwell para el caso general.

Sea $X_1, X_2, \dots, X_n, \dots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con espacio muestral Ω . Sea N la variable aleatoria que denota el número de observaciones necesarias para tomar una decisión en un procedimiento secuencial, por lo tanto $N = 1, 2, \dots$ con la propiedad que para cada $n \geq 1$ el evento $\{N = n\}$ solo depende de las observaciones X_1, X_2, \dots, X_n y supongamos que $P(N < \infty) = 1$.

Sea $X^n = (X_1, X_2, \dots, X_n)$ y

Z_i^n la i -ésima estadística de orden de X^n

Se tendrán las siguientes hipótesis:

- 1.- Para cada n , podemos obtener V_n una estadística suficiente.
- 2.- Para cada n , V_n es una función de Z_n^n (esto se cumple si

*(1969).

$V_n = Z_n^n$ o si V_n es una estadística suficiente minimal).

- 3.- Se puede construir un estimador insesgado T para θ , -
es decir,

$$T = T(X_1, X_2, \dots) \quad \text{tal que} \quad E(T) = \theta$$

El estimador T depende solo de un número finito de obser-
vaciones (esto se desprende de la suposición de que $P(N < \infty) = 1$); hay una regla de decisión con M (número -
de observaciones necesarias para tomar una decisión) pa-
ra T tal que si $M=n$, T es una función de X_1, \dots, X_n sola-
mente.

- 4.- Dado un conjunto de reglas de decisión para un procedi-
miento secuencial, existe con probabilidad 1 una regla -
tal que su tamaño de muestra requerido para tomar una --
decisión (M) es menor o igual que el tamaño de muestra -
requerido (N) para cualquier otra regla del conjunto, es
to es:

$$P(N \geq M) = 1$$

Cuando $N=n$ implica que $V_n = V_N = V$. Si se está trabajando con
una sucesión de reglas de decisión, se tendrá N_i para la i -ésima regla
de decisión por lo que se denotará

$$V_{N_i} = V^i$$

Ahora se verá la construcción de los estimadores de Rao-----
Blackwell. Dado un estimador insesgado T y (N, V) estadística suficien-
te definiremos el nuevo estimador de θ como:

$$U = U(N) = E(T/N, V)$$

o equivalentemente

$$U = U_n(v) \quad \text{cuando } N=n \text{ y } V_n = v \text{ donde}$$

$$U_n(v) = E(T / N=n, V_n = v)$$

$$= E(T I(N=n)^* / V_n = v) \\ = E(T(N=n) / V_n = v)$$

Las suposiciones 1 y 4 implican que $U_n(v)$ es una función de X_1, X_2, \dots, X_n . Entonces U es también un estimador insesgado para θ con varianza menor a la correspondiente varianza de T . Por 2, U es una función simétrica de X_1, \dots, X_n . También es importante señalar que U es el único estimador insesgado de mínima varianza cuando (N, V) es una estadística completa.

Wolfowitz demostró que $\{U_n(V)\}$ es una sucesión consistente cuando se estima el parámetro p de una distribución binomial, esta consistencia se refiere a que $\lim_{i \rightarrow \infty} n_{0,i} = \infty$ donde $n_{0,i}$ es el valor más pequeño de n tal que

$$P(N_i = n) > 0$$

teniendo una sucesión de reglas de decisión Loynes** después demostró que en el caso general la consistencia es muy débil a menos de que se tengan más condiciones sobre las N_i 's, estas condiciones dependen de la distribución en cuestión.

Estimación por doble muestreo.

Supongamos que estamos interesados en el problema de estima-

*Si A es un evento, $I(A)$ denotará la función indicadora de A , e.d., va a ser 1 en A y 0 en otro lado.

**Loynes. The consistency of certain sequential estimators. *Annals M. Statist.*, 40, 568 - 574.

ción de un parámetro desconocido θ con acuracidad* asignada usando el mínimo de observaciones posible. La acuracidad estaría en términos de la desviación estandar, esto es, el estimador tendría varianza $v(\theta)$, alguna función dada de θ . Otra posibilidad es estimar por medio de un intervalo de confianza al $(1 - \alpha)$ por ciento suponiendo una longitud fija (esta posibilidad se discutirá en secciones posteriores). En la estadística clásica es por lo general imposible obtener un estimador con las propiedades deseadas debido a que se tiene una muestra de tamaño fijo, sin embargo, al usar métodos secuenciales también se tienen algunas dificultades como el construir un esquema de muestreo secuencial que cumpla ciertas características, otra desventaja es el cálculo y análisis que se hacen en cada paso; para evitar este último problema sobre todo, Cox** propone un esquema de doble muestreo. La idea básica es usar una muestra preliminar para determinar que tan grande debe ser la muestra total para obtener las propiedades deseadas. Los métodos de doble muestreo de Cox son diferentes a aquellos que se usan en la inspección industrial, ya que en el último caso la segunda muestra, si se toma, es de tamaño fijo.

La teoría de doble muestreo desarrollada por Cox está hecha en base a muestras grandes, aunque en la práctica las aproximaciones son buenas a pesar de tener muestras pequeñas, por ejemplo para cualquier valor de θ el tamaño esperado de muestra en este esquema es más grande que el correspondiente tamaño esperado de muestra del "mejor" procedimiento secuencial.

Supongamos que θ es un parámetro desconocido el cual queremos estimar teniendo una varianza igual a una función dada de θ $v(\theta)$. Sea una muestra aleatoria de tamaño m fija, supongamos que podemos construir $T^{(m)}$ un estimador de θ tal que:

$$1. \quad T^{(m)} \text{ es un estimador insesgado de } \theta \text{ con varianza } v(\theta) / m.$$

*La acuracidad o exactitud se refiere al tamaño de las desviaciones respecto a la media verdadera.

** (1952).

2.- El coeficiente de asimetría γ_1 de $T^{(m)}$ es asintóticamente $\gamma_1(\theta)/m^{1/2}$ y γ_2 la kurtosis de $T^{(m)}$ es $O(m^{-1})^*$.

3.- Las medias y los errores estándar asintóticos pueden ser obtenidos para $a(T^{(m)})$, $v(T^{(m)})$ y para cualquier combinación de estas funciones por expansión en series.

Sean n_0 el tamaño de muestra inicial.

$$n(\theta) = v(\theta) / a(\theta)$$

$$m(\theta) = 1 / n(\theta)$$

y T_1 el estimador de θ para la muestra inicial.

entonces $n(T_1)$ nos da el tamaño de muestra necesario para obtener la acuracidad deseada. Ahora se toma una segunda muestra de tamaño $\max(0, n(T_1) - n_0)^{**}$. Sea T_2 el estimador de θ para la segunda muestra. Sea

$$T = \begin{cases} (n_0 T_1 + (n(T_1) - n_0) T_2) / n(T_1) & \text{si } n(T_1) \geq n_0 \\ T_1 & \text{si } n(T_1) < n_0 \end{cases}$$

En particular si $T^{(m)}$ es la media muestral, T es la media de la muestra conjunta.

*Una función f se dice que es $O(\varepsilon)$ cuando $\varepsilon \rightarrow \varepsilon_0$ si existe $M \in \mathbb{R}^+$ tal que

$$\left| \lim_{\varepsilon \rightarrow \varepsilon_0} f(\varepsilon) / \varepsilon \right| < M$$

**Si $n(T_1) - n_0$ no es un entero, se toma el entero más cercano a este valor.

Sean $E_1(A) = E(A / n(T_1) \geq n_0)$

$$E_2(A) = E(A / n(T_1) < n_0)$$

entonces

$$\begin{aligned} E(T) &= E_1((n_0 T_1 + (n(T_1) - n_0) T_2) / n(T_1)) P(n(T_1) \geq n_0) \\ &\quad + E_2(T_1) P(n(T_1) < n_0) \\ &= E_1(n_0 T_1 / n(T_1) + T_2 - n_0 T_2 / n(T_1)) P(n(T_1) \geq n_0) \\ &\quad + E_2(T_1) P(n(T_1) < n_0) \\ &= (n_0 E(T_1 / n(T_1)) + E_1(T_2) - n_0 E(T_2 / n(T_1))) P(n(T_1) \geq n_0) \\ &\quad + E_2(T_1) P(n(T_1) < n_0) \end{aligned}$$

para encontrar una expresión más sencilla de $E(T)$ se hará la siguiente hipótesis:

- 4.- $n_0 < n(\theta)$ y la distribución de $m(T_1)$ es tal que el evento $n(T_1) < n_0$ tiene probabilidad 0.

entonces

$$E(T) = n_0 E_1(m(T_1) T_1) + E_1(T_2) - n_0 E_1(T_2 m(T_1))$$

pero $E_1(T_2) = E(T_2 / T_1) = E(T_2) = \theta$ pues las muestras son independientes.

por lo tanto

$$E(T) = n_0 E_1(T_1 m(T_1)) + \theta - n_0 E_1(T_2 m(T_1))$$

$$= n_0 E_1 (T_1 a(T_1) / v(T_1)) + \theta - n_0 E_1 (T_2 a(T_2) / v(T_2))$$

En esta expresión se tienen funciones de $a(T^{(m)})$ y $v(T^{(m)})$, por lo que aplicando la hipótesis 3 se tiene que:

$$E_1 (T_1 a(T_1) / v(T_1)) = \theta a(\theta) / v(\theta) + R_1$$

y

$$E_1 (T_2 a(T_2) / v(T_2)) = \theta a(\theta) / v(\theta) + R_2$$

por lo que

$$\begin{aligned} E(T) &= n_0 (\theta a(\theta) / v(\theta) + R_1) + \theta - n_0 (\theta a(\theta) / v(\theta) + R_2) \\ &= \theta + (R_1 - R_2) n_0 \end{aligned}$$

donde $(R_1 - R_2) n_0$ es una función $O(n_0^2)$

por lo tanto

$$E(T) = \theta + O(n_0^2)$$

Análogamente se obtiene que:

$$\text{Var}(T) = a(\theta) n_0^2 (1 + O(n_0^2))$$

Ejemplo.- Estimación de la media (θ) de una población normal con una varianza dada a .

Supongamos que la varianza poblacional σ^2 es conocida. Para cada m fija, el mejor estimador de la media es: \bar{X} donde

$$\bar{X} = \sum_{i=1}^m X_i / m \quad \text{con } \text{Var}(\bar{X}) = \sigma^2/m$$

en este caso $n(\theta) = \sigma^2/a$, por lo que el tamaño de la muestra es constante. Este procedimiento es equivalente a tomar una sola muestra de tamaño el entero más cercano a σ^2/a .

A continuación se demostrará que el procedimiento de doble muestreo tiene un tamaño esperado de muestra más grande que el mejor procedimiento secuencial bajo la suposición que n_0 es grande. Para mayor facilidad se considerará que el parámetro en cuestión es unidimensional.

Sea T definido como antes, entonces $n(T) = n(T_1)$ por lo que

$$E(T) = E(n(T_1)) = E(v(T_1) / a(T_1))$$

usando la hipótesis 3 tenemos que:

$$E(v(T_1)/a(T_1)) \approx v(\theta) / a(\theta) = n(\theta) + O(n_0^2)$$

Para cualquier procedimiento de estimación secuencial, dado un estimador $\hat{\theta}$ insesgado para θ , se tiene que:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{E(N) E\left(\frac{\partial}{\partial \theta} \ln f\right)^2}$$

donde f es la función de densidad de la población. Ahora si tomamos un tamaño de muestra fijo m , para las distribuciones Normal, Poisson y Binomial, la varianza de la media muestral (mejor estimador* de la media poblacional en los casos anteriores) es $v(\theta)/m$ donde $v(\theta)$ es:

$$v(\theta) = \frac{1}{E\left(\frac{\partial}{\partial \theta} \ln f\right)^2}$$

*En el sentido que es insesgado y tiene mínima varianza.

entonces considerando estimadores tales que $\text{Var}(\hat{\theta}) \leq a(\theta)$, para las distribuciones anteriores se tiene:

$$E(N) \geq \frac{1}{\text{Var}(\hat{\theta}) E\left(\frac{\partial \ln f}{\partial \theta}\right)^2} = \frac{v(\theta)}{\text{Var}(\hat{\theta})} \geq \frac{v(\theta)}{a(\theta)} = n(\theta)$$

la cual se satisface asintóticamente para todas las distribuciones.

Estimación secuencial del tamaño de una población finita.

Un problema que ha sido atacado desde diferentes puntos de vista es el que se refiere a la estimación del tamaño de una población finita, este problema desde el punto de vista secuencial se ha tratado principalmente utilizando 5 reglas de decisión diferentes, las cuales se verán en esta sección. Estas reglas han dado lugar a resultados sobre N (tamaño de la población en estudio), dichos resultados solamente serán mencionados pues la demostración de los mismos se omitirá pero se darán referencias acerca de cada una de las demostraciones necesarias.

Para una mejor comprensión de este método, se hará una equivalencia entre la población en estudio y una urna con un número desconocido N de bolas blancas. La estimación de N que se desea estará basada en el siguiente esquema de muestreo:

- 1er. paso.- Se saca una bola blanca y se pinta de negro.
- 2o. paso.- Se regresa a la urna y se saca otra bola.
- 3er. paso.- Si la bola extraída es negra se repite el segundo paso.

Si la bola extraída es blanca se pinta de negro y se regresa al segundo paso.

En cada una de las extracciones se anota el tipo de bola que se obtuvo.

Estamos interesados en dos problemas:

- 1.- Encontrar un regla de decisión t .
- 2.- Estimar N después de haber terminado de muestrear.

Sean b_i, n_i el número de bolas blancas y bolas negras respectivamente, observadas en las primeras i extracciones ($b_i + n_i = i$). Entonces considerar las siguientes reglas de decisión:

- A) Sea $A > 0$ un entero fijo, entonces $t_A = A$
Esta primera regla se refiere a tomar una muestra de tamaño fijo A independientemente de los resultados obtenidos.
- B) Sea $B > 0$ un entero fijo, entonces $t_B = \inf \{ i/n_i = B \}$.
Aquí se irá muestreando las bolas hasta obtener B (valor fijo) bolas negras y entonces pasar a estimar N .
- C) Sea $C > 0$ un número fijo, entonces
 $t_C = \inf \{ i/n_i > C b_i \} = \inf \{ i > (C+1)b_i \}$
En este caso se terminará de muestrear hasta que la proporción entre bolas negras y bolas blancas exceda a una cantidad C fija.
- D) Sea $-\infty < D < \infty$ fijo, entonces
 $t_D = \inf \{ i/n_i > \max(1, b_i \ln b_i + b_i D) \}$
 $= \inf \{ i/i > \max(b_i + 1, b_i \ln b_i + b_i (D+1)) \}$

E) Sea $\{D_j\}$ una sucesión de reales tales que

$$\lim_{j \rightarrow \infty} D_j = \infty \quad \text{entonces}$$

$$t_E = \inf \left\{ i/n_i > \max(1, b_i \ln b_i + b_i D_{b_i}) \right\}$$

Desde que $b_i \leq N$ para toda i , cada una de estas reglas es cerrada por lo que con probabilidad 1 terminan en un número finito de pasos. De estas reglas, la regla B es la que más se ha estudiado*. La regla D ha sido analizada por Darling y Robbins† quienes demostraron que para cualquier α tal que $0 < \alpha < 1$ y una elección apropiada de D se tiene que:

$$P_N(B_D = N) \gg 1 - \alpha$$

uniformemente en N , donde B_D es el total de bolas blancas observadas después de terminar de muestrear.

De acuerdo a las relaciones entre i y $N \rightarrow \infty$ se establecerán las distribuciones límites de b_i . Sea $U_i = N - b_i$ el número de bolas blancas no observadas en la muestra de tamaño i (= número de bolas blancas en la urna después de la i -ésima extracción). Dado que las relaciones entre b_i, n_i, U_i son lineales, basta obtener la distribución límite de una de ellas. Sea Φ la función de distribución normal con media 0 y varianza 1.

Teorema 2.20.-

Sea $N \rightarrow \infty$ entonces

a) Si $i = (2N\lambda_N)^{1/2}$ donde $\lambda_N \rightarrow 0$

$$\Rightarrow P_N(n_i = 0) \rightarrow 1$$

*Goodman (1953), Chapman (1954) y Darroch (1958).
† (1967).

b) Si $i = (2N \lambda_N)^{1/2}$ donde $\lambda_N \rightarrow \lambda$ y $0 < \lambda < \infty$

$$\Rightarrow P_N(n_i = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad k=0,1,2,\dots$$

c) Si $i = Na_N$ donde $\{N^{-1/2} a_N\} \rightarrow \mu$ y $-\infty < \mu < \infty$

$$P_N \left[\frac{b_i - E(b_i)}{(\text{Var}(b_i))^{1/2}} \leq x \right] \rightarrow \Phi(x) \quad -\infty < x < \infty$$

d) Si $i = N \ln N + Na_N$ donde $a_N \rightarrow a$ y $-\infty < a < \infty$

$$\Rightarrow P_N(u_i = k) \rightarrow \frac{e^{-\lambda} \lambda^k}{k!} \quad k=0,1,2,\dots$$

$$\text{donde } \lambda = e^{-a}$$

e) Si $i = N \ln N + Na_N$ donde $a_N \rightarrow \infty$

$$\Rightarrow P_N(b_i = N) \rightarrow 1$$

Los incisos anteriores han sido probados por diferentes autores. Por ejemplo Renyi además de los ya mencionados anteriormente.

Estimación de N por máxima verosimilitud.

Samuel consideró la estimación de N por máxima verosimilitud (N), dice que si al terminar de observar las bolas se tienen b blancas y n bolas negras entonces $N = N(b, n)$ no depende de la regla utilizada y por lo tanto la distribución de N tampoco dependerá de la regla de decisión. Para cualquier regla de decisión t , sea $P_N(b, n/t)$ la probabilidad de obtener exactamente b bolas blancas y n bolas negras cuando se

haya terminado de muestrear. Se puede demostrar que:

$$P(b, n / t) = \frac{(N)_b h(b, n)}{N^{b+n}} \quad \begin{array}{l} b = 1, 2, \dots \\ n = 0, 1, \dots \end{array}$$

donde $(N)_i = N(N-1)(N-2)\dots(N-i+1)$ y $h(b, n)$ depende solamente de t .
Por ejemplo para la regla A:

$$h(b, n) = \sum_{j=0}^b \frac{(-1)^{b-j} j^A}{j! (b-j)!} \quad A = b + n$$

para la regla B:

$$P_N(t_B = k) = \frac{(N)_{k-B} h(k-B)}{N^k} \sum_{j=0}^{k-B} \frac{(-1)^{k-B-j} j^{k-1}}{j! (k-B-j)!} \quad k = B+1, B+2, \dots$$

y para la regla C:

$$\begin{aligned} P_N(B_C = k) &= P_N(t_C = [\gamma k] + 1) \\ &= \frac{(N)_k h_C(k)}{N^{[\gamma k] + 1}} \quad k = 1, 2, \dots, N \end{aligned}$$

donde $\gamma = C + 1$, $[\gamma k]$ el mayor entero menor o igual que γk y $h_C(k)$ son constantes definidas por:

$$h_C(1) = 1$$

$$h_C(k) = \left(1 - \sum_{i=1}^{k-1} \binom{k}{i} h_C(i) k^{-([\gamma i] + 1)} \right) k^{[\gamma k] + 1} (k!)^{-1} \quad k \geq 2$$

Las demostraciones de las afirmaciones anteriores pueden verse en los libros de Samuel[†] Darroch*, Lewontin y Prout.

[†] (1968)
^{*} (1958)

Sea $N(b, n)$ el estimador máximo-verosímil de N , por lo tanto $N(b, n)$ es un entero positivo tal que maximiza $(N)_b / N^{b+n}$ (si b y n son tales que $h(b, n) \neq 0$). Mediante un breve análisis obtenemos que:

$$N(b, 0) = \infty \quad \text{para } b \geq 2$$

$$N(1, n) = 1 \quad \text{para } n \geq 1$$

(el caso $b=1, n=0$ no es de interés pues la primera extracción siempre es una bola blanca).

Si vemos a $(N)_b / N^{b+n}$ como un función de variable positiva real* N ($N > b-1$), entonces para obtener el máximo podemos usar el método de la primera derivada, es decir:

$$\begin{aligned} \frac{(N)_b}{N^{b+n}} &= \frac{N(N-1)(N-2)\dots(N-b+1)}{N^{b+n}} \\ &= N^{1-b-n} (N-1)(N-2)\dots(N-b+1) \end{aligned}$$

$$\frac{d}{dN} \left[\frac{(N)_b}{N^{b+n}} \right] = (1-b-n) N^{-b-n} (N-1)(N-2)\dots(N-b+1)$$

$$+ N^{1-b-n} \frac{d}{dN} [(N-1)(N-2)\dots(N-b+1)]$$

$$\frac{d}{dN} [(N-1)\dots(N-b+1)] = (N-2)(N-3)\dots(N-b+1)$$

$$+ (N-1) \frac{d}{dN} [(N-2)(N-3)\dots(N-b+1)]$$

*Al finalizar el método nos podemos quedar con $[N]$ como una buena estimación de N .

$$\begin{aligned}
&= (N-2)(N-3) \dots (N-b+1) + \\
&\quad (N-1) \left[(N-3) \dots (N-b+1) + (N-2) \frac{d}{dN} [(N-3) \dots (N-b+1)] \right] \\
&= (N-2) \dots (N-b+1) + (N-1)(N-3) \dots (N-b+1) \\
&\quad + (N-1)(N-2) \frac{d}{dN} [(N-3)(N-4) \dots (N-b+1)] \\
&= \frac{(N)_b}{N(N-1)} + \frac{(N)_b}{N(N-2)} + (N-1)(N-2) \frac{d}{dN} [(N-3) \dots (N-b+1)] \\
&= \frac{(N)_b}{N(N-1)} + \frac{(N)_b}{N(N-2)} + \frac{(N)_b}{N(N-3)} + \\
&\quad (N-1)(N-2)(N-3) \frac{d}{dN} [(N-4) \dots (N-b+1)] \\
&= \dots = \frac{(N)_b}{N(N-1)} + \frac{(N)_b}{N(N-2)} + \dots + \frac{(N)_b}{N(N-b+2)} + \\
&\quad (N-1)(N-2) \dots (N-b+2) \frac{d}{dN} [N-b+1] \\
&= \frac{(N)_b}{N(N-1)} + \frac{(N)_b}{N(N-2)} + \dots + \frac{(N)_b}{N(N-b+1)} \\
&= \sum_{i=1}^{b-1} \frac{(N)_b}{N(N-i)} = \frac{(N)_b}{N} \sum_{i=1}^{b-1} (N-i)^{-1}
\end{aligned}$$

$$\frac{d}{dN} \left[\frac{(N)_b}{N^{b+n}} \right] = (1-b-n) N^{-1-b-n} (N)_b + N^{-1-b-n} \frac{(N)_b}{N} \sum_{i=1}^{b-1} (N-i)^{-1}$$

$$\begin{aligned}
&= (1-b-n)N^{-1-(b+n)}(N)_b + N^{-(b+n)}(N)_b \sum_{i=1}^{b-1} (N-i)^{-1} \\
&= N^{-(b+n)}(N)_b \left[(1-b-n)N^{-1} + \sum_{i=1}^{b-1} (N-i)^{-1} \right] \\
&= N^{-(b+n)}(N)_b \left[N^{-1} - (b+n)N^{-1} + \sum_{i=1}^{b-1} (N-i)^{-1} \right] \\
&= N^{-(b+n)}(N)_b \left[\sum_{i=0}^{b-1} (N-i)^{-1} - (b+n)N^{-1} \right]
\end{aligned}$$

igualando a cero esta expresi3n obtenemos que

$$N^{-(b+n)}(\tilde{N})_b \left[\sum_{i=0}^{b-1} (\tilde{N}-i)^{-1} - (b+n)\tilde{N}^{-1} \right] = 0$$

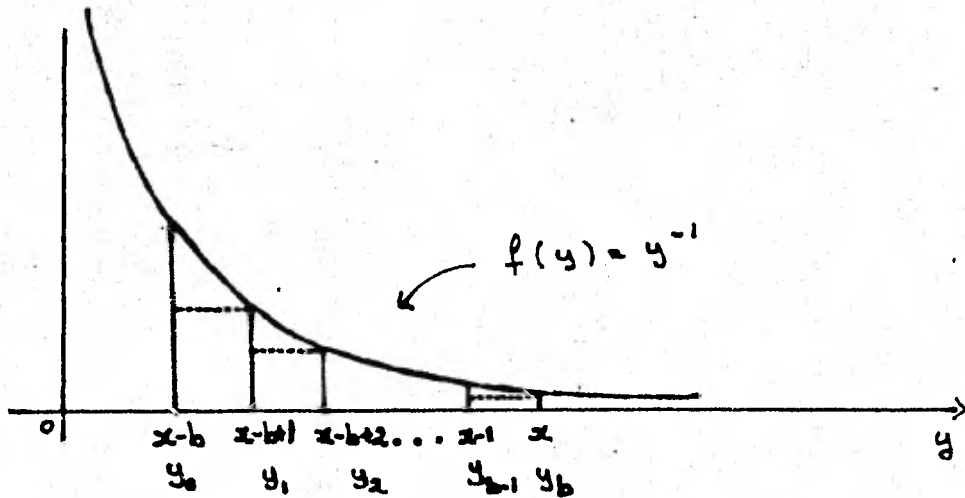
$$\therefore \sum_{i=0}^{b-1} (\tilde{N}-i)^{-1} - (b+n)\tilde{N}^{-1} = 0$$

$$\therefore \sum_{i=0}^{b-1} (\tilde{N}-i)^{-1} = (b+n)\tilde{N}^{-1} \quad \dots (2.21)$$

ahora para encontrar N consideremos:

$$\sum_{i=0}^{b-1} (x-i)^{-1} = \sum_{y=x-b+1}^x y^{-1}$$

$$f(y) = \frac{1}{y} \quad y > 0$$



Sea

$$\sum_{i=1}^b f(y_i) |y_i - y_{i-1}| = I$$

por lo tanto, utilizando la definición de integral, I es una suma inferior de:

$$\int_{x-b}^x f(y) dy$$

\therefore

$$I < \int_{x-b}^x f(y) dy$$

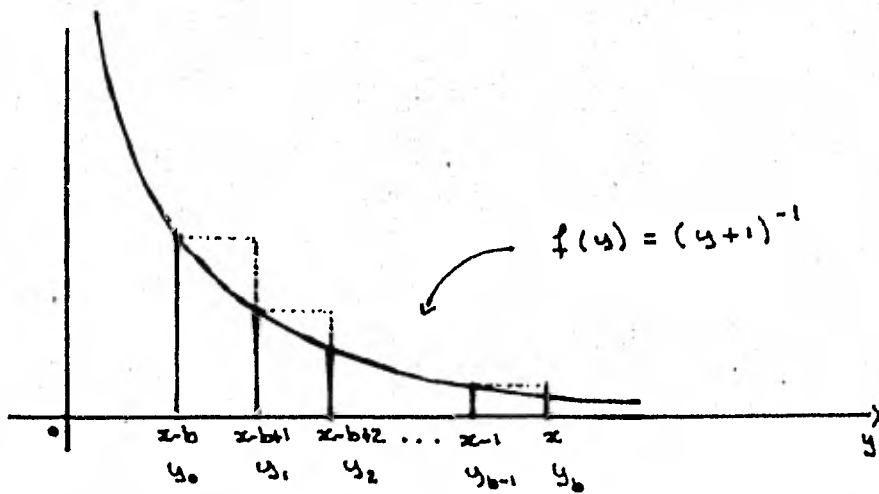
$$\begin{aligned} I &= \sum_{i=1}^b f(y_i) |y_i - y_{i-1}| = \sum_{i=1}^b f(x-b+i) |x-b+i - x-b+i-1| \\ &= \sum_{i=1}^b \frac{1}{x-b+i} = \sum_{i=0}^{b-1} \frac{1}{x-i} \end{aligned}$$

$$\int_{x-b}^x \frac{1}{y} dy = \ln y \Big|_{x-b}^x = \ln x - \ln(x-b) = \ln \frac{x}{x-b}$$

$$\therefore \sum_{i=0}^{b-1} \frac{1}{x-i} < \ln \frac{x}{x-b}$$

ahora sea

$$f(y) = \frac{1}{y+1}$$



Si

$$\sum_{i=0}^{b-1} f(y_i) |y_{i+1} - y_i| = I,$$

entonces, en este caso se tiene que I_1 es una suma superior de:

$$\int_{x-b}^x f(y) dy$$

$$\therefore \int_{x-b}^x f(y) dy < I_1$$

$$\begin{aligned} \int_{x-b}^x \frac{1}{y+1} dy &= \ln(y+1) \Big|_{x-b}^x = \ln(x+1) - \ln(x-b+1) \\ &= \ln\left(\frac{x+1}{x-b+1}\right) \end{aligned}$$

$$\begin{aligned} I_1 &= \sum_{i=0}^{b-1} f(x-b+i) |x-b+i+1 - x+b-i| \\ &= \sum_{i=0}^{b-1} \frac{1}{x-b+i+1} = \frac{1}{x-b+1} + \frac{1}{x-b+2} + \dots + \frac{1}{x} \\ &= \sum_{i=0}^{b-1} \frac{1}{x-i} \end{aligned}$$

$$\therefore \ln\left(\frac{x+1}{x-b+1}\right) < \sum_{i=0}^{b-1} \frac{1}{x-i}$$

Finalmente tenemos que:

$$\ln\left(\frac{x+1}{x-b+1}\right) < \sum_{i=0}^{b-1} \frac{1}{x-i} < \ln\left(\frac{x}{x-b}\right)$$

De aquí que para resolver (2.21) podemos utilizar como alter
nativa la ecuación:

$$\ln \frac{\tilde{N}}{\tilde{N} - b} = \frac{b + n}{\tilde{N}} \quad \dots \quad (2.22)$$

La solución de (2.22) está dada por:

$$N = \frac{b + n}{m(s)}$$

donde $s = \frac{b}{b + n}$

y $m(s)$ es la solución de $s = \frac{1 - e^{-m}}{m}$

La cual puede obtenerse de manera aproximada por métodos numéricos.

Por lo tanto, aproximadamente tendremos que:

$$N(b, n) = \frac{b + n}{m(s)} \quad \dots \quad (2.23)$$

De (2.23) se ve que el estimador máximo-verosímil es aproxi-
madamente proporcional al tamaño de muestra, en donde el factor de pro-
porcionalidad es una función únicamente de la proporción de bolas blan-
cas en la muestra.

Intervalos de confianza secuenciales.

Introducción.

El problema de estimación secuencial se reduce principalmente a encontrar una regla de muestreo tal que el estimador del parámetro en estudio tenga una cierta acuracidad, y además asegurar que se minimice el tamaño esperado de muestra. La acuracidad deseada puede darse en términos de la longitud de un intervalo de confianza, del error estándar o del coeficiente de variación del estimador.

Al igual que en la estadística clásica, la estimación de parámetros desde el punto de vista secuencial no solo se reduce al caso puntual sino que se lleva a la estimación por conjuntos (los cuales -- por lo general son intervalos de confianza cuando el parámetro es unidimensional). Este problema fue considerado por Stein y Wald* para la media de una población normal con varianza conocida y cierta longitud del intervalo deseado (encontraron que el mejor procedimiento secuencial, en términos del tamaño de muestra, es precisamente un procedimiento con tamaño de muestra fijo), más tarde Anscombe trata este problema pero considerando que la varianza es desconocida dando el coeficiente de confianza y la longitud del intervalo.

Intervalos de confianza para la media de una población normal.

Sean X_1, X_2, \dots v.a.i.i.d. como una normal con media μ y varianza σ^2 , se quiere estimar la media de la población por medio de un intervalo de confianza cuya longitud sea $2d$ y coeficiente de confianza $1 - \alpha$ ($0 < \alpha < 1$)

Para un tamaño de muestra fijo n , se tiene que si

* (1947)

$$y_n = \frac{\sum_{i=1}^n x_i}{n} \Rightarrow y_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

y de aquí utilizando el método pivotal de la estadística clásica, el intervalo de confianza para μ está dado por:

$$y_n - u_\alpha \frac{\sigma}{\sqrt{n}} < \mu < y_n + u_\alpha \frac{\sigma}{\sqrt{n}}$$

donde u_α es tal que

$$\int_{-u_\alpha}^{u_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 - \alpha$$

Si σ^2 es conocida entonces se tiene que:

$$d = u_\alpha \frac{\sigma}{\sqrt{n}}$$

$$d^2 = \frac{u_\alpha^2 \sigma^2}{n}$$

de donde podemos despejar n y obtener que:

$$n = \frac{u_\alpha^2 \sigma^2}{d^2} \quad \dots \quad (2.24)$$

por lo que se debería tomar observaciones de manera secuencial hasta que (2.24) se cumpla, pero como puede verse esta relación no depende del valor de las observaciones, es decir, bastará tomar una muestra de tamaño $n' = [n]$ para tener las características deseadas. Esto demuestra que el intervalo de confianza secuencial para μ coincide con el intervalo de confianza construido a partir de una muestra de tamaño fijo n' .

Si σ^2 es desconocida no podemos despejar n como en el caso anterior por lo que se han propuesto distintos procedimientos para poder tomar una decisión.

Un primer procedimiento es: Para cada valor fijo de n , estimar σ^2 , es decir, tomar:

$$\hat{\sigma}_n^2 = S_n^2 = \sum_{i=1}^n \frac{(X_i - \bar{Y}_n)^2}{n-1}$$

(el cual es insesgado para cada n) y fijar la regla de decisión de tal manera que se tomen observaciones hasta tener que:

$$S_n^2 \leq \frac{d_n^2}{u_\alpha^2}$$

Para relacionar este primer procedimiento con el segundo procedimiento propuesto se propone la siguiente transformación:

Para cada n y las respectivas X_1, X_2, \dots, X_n sean:

$$U_i = \frac{(iX_{i+1} - \sum_{j=1}^i X_j)^2}{i(i+1)} \quad i = 1, 2, \dots, n-1$$

$$\begin{aligned} \therefore U_i &= \frac{(iX_{i+1} - iY_i)^2}{i(i+1)} = \frac{i^2(X_{i+1} - Y_i)^2}{i(i+1)} \\ &= \frac{i(X_{i+1} - Y_i)^2}{i+1} \end{aligned}$$

$$X_{i+1} \sim N(\mu, \sigma^2)$$

$$Y_i = \sum_{j=1}^i \frac{X_j}{\lambda} \sim N\left(\mu, \frac{\sigma^2}{\lambda}\right)$$

como X_{i+1} y Y_i son independientes

$$X_{i+1} - Y_i \sim N\left(0, \sigma^2 + \frac{\sigma^2}{\lambda}\right)$$

$$\therefore \frac{X_{i+1} - Y_i}{\sigma\sqrt{1 + \frac{1}{\lambda}}} \sim N(0, 1)$$

$$\frac{(X_{i+1} - Y_i)^2}{\sigma^2\left(1 + \frac{1}{\lambda}\right)} \sim \chi^2(1)$$

$$\therefore \frac{i}{i+1} \frac{(X_{i+1} - Y_i)^2}{\sigma^2} = \frac{U_i}{\sigma^2} \sim \chi^2(1)$$

ahora

$$\begin{aligned} \sum_{i=1}^{n-1} U_i &= \sum_{i=1}^{n-1} \frac{i}{i+1} (X_{i+1} - Y_i)^2 \\ &= \sum_{i=1}^{n-1} \frac{i}{i+1} (X_{i+1} - Y_n + Y_n - Y_i)^2 \\ &= \sum_{i=1}^{n-1} \frac{i}{i+1} \left[(X_{i+1} - Y_n)^2 - 2(X_{i+1} - Y_n)(Y_i - Y_n) + (Y_i - Y_n)^2 \right] \end{aligned}$$

$$= \sum_{k=1}^{n-1} \frac{k}{k+1} (X_{k+1} - Y_n)^2 - 2 \sum_{k=1}^{n-1} \frac{k}{k+1} (X_{k+1} - Y_n)(Y_k - Y_n) +$$

$$+ \sum_{k=1}^{n-1} \frac{k}{k+1} (Y_k - Y_n)^2$$

$$\sum_{k=1}^{n-1} \frac{k}{k+1} (X_{k+1} - Y_n)^2 = \sum_{k=2}^n \frac{k-1}{k} (X_k - Y_n)^2 = \sum_{k=1}^n \frac{k-1}{k} (X_k - Y_n)^2$$

$$= \sum_{k=1}^n (X_k - Y_n)^2 - \sum_{k=1}^n \frac{(X_k - Y_n)^2}{k}$$

$$\sum_{k=1}^{n-1} \frac{k}{k+1} (X_{k+1} - Y_n)(Y_k - Y_n) = \sum_{k=1}^{n-1} \frac{k}{k+1} (Y_k - Y_n) [(k+1)Y_{k+1}$$

$$- kY_k - Y_n]$$

pues $X_{k+1} = (k+1)Y_{k+1} - kY_k$

$$= \sum_{k=1}^{n-1} \frac{k}{k+1} [(k+1)Y_{k+1} - kY_k - Y_k + Y_k - Y_n](Y_k - Y_n)$$

$$= \sum_{k=1}^{n-1} \frac{k}{k+1} [(k+1)(Y_{k+1} - Y_k) + (Y_k - Y_n)](Y_k - Y_n)$$

$$= \sum_{k=1}^{n-1} \frac{k}{k+1} [(k+1)(Y_{k+1} - Y_k)(Y_k - Y_n) + (Y_k - Y_n)^2]$$

$$= \sum_{k=1}^{n-1} k (Y_{k+1} - Y_k)(Y_k - Y_n) + \sum_{k=1}^{n-1} \frac{k}{k+1} (Y_k - Y_n)^2$$

$$\begin{aligned}
\therefore \sum_{i=1}^{n-1} u_i &= \sum_{i=1}^n (x_i - y_n)^2 - \sum_{i=1}^n \frac{(x_i - y_n)^2}{i} + \sum_{i=1}^{n-1} \frac{i}{i+1} (y_i - y_n)^2 \\
&\quad - 2 \sum_{i=1}^{n-1} i (y_{i+1} - y_i)(y_i - y_n) - 2 \sum_{i=1}^{n-1} \frac{i}{i+1} (y_i - y_n)^2 \\
&= \sum_{i=1}^n (x_i - y_n)^2 - \sum_{i=1}^n \frac{(x_i - y_n)^2}{i} - \sum_{i=1}^{n-1} \frac{i}{i+1} (y_i - y_n)^2 \\
&\quad - 2 \sum_{i=1}^{n-1} i (y_{i+1} - y_i)(y_i - y_n)
\end{aligned}$$

pero

$$\begin{aligned}
\sum_{i=1}^n \frac{(x_i - y_n)^2}{i} &= (x_1 - y_n)^2 + \sum_{i=2}^n \frac{(x_i - y_n)^2}{i} \\
&= (x_1 - y_n)^2 + \sum_{i=1}^{n-1} \frac{(x_{i+1} - y_n)^2}{i+1} \\
&= (x_1 - y_n)^2 + \sum_{i=1}^{n-1} \left[\frac{((i+1)y_{i+1} - i y_i - y_n)^2}{i+1} \right] \\
&= (x_1 - y_n)^2 + \sum_{i=1}^{n-1} \left[\frac{((i+1)(y_{i+1} - y_i) + (y_i - y_n))^2}{i+1} \right] \\
&= (x_1 - y_n)^2 + \sum_{i=1}^{n-1} \left[\frac{(i+1)^2 (y_{i+1} - y_i)^2}{i+1} + \right. \\
&\quad \left. + \frac{2(i+1)(y_{i+1} - y_i)(y_i - y_n)}{i+1} + \frac{(y_i - y_n)^2}{i+1} \right]
\end{aligned}$$

$$= (x_1 - y_n)^2 + \sum_{i=1}^{n-1} (i+1) (y_{i+1} - y_i)^2 + \sum_{i=1}^{n-1} \frac{(y_i - y_n)^2}{i+1} \\ + 2 \sum_{i=1}^{n-1} (y_{i+1} - y_i)(y_i - y_n)$$

$$\begin{aligned} \sum_{i=1}^{n-1} u_i &= \sum_{i=1}^n (x_i - y_n)^2 - (x_1 - y_n)^2 - \sum_{i=1}^{n-1} (i+1) (y_{i+1} - y_i)^2 - \\ &- 2 \sum_{i=1}^{n-1} (y_{i+1} - y_i)(y_i - y_n) - \sum_{i=1}^{n-1} \frac{(y_i - y_n)^2}{i+1} - \\ &- 2 \sum_{i=1}^{n-1} i (y_{i+1} - y_i)(y_i - y_n) - \sum_{i=1}^{n-1} \frac{i}{i+1} (y_i - y_n)^2 \\ &= \sum_{i=1}^n (x_i - y_n)^2 - (x_1 - y_n)^2 - \sum_{i=1}^{n-1} (i+1) (y_{i+1} - y_i)^2 - \\ &- 2 \sum_{i=1}^{n-1} (i+1) (y_{i+1} - y_i)(y_i - y_n) - \sum_{i=1}^{n-1} (y_i - y_n)^2 \\ &= \sum_{i=1}^n (x_i - y_n)^2 - (x_1 - y_n)^2 - \sum_{i=1}^{n-1} (y_i - y_n)^2 \\ &- \sum_{i=1}^{n-1} (i+1) \left[(y_{i+1} - y_i)^2 + 2(y_{i+1} - y_i)(y_i - y_n) \right] \\ &= \sum_{i=1}^n (x_i - y_n)^2 - (x_1 - y_n)^2 - \sum_{i=1}^{n-1} (y_i - y_n)^2 \\ &- \sum_{i=1}^{n-1} (i+1) \left[((y_{i+1} - y_i) + (y_i - y_n))^2 - (y_i - y_n)^2 \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n (x_i - \gamma_n)^2 - (x_1 - \gamma_n)^2 - \sum_{i=1}^{n-1} (\gamma_i - \gamma_n)^2 - \\
&\quad - \sum_{i=1}^{n-1} (i+1) [(\gamma_{i+1} - \gamma_n)^2 - (\gamma_i - \gamma_n)^2] \\
&= \sum_{i=1}^n (x_i - \gamma_n)^2 - (x_1 - \gamma_n)^2 - \sum_{i=1}^{n-1} (\gamma_i - \gamma_n)^2 - \\
&\quad - \sum_{i=1}^{n-1} (i+1)(\gamma_{i+1} - \gamma_n)^2 + \sum_{i=1}^{n-1} (i+1)(\gamma_i - \gamma_n)^2 \\
&= \sum_{i=1}^n (x_i - \gamma_n)^2 - (x_1 - \gamma_n)^2 - \sum_{i=2}^n i(\gamma_i - \gamma_n)^2 \\
&\quad + \sum_{i=1}^{n-1} i(\gamma_i - \gamma_n)^2 \\
&= \sum_{i=1}^n (x_i - \gamma_n)^2 - [(x_1 - \gamma_n)^2 + \sum_{i=2}^n i(\gamma_i - \gamma_n)^2] \\
&\quad + \sum_{i=1}^{n-1} i(\gamma_i - \gamma_n)^2
\end{aligned}$$

però

$$\gamma_n = \frac{\sum_{j=1}^n x_j}{1} = x_1$$

$$= \sum_{i=1}^n (x_i - \gamma_n)^2 - \sum_{i=1}^{n-1} i(\gamma_i - \gamma_n)^2 + \sum_{i=1}^{n-1} i(\gamma_i - \gamma_n)^2$$

$$\therefore \sum_{i=1}^{n-1} u_i = \sum_{i=1}^n (x_i - \gamma_n)^2$$

$$S_n^2 = \frac{\sum_{i=1}^{n-1} u_i}{n-1}$$

Por lo tanto el primer procedimiento queda expresado como si que: Se van a tomar observaciones hasta que

$$\sum_{i=1}^{n-1} u_i \leq \frac{d^2 n(n-1)}{u_{\alpha/2}^2}$$

para obtener finalmente el intervalo deseado como:

$$[Y_n - d, Y_n + d]$$

Otro procedimiento utilizado es el propuesto por Anscombe -- cuando n es pequeño. Se muestreará hasta tener que:

$$\sum_{i=1}^{n-1} u_i \leq f(n) \left(n - 2.676 - \frac{1}{2} u_{\alpha/2}^2 \right) \quad n \geq 4$$

donde $f(n)$ es una función $O(n^{1/2})$ cuando $n \rightarrow \infty$. La función $f(n)$ más común es:

$$f(n) = \frac{d^2 n}{u_{\alpha/2}^2}$$

basándose en el primer procedimiento descrito, es decir:

$$\sum_{i=1}^{n-1} u_i \leq \frac{d^2 n}{u_{\alpha/2}^2} \left(n - 2.676 - \frac{1}{2} u_{\alpha/2}^2 \right)$$

Para este caso se tiene que:

$$E(N) = \frac{\sigma^2 U^2}{d^2} + \frac{1}{2}(1 + U^2)$$

donde $\frac{\sigma^2 U^2}{d^2}$ es el número de observaciones que se toman si σ^2 es conocida. Entonces el método de Anscombe no incrementa demasiado el tamaño de muestra requerido en promedio.

Ray también trata de encontrar un intervalo de confianza para la media μ considerando que σ^2 es desconocida. Su método puede describirse como sigue:

Sea X_1, X_2, \dots v.a.i.i.d. como $N(\mu, \sigma^2)$ por lo tanto

$$Y_n = \frac{\sum_{i=1}^n X_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{Y_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$

Si $S_n^2 = \sum_{i=1}^n \frac{(X_i - Y_n)^2}{n-1}$ entonces

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$\frac{\frac{Y_n - \mu}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{(n-1)S_n^2}{(n-1)\sigma^2}}} \sim t_{(n-1)}$$

pues y_n y S_n^2 son independientes para cada n .

∴

$$t^* = \frac{y_n - \mu}{\sqrt{\frac{S_n^2}{n}}} \sim t_{(n-1)}$$

$$P(-t_{n-1}^{1-\alpha/2} < t^* < t_{n-1}^{1-\alpha/2}) = 1 - \alpha$$

$$P\left(y_n - t_{n-1}^{1-\alpha/2} \sqrt{\frac{S_n^2}{n}} < \mu < y_n + t_{n-1}^{1-\alpha/2} \sqrt{\frac{S_n^2}{n}}\right) = 1 - \alpha$$

Dada la longitud ($2d$) del intervalo y el nivel de confianza:

$$2d = 2t_{n-1}^{1-\alpha/2} \sqrt{\frac{S_n^2}{n}}$$

$$S_n^2 = \frac{nd^2}{(t_{n-1}^{1-\alpha/2})^2}$$

de aquí que Ray* propone tomar observaciones hasta que:

$$S_n^2 \leq \frac{nd^2}{(t_{n-1}^{1-\alpha/2})^2}$$

Si se definen u_i 's como antes, entonces este método se simplifica en muestrear hasta que

$$\sum_{i=1}^{n-1} u_i \leq \frac{d^2 n(n-1)}{(t_{n-1}^{1-\alpha/2})^2}$$

Otro procedimiento es el propuesto por Starr*. Este procedimiento es una modificación del método de Wald cuando la varianza es desconocida, mencionada al principio de esta sección.

Sea X_1, X_2, \dots una sucesión de v.a.i.i.d. como una normal -- con media μ y varianza σ^2 . Para $n \geq 2$ se calculan Y_n y S_n^2 (la media y la varianza muestral de una muestra de tamaño n).

Sea α fijo y u tal que:

$$\int_{-u}^u \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 1 - \alpha$$

Se sabe que al tomar una distribución t con n grados de libertad cuando n tiende a infinito dicha distribución tenderá a una normal estandar (ver Johnson & Kotz) utilizando este hecho es como Starr desarrolla su método:

$$\text{Sea } \psi_n(x) = \int_{-\infty}^x f(t|n) dt \quad \text{con}$$

$$f(t|n) = \begin{cases} \frac{\Gamma(\frac{n+1}{2})}{(\sqrt{n}\pi)^{1/2} \Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{(n+1)}{2}} & t \in \mathbb{R} \\ 0 & \text{s.o.l.} \end{cases}$$

Para cada n podemos encontrar u_n tal que:

$$\psi_n(u_n) = 1 - \alpha$$

Por lo tanto $\{u_n\}_1^\infty$ es una sucesión que converge a u es decir:

$$\lim_{n \rightarrow \infty} u_n = u$$

Se muestreará hasta que el tamaño n sea tal que:

$$S_n^2 \leq \frac{d^2 n}{u_n^2}$$

Dado que no existe un procedimiento de tamaño de muestra fijo cuando σ^2 es desconocida, Stein propone un método de construcción basado en un muestreo en 2 etapas:

Tomar una muestra inicial de tamaño n_1 y estimar σ^2 mediante $S_{n_1}^2$.

Considerar ahora el método de Wald cuando σ^2 es conocida, - es decir, tomar una muestra de tamaño n tal que:

$$n \geq \frac{u_n^2 S_{n_1}^2}{d^2}$$

donde $u_{n_1}^2$ es el cuantil $(1 - \alpha/2)$ de una distribución t con n_1 grados de libertad. Sin embargo este método es intuitivamente ineficiente pues no utiliza toda la información de la muestra total (para estimar σ^2).

Hay un procedimiento que puede llamarse el Procedimiento de Ray modificado* que consiste en:

Fijar un $n_0 \geq 2$ y muestrear hasta obtener X_N donde N es el primer entero $n \geq n_0$ tal que:

$$S_n^2 \leq \frac{n d^2}{u_{n-1}^2} \quad \dots \quad (2.25)$$

entonces el intervalo de confianza para μ será:

$$[y_N - d, y_N + d] = I_N$$

Dicho intervalo es tal que:

$$P(\mu \in I_N) \geq 1 - \alpha$$

Esta aseveración se justifica con el siguiente desarrollo: ---

sean

$$I_n = [y_n - d, y_n + d]$$

$$y \quad J_n = \left[y_n - \frac{u_{n-1} S_n}{n^{1/2}}, y_n + \frac{u_{n-1} S_n}{n^{1/2}} \right] \quad \text{dos intervalos.}$$

como

$$d > \frac{u_{n-1} S}{n^{1/2}} \quad \text{se tiene que } J_n \subset I_n \text{ por lo que:}$$

$$P(\mu \in J_n) \leq P(\mu \in I_n)$$

pero

$$P(\mu \in J_n) = P\left(y_n - \frac{u_{n-1} S_n}{n^{1/2}} \leq \mu \leq y_n + \frac{u_{n-1} S_n}{n^{1/2}} \right)$$

$$= P\left(|y_n - \mu| \leq \frac{u_{n-1} S_n}{n^{1/2}} \right)$$

$$= P\left(\frac{|y_n - \mu|}{\frac{S_n}{n^{1/2}}} \leq u_{n-1} \right)$$

$$= \Psi_{n-1}(u_{n-1}) = 1 - \alpha$$

Como $\lim_{n \rightarrow \infty} S^2 = \sigma^2$ con probabilidad uno, entonces de (2.25) se sigue que $\sigma^2 \approx \frac{d^2 N}{u^2}$ al menos para N grande. por lo tanto:

$$E(N) \approx \frac{u^2 \sigma^2}{d^2}$$

Intervalos de confianza para la diferencia de medias.

Para construir un intervalo de confianza para la diferencia de medias de 2 poblaciones normales independientes, al igual que en estadística clásica es necesario suponer la igualdad de varianzas cuando estas son desconocidas,

Supongamos que las dos poblaciones en estudio tienen varianza $\sigma^2/2$. Entonces en este caso se obtendrán secuencialmente parejas de observaciones (X_i, X_i') , donde $X_i \sim N(\mu_1, \sigma^2/2)$ y $X_i' \sim N(\mu_2, \sigma^2/2)$. Para $n \geq 3$ definimos:

$$U_i = \left(i X_{i+1} - \sum_{j=1}^i X_j \right)^2 / i(i+1) \quad i=1, 2, \dots, n-1$$

y

$$U_i' = \left(i X_{i+1}' - \sum_{j=1}^i X_j' \right)^2 / i(i+1) \quad i=1, 2, \dots, n-1$$

Por lo tanto para cada sucesión de observaciones de las 2 poblaciones, tendremos otra sucesión a saber (U_i, U_i') .

Sea $V_i = U_i + U_i'$ para $i=1, 2, \dots$, por lo visto anteriormente se tiene que:

$$\frac{U_i}{\sigma^2/2} \sim \chi_{(1)}^2 \quad y \quad \frac{U_i'}{\sigma^2/2} \sim \chi_{(1)}^2$$

$$E(u_i) = \sigma^2/2 \quad \text{y} \quad E(u_i') = \sigma^2/2$$

por lo tanto $E(V_i) = \sigma^2/2 + \sigma^2/2 = \sigma^2$

es decir V_i es un estimador insesgado de σ^2 tal que:

$$\frac{2V_i}{\sigma^2} \sim \chi^2(2)$$

Sea $y_n = \bar{X} - \bar{X}'$ la diferencia de las medias muestrales para las 2 poblaciones. El intervalo de confianza para $\mu_1 - \mu_2$ con una longitud $2d$ y coeficiente de confianza $1 - \alpha$ será:

$$[y_N - d, y_N + d]$$

donde N es el primer entero $n \geq 3$ tal que:

$$\sum_{i=1}^{n-1} V_i \leq \frac{d^2 n (n - 1.505 - \frac{1}{4} u^2)}{u^2}$$

Esta regla da como resultado que:

$$E(N) = \frac{\sigma^2 u^2}{d^2} + \frac{(1 + u^2)}{4}$$

Tanto la regla de decisión como el tamaño esperado de muestra fueron obtenidos por Anscombe.

Ahora consideremos un caso más general, es decir, sean X_1, \dots, X_2, \dots y Y_1, Y_2, \dots dos sucesiones independientes de variables aleatorias. Las X 's normales con media μ_1 y varianza σ_1^2 y las Y 's normales con media μ_2 y varianza σ_2^2 . Se desea encontrar un intervalo de

confianza I de longitud $2d$ con un coeficiente de confianza $1 - \alpha$ para el parámetro $\mu = \mu_1 - \mu_2$.

Si σ_1^2 y σ_2^2 son conocidas, se procederá como sigue:

Se toman n observaciones de X y t observaciones de Y , se definen \bar{X}_n y \bar{Y}_t como siempre, es decir:

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \quad \text{y} \quad \bar{Y}_t = \frac{\sum_{i=1}^t Y_i}{t}$$

$$\text{Si } I = [\bar{X}_n - \bar{Y}_t - d, \bar{X}_n - \bar{Y}_t + d]$$

se tiene un intervalo de longitud $2d$ centrado en $\bar{X}_n - \bar{Y}_t$ entonces:

$$\begin{aligned} P(\mu \in I) &= P(\bar{X}_n - \bar{Y}_t - d < \mu < \bar{X}_n - \bar{Y}_t + d) \\ &= P(-d < \bar{X}_n - \bar{Y}_t - \mu < d). \end{aligned}$$

$$\begin{aligned} \text{dado que } \bar{X}_n - \bar{Y}_t &\sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{t}) \\ &= N(\mu, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{t}) \end{aligned}$$

$$\begin{aligned} P(\mu \in I) &= P(|\bar{X}_n - \bar{Y}_t - \mu| < d) \\ &= P\left[\frac{|\bar{X}_n - \bar{Y}_t - \mu|}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{t}}} < \frac{d}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{t}}} \right] \\ &= 2 \Phi\left[\frac{d}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{t}}} \right] - 1 \end{aligned}$$

donde Φ es la función de distribución de un normal estándar.

Sean u y v tales que:

$$2 \Phi(u) - 1 = 1 - \alpha$$

$$v = \frac{u^2}{d^2}$$

si deseamos que $P(u \in I) \geq 1 - \alpha$, entonces

$$2 \Phi \left[\frac{d}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{t}}} \right] - 1 \geq 2 \Phi(u) - 1$$

\therefore

$$\frac{d}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{t}}} \geq u$$

$$\frac{d}{u} \geq \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{t}}$$

$$\frac{1}{v} = \frac{d^2}{u^2} \geq \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{t}$$

de aquí n y t deben de satisfacer:

$$\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{t} \leq \frac{1}{v} \quad \dots \quad (2.26)$$

pero no son únicos, por lo que tendremos que encontrar n^* y t^* tal -- que son mínimos y satisfacen (2.26).

Si se trata a n y t como variables continuas, se querrá minimizar

$$n = n + t \quad \text{sujeto a} \quad (2.26)$$

obteniendose

$$n^* = v \sigma_1 (\sigma_1 + \sigma_2)$$

$$t^* = v \sigma_2 (\sigma_1 + \sigma_2)$$

por lo que $n^* / t^* = \sigma_1 / \sigma_2$

y el tamaño total de muestra es: $n^* = v(\sigma_1 + \sigma_2)^2$

Si las varianzas σ_1^2 y σ_2^2 son desconocidas, se utiliza el procedimiento de Robbins*. Este procedimiento consiste en:

- i) Un esquema de muestreo que en cada paso nos diga de que población tomar la siguiente observación,
- ii) Una regla de terminación que proporcione n y t (y de aquí construir I como ya se definió).

$$\text{Sean } S_{1i}^2 = \sum_{k=1}^i \frac{(x_k - \bar{x}_i)^2}{i+1} \quad \text{y}$$

$$S_{2i}^2 = \sum_{k=1}^i \frac{(y_k - \bar{y}_j)^2}{j-1}$$

Se toman n_0 (fija) ≥ 2 observaciones de X y de Y . En cualquier paso se han tomado i observaciones de X y j observaciones de Y , es decir, $n = i + j \geq 2n_0$

Si $\frac{i}{j} \leq \frac{S_{1i}}{S_{2j}}$ se toma otra observación de X .

Si $\frac{i}{j} > \frac{S_{1i}}{S_{2j}}$ se toma otra observación de Y .

Para este caso se tienen 3 posibles reglas de terminación (más o menos equivalentes). Sean $\{u_n\}$ la sucesión de cuantiles definida anteriormente y $\{v_n\}$ otra sucesión, tal que:

* (1967).

$$V_n = \frac{u_n^2}{d^2}$$

Regla 1.- Terminar de muestrear con el primer $n \geq 2n_0$ tal -- que si r y t son el número de observaciones tomadas de cada población, con $n = r + t$ se tiene que:

$$n \geq V_n (S_{1r} + S_{2t})^2$$

Regla 2.- Lo mismo que en la regla 1 solo que se terminará de muestrear si:

$$\frac{S_{1r}^2}{r} + \frac{S_{2t}^2}{t} \leq \frac{1}{V_n}$$

Regla 3.- Lo mismo que en las anteriores pero con:

$$r \geq V_n S_{1r} (S_{1r} + S_{2t})$$

$$y \quad t \geq V_n S_{2t} (S_{1r} + S_{2t})$$

Como puede verse estas reglas están basadas en las características que se tenían cuando las varianzas eran conocidas. Además si se define n_i el tamaño de muestra determinado por cada una de las reglas entonces se puede checar fácilmente que:

$$n_1 \leq n_2 \leq n_3$$

Un método general para la construcción de intervalos de confianza secuenciales de longitud acotada.

Sea X_1, X_2, \dots, X_n una muestra aleatoria fija de tamaño n de una población con función de densidad $f(x/\theta)$ con θ el parámetro de

la población. Se desea construir un intervalo de confianza para θ cuya longitud sea menor o igual que $2d$. Para cada n , se consideran 2 estadísticas L_n y U_n (que no dependen de d) tales que $L_n < U_n$ casi seguramente y además se tenga que:

$$\lim_{n \rightarrow \infty} P(L_n \leq \theta \leq U_n) = 1 - \alpha$$

(es decir, para n grande (L_n, U_n) es un intervalo de confianza para al $(1 - \alpha) \cdot 100\%$ aproximadamente).

Se define N como la variable aleatoria que toma el valor n si n es el primer entero mayor o igual que n_0 (entero fijo) tal que:

$$U_n - L_n \leq 2d$$

y de aquí se tomará (L_N, U_N) como el intervalo de confianza deseado.

CAPITULO III

PRUEBAS DE HIPOTESIS
SECUENCIALES

Método General.

Desde el punto de vista secuencial también se presenta el problema de probar para una población X , con función de densidad $f(x / \theta)$ las hipótesis:

$$H_0: \theta \in \omega_0 \quad \text{vs.} \quad H_1: \theta \in \omega_1$$

mediante un proceso de este tipo en base a la información que se obtenga a partir de muestras.

De manera general se buscará un esquema de muestreo y una regla de decisión tales que para cada m fijo dividen al espacio muestral en 3 regiones mutuamente excluyentes:

- a) La región R_m^0 , tal que si la muestra cae dentro de esta región se aceptará H_0 .
- b) La región R_m^1 , donde se aceptará H_1 si la muestra cae dentro de esta región.
- y c) La región $R_m^c = (R_m^0 \cup R_m^1)^c$, que es el conjunto de muestras tales que no proporcionan suficiente información para tomar una decisión.

Función Característica Operante.

Por lo anterior la decisión final (aceptar H_0 o aceptar H_1) está en función de las observaciones obtenidas, lo cual hace razonable el pensar que el aceptar H_0 al final de una prueba secuencial es un evento por lo que tiene sentido preguntarse por su probabilidad.

Además las observaciones al provenir de una población con función de densidad $f(x / \theta)$ van a depender de θ , y por lo tanto el evento anterior va a depender también del valor del parámetro θ . Entonces la probabilidad en cuestión es:

$$P(\text{aceptar } H_0 \text{ al final de la prueba} / \theta) = L(\theta)$$

la cual podrá ser evaluada para todo $\theta \in \Omega$. A $L(\theta)$ se le conoce como la función característica operante, que de alguna manera "es un indicador del comportamiento de la prueba respecto a los errores y aciertos que se pueden cometer".*

En un proceso secuencial se pedirá que el esquema de muestreo y la regla de decisión sean tales que:

$$L(\theta_0) \geq 1 - \alpha \quad \forall \theta_0 \in \omega_0$$

$$L(\theta_1) \leq \beta \quad \forall \theta_1 \in \omega_1$$

donde α y β son las cotas de las probabilidades de cometer el error tipo I y el error tipo II respectivamente.

Tamaño Esperado de Muestra.

Una segunda función en la cual se describe el comportamiento de un proceso secuencial es el tamaño esperado de muestra, que es el valor medio del tamaño de muestra necesario para tomar una decisión, por lo cual también dependerá del valor θ .

Esta función puede ser utilizada para comparar 2 o más métodos secuenciales y obtener cual es el "mejor" de ellos en el sentido que necesita menos observaciones para tomar una decisión.

*Ver Mendoza.

Prueba de Wald para hipótesis simples:

Supongamos que tenemos una población con función de densidad $f(x / \theta)$. Se desea probar:

$$H_0: \theta = \theta_0 \quad \text{vs.} \quad H_1: \theta = \theta_1$$

Si se cuenta con una muestra de m observaciones que se han obtenido de manera secuencial, se puede utilizar el método de Wald -- para tomar una decisión. Este método consiste en:

1o.- Construir el cociente Λ_m definido como:

$$\Lambda_m = \frac{f(x_1, x_2, \dots, x_m / \theta_0)}{f(x_1, x_2, \dots, x_m / \theta_1)}$$

2o.- Seleccionar dos números A y B tales que $0 < A < B$ que estarán en función de α y β .

3o.- Definir la partición de Ω_m como sigue:

$$R_m = \{ (x_1, x_2, \dots, x_m) \mid A < \Lambda_m < B \}$$

$$R_m^0 = \{ (x_1, x_2, \dots, x_m) \mid \Lambda_m \geq B \}$$

$$R_m^1 = \{ (x_1, x_2, \dots, x_m) \mid \Lambda_m \leq A \}$$

4o.- El criterio de decisión es el descrito al principio de esta sección.

Lo único que hace falta para definir por completo la prueba anterior es determinar las funciones de α y β que nos darán una elección adecuada de A y B .

Dados α y β fijos, mediante un sencillo desarrollo* se verifica que:

$$B \leq \frac{1-\alpha}{\beta} \quad \text{y} \quad A \geq \frac{\alpha}{1-\beta}$$

pero aún no se determinan de manera única A y B . Si se toma como medida auxiliar los valores A' y B' donde:

$$B' = \frac{1-\alpha}{\beta} \quad \text{y} \quad A' = \frac{\alpha}{1-\beta}$$

Wald demuestra que la probabilidad de cometer un error en la decisión terminal ($\alpha + \beta$) no aumenta, además si α y β son pequeños no hay mucha diferencia entre estos y los nuevos errores (que se cometen en la decisión terminal al definir R_m, R_m^0, R_m^1 con A' y B').

Dado este tipo de prueba (y su correspondiente regla de decisión) es de interés conocer la forma de $L(\theta)$, para esto Wald demuestra la existencia de una función $h(\theta)$ con las siguientes propiedades:

$$1) h(\theta) \neq 0 \quad \forall \theta \in \Theta \quad (1)$$

$$2) E \left[\left(\frac{f(x|\theta_0)}{f(x|\theta)} \right)^{h(\theta)} \right] = 1$$

con las cuales se llega* a que:

$$L(\theta) = \frac{1 - A^{h(\theta)}}{B^{h(\theta)} - A^{h(\theta)}} \quad \forall \theta \in \Theta \quad (2)$$

Como se vio en un principio la función tamaño esperado de muestra nos da una idea del número de observaciones necesarias para tomar una decisión, para la prueba de Wald es posible encontrar su forma explícita aproximadamente si se hacen 2 suposiciones:

*Ver Mendoza y/o Wald.

- 1.- La probabilidad de que el proceso termine con un número finito de pasos es 1 (es decir, es un procedimiento cerrado).
- 2.- La cota de la desigualdad que se cumple es una buena aproximación al valor que toma el cociente cuando termina la prueba.

Con lo anterior se obtiene (ver Wald) que:

$$E(N / \theta) = \frac{\ln B L(\theta) + \ln A(1 - L(\theta))}{E(Z / \theta)}$$

donde

N es la v.a. que denota el número de observaciones necesario para aceptar alguna de la 2 hipótesis.

$$Z = \ln \left[\frac{f(x|\theta_0)}{f(x|\theta_1)} \right]$$

$A, B, L(\theta)$ definidos como antes.

Como ya se dijo anteriormente, la función tamaño esperado de muestra puede utilizarse para comparar 2 procedimientos secuenciales, Wald demuestra finalmente que para cualquier otro procedimiento secuencial basado en A_m y tamaño esperado de muestra $E^*(N^* / \theta)$ se tiene que:

$$E(N / \theta = \theta_i) \leq E^*(N^* / \theta = \theta_i) \quad i=0,1$$

es decir, la prueba de Wald es más eficiente en este sentido.

Pruebas secuenciales para hipótesis compuestas.

En la sección anterior se consideró la prueba de una hipótesis simple contra una hipótesis alternativa también simple. Sin embargo en la práctica se presenta también el problema de probar hipótesis compuestas, dichas hipótesis pueden surgir a partir de 2 situaciones:

- La hipótesis compuesta se refiere a los parámetros de interés y no hay parámetros de estorbo*.
- La hipótesis puede ser simple o compuesta en presencia de parámetros de estorbo.

Supongamos que X_1, X_2, \dots son v.a.i.i.d. con función de densidad $f(x / \theta)$, donde $\theta = (\theta_1, \theta_2, \dots, \theta_k)$. Se desea probar

$$H_0: \theta \in \omega_0 \quad \text{vs.} \quad H_1: \theta \in \omega_1$$

(en este caso, al igual que en la prueba de hipótesis simples, el espacio parametral Θ es dividido en 3 regiones, ω_0 en donde H_0 se prefiere, ω_1 en donde H_1 se prefiere y $\omega_2 = \omega_0 \cup \omega_1$ una región de indiferencia).

Para esta prueba se requiere que:

- $\max_{\theta \in \omega_0} P(\text{rechazar } H_0 \text{ al final de la prueba} / \theta) \leq \alpha$
 $= \max_{\theta \in \omega_0} (1 - L(\theta)) \leq \alpha$
- $\max_{\theta \in \omega_1} P(\text{aceptar } H_0 \text{ al final de la prueba} / \theta) \leq \beta$
 $= \max_{\theta \in \omega_1} L(\theta) \leq \beta$

*parámetro desconocido que carece de interés.

- La prueba termine en un número finito de pasos con probabilidad 1.

Prueba de Wald.

Wald propone el método de funciones ponderadas para probar dichas hipótesis. Supóngase que se tienen 2 distribuciones sobre θ , $w_0(\theta)$ y $w_1(\theta)$ tales que:

$$\int_{w_0} w_0(\theta) d\theta = \int_{w_1} w_1(\theta) d\theta = 1$$

entonces si

$$\Lambda_m = \frac{f^*(x|H_0)}{f^*(x|H_1)} = \frac{\int_{w_0} w_0(\theta) \prod_{i=1}^m f(x_i|\theta) d\theta}{\int_{w_1} w_1(\theta) \prod_{i=1}^m f(x_i|\theta) d\theta}$$

la regla de decisión es análoga a la utilizada en la sección anterior, es decir, dadas dos constantes A, B tales que $0 < A < B$ entonces:

Si $\Lambda_m > B$ se acepta H_0

$\Lambda_m \leq A$ se rechaza H_0

$A < \Lambda_m < B$ se necesita más información

Si $\alpha(\theta)$ y $\beta(\theta)$ son las probabilidades de error para cada θ con la regla de decisión anterior entonces

$$\int_{w_0} w_0(\theta) \alpha(\theta) d\theta \leq \int_{w_0} w_0(\theta) \alpha d\theta = \alpha$$

$$\int_{w_1} w_1(\theta) \beta(\theta) d\theta \leq \int_{w_1} w_1(\theta) \beta d\theta = \beta$$

De acuerdo con lo anterior se obtienen las siguientes relaciones:

$$\frac{1-A}{B-A} \leq \alpha \quad y \quad \frac{A(B-1)}{B-A} \leq \beta$$

o equivalentemente:

$$A \geq \frac{\alpha}{1-\beta} \quad B \leq \frac{1-\alpha}{\beta}$$

Ejemplo 1.-

Sea X que se distribuye como Bernoulli de parámetro p . Se desea probar la hipótesis:

$$H_0: p=1/2 \quad \text{vs.} \quad H_1: |p - 1/2| \geq \delta$$

Como H_0 es una hipótesis simple se tiene que:

$$w_0(p) = \begin{cases} 1 & \text{si } p=1/2 \\ 0 & \text{si } p \neq 1/2 \end{cases}$$

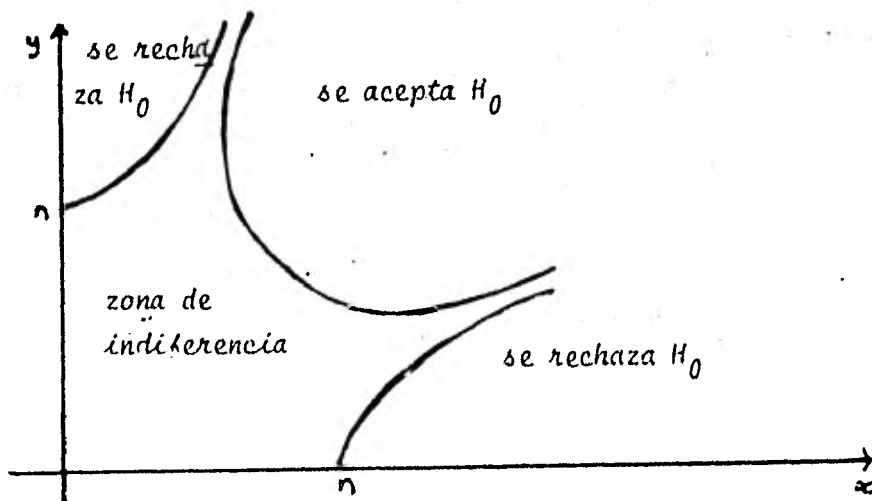
si se escoge $w_1(p)$ como sigue:

$$w_1(p) = \begin{cases} 1/2 & \text{si } p = 1/2 \pm \delta \\ 0 & \text{e.o.l.} \end{cases}$$

entonces se tiene que si se han hecho n observaciones de las cuales x han sido éxitos

$$\begin{aligned} \frac{f^*(x|H_0)}{f^*(x|H_1)} &= \frac{1 \cdot \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{n-x}}{\frac{1}{2} \left(\frac{1}{2} + \delta\right)^x \left(1 - \left(\frac{1}{2} + \delta\right)\right)^{n-x} + \frac{1}{2} \left(\frac{1}{2} - \delta\right)^x \left(1 - \left(\frac{1}{2} - \delta\right)\right)^{n-x}} \\ &= \frac{2 \left(\frac{1}{2}\right)^n}{\left(\frac{1}{2} + \delta\right)^x \left(\frac{1}{2} - \delta\right)^{n-x} + \left(\frac{1}{2} - \delta\right)^x \left(\frac{1}{2} + \delta\right)^{n-x}} \\ &= \frac{1}{\left[\left(\frac{1}{2} + \delta\right)^x \left(\frac{1}{2} - \delta\right)^{n-x} + \left(\frac{1}{2} - \delta\right)^x \left(\frac{1}{2} + \delta\right)^{n-x}\right] 2^{n-1}} \end{aligned}$$

Para n fija si $y=x$ y $x=n-x$ las regiones definidas para la regla de decisión quedan definidas por:



Ejemplo 2.-

Sean X_1, X_2, \dots v.a.i.i.d. como $N(\mu, \sigma^2)$ con μ y σ^2 desconocidos. Se desea probar:

$$H_0: \sigma^2 = \sigma_0^2 \quad \text{vs.} \quad H_1: \sigma^2 = \sigma_1^2 \quad (\sigma_1 > \sigma_0)$$

En este caso $\theta = (\mu, \sigma^2)$ donde μ es un parámetro de estorbo. Sean

$$w_0(\theta) = \begin{cases} 1/2c & -c \leq \mu \leq c \quad \sigma^2 = \sigma_0^2 \\ 0 & \text{e.o.l.} \end{cases}$$

$$w_1(\theta) = \begin{cases} 1/2c & -c \leq \mu \leq c \quad \sigma^2 = \sigma_1^2 \\ 0 & \text{e.o.l.} \end{cases}$$

$$\begin{aligned} \frac{f^*(x|H_0)}{f^*(x|H_1)} &= \frac{\int_{-c}^c \frac{1}{2c} \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma_0^2}} d\mu}{\int_{-c}^c \frac{1}{2c} \left(\frac{1}{\sqrt{2\pi}\sigma_1}\right)^n e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma_1^2}} d\mu} \\ &= \frac{\int_{-c}^c \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n e^{-\frac{\sum x_i^2 - 2\mu \sum x_i + n\mu^2}{2\sigma_0^2}} d\mu}{\int_{-c}^c \left(\frac{1}{\sqrt{2\pi}\sigma_1}\right)^n e^{-\frac{\sum x_i^2 - 2\mu \sum x_i + n\mu^2}{2\sigma_1^2}} d\mu} \\ &= \left(\frac{\sigma_1}{\sigma_0}\right)^n e^{\frac{\sum x_i^2}{2} \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)} \frac{\int_{-c}^c e^{-\frac{n}{2\sigma_0^2}(\mu - \bar{x})^2} e^{\frac{n\bar{x}^2}{2\sigma_0^2}} d\mu}{\int_{-c}^c e^{-\frac{n}{2\sigma_1^2}(\mu - \bar{x})^2} e^{\frac{n\bar{x}^2}{2\sigma_1^2}} d\mu} \end{aligned}$$

Si $c \rightarrow \infty$ entonces

$$\begin{aligned} \frac{f^*(z|H_0)}{f^*(z|H_1)} &\rightarrow \left(\frac{\sigma_1}{\sigma_0}\right)^n e^{\frac{z \sum X_i^2}{2} \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)} \frac{\sqrt{2\pi} \frac{\sigma_0}{\sqrt{n}} e^{-\frac{n \bar{X}^2}{2\sigma_0^2}}}{\sqrt{2\pi} \frac{\sigma_1}{\sqrt{n}} e^{-\frac{n \bar{X}^2}{2\sigma_1^2}}} \\ &= \left(\frac{\sigma_1}{\sigma_0}\right)^{n-1} e^{-\frac{1}{2\sigma_0^2} [\sum X_i^2 - n\bar{X}^2]} e^{-\frac{1}{2\sigma_1^2} [\sum X_i^2 - n\bar{X}^2]} \\ &= \left(\frac{\sigma_1}{\sigma_0}\right)^{n-1} e^{\sum (X_i - \bar{X})^2 \left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}\right)} \\ &= \left(\frac{\sigma_1}{\sigma_0}\right)^{n-1} e^{(n-1)S_n^2 \left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}\right)}. \end{aligned}$$

Alguna veces este método es modificado ligeramente, dicha modificación consiste en pedir que w_0 y w_1 sean tales que:

$$\int_{w_0} w_0(\theta) d\theta = 1$$

y

$$\int_{S_1} w_1(\theta) d\theta = 1$$

donde S_1 es la frontera de w_1 , por lo que el cociente que se construye es:

$$\frac{f^*(z|H_0)}{f^*(z|H_1)} = \frac{\int_{w_0} w_0(\theta) \prod_{i=1}^n f(x_i|\theta) d\theta}{\int_{S_1} w_1(\theta) \prod_{i=1}^n f(x_i|\theta) dS_1}$$

Prueba de Cox .

Cox desarrolla un método basado en el de Bartlett (se verá posteriormente), para probar hipótesis compuestas. Si el tamaño de muestra no es muy pequeño, se puede construir esta prueba secuencial basada en estimadores máximo-verosímiles suponiendo además una distribución asintótica.

Supongamos que X_1, X_2, \dots son v.a.i.i.d. con densidad $f(x | \theta, \phi)$ donde θ es el (los) parámetro (s) de interés y ϕ es el (los) parámetro (s) de estorbo. Se desea probar:

$$H_0: \theta = \theta_0 \quad \text{vs.} \quad H_1: \theta = \theta_1$$

Si ϕ es conocido, la prueba está basada en:

$$L^\circ(\underline{x} | \theta_0, \phi) - L^\circ(\underline{x} | \theta_1, \phi) \dots \quad (3.1)$$

donde L° denota el logaritmo de la función de verosimilitud de la muestra.

Si se expande (3.1) hasta los términos de 2o. grado alrededor de (θ, ϕ) se tiene:

$$(\theta_0 - \theta_1) \frac{\partial L^\circ(\underline{x} | \theta, \phi)}{\partial \theta} + \frac{1}{2} (\theta_0 - \theta_1) (\theta_0 + \theta_1 - 2\theta) \frac{\partial^2 L^\circ(\underline{x} | \theta, \phi)}{\partial \theta^2}$$

Si ϕ es desconocido, la prueba estará basada en:

$$L^\circ(\underline{x} | \theta_0, \hat{\phi}) - L^\circ(\underline{x} | \theta_1, \hat{\phi}) \dots \quad (3.2)$$

donde $\hat{\phi}$ es el estimador máximo-verosímil de ϕ . Ahora si se expande (3.2) de la misma manera que antes, alrededor del punto $(\theta, \hat{\phi})$ se tiene:

$$\begin{aligned}
 & (\theta_0 - \theta_1) \frac{\partial L^\circ(\mathbf{x} | \theta, \phi)}{\partial \theta} + \frac{1}{2} (\theta_0 - \theta_1) (\theta_0 + \theta_1 - 2\theta) \frac{\partial^2 L^\circ(\mathbf{x} | \theta, \phi)}{\partial \theta^2} \\
 & + (\theta_0 - \theta_1) (\hat{\phi} - \phi) \frac{\partial^2 L^\circ(\mathbf{x} | \theta, \phi)}{\partial \theta \partial \phi}
 \end{aligned}$$

Si $\hat{\theta}$ es el estimador máximo-verosímil de θ y si además se cumple que

$$\frac{\partial^2 L^\circ(\mathbf{x} | \hat{\theta}, \hat{\phi})}{\partial \theta \partial \phi} = 0$$

entonces

$$\begin{aligned}
 L^\circ(\mathbf{x} | \theta_0, \hat{\phi}) - L^\circ(\mathbf{x} | \theta_1, \hat{\phi}) &= \frac{1}{2} (\theta_0 - \theta_1) (\theta_0 + \theta_1 - 2\hat{\theta}) \frac{\partial^2 L^\circ(\mathbf{x} | \hat{\theta}, \hat{\phi})}{\partial \theta^2} \\
 &= (\theta_0 - \theta_1) \left(\frac{\theta_0 + \theta_1}{2} - \hat{\theta} \right) \frac{\partial^2 L^\circ(\mathbf{x} | \hat{\theta}, \hat{\phi})}{\partial \theta^2}
 \end{aligned}$$

Para muestras grandes se tiene que:

$$\begin{aligned}
 E \left[L^\circ(\mathbf{x} | \theta_0, \hat{\phi}) - L^\circ(\mathbf{x} | \theta_1, \hat{\phi}) \right] &= (\theta_0 - \theta_1) \left(\frac{\theta_0 + \theta_1}{2} - \hat{\theta} \right) E \left(\frac{\partial^2 L^\circ(\mathbf{x} | \theta, \phi)}{\partial \theta^2} \right) \\
 &= (\theta_0 - \theta_1) \left(\frac{\theta_0 + \theta_1}{2} - \theta \right) n E \left(\frac{\partial^2 \ln f(\mathbf{x} | \theta, \phi)}{\partial \theta^2} \right)
 \end{aligned}$$

Sea

$$I_{\theta\theta} = E \left(\frac{\partial^2 \ln f(\mathbf{x} | \theta, \phi)}{\partial \theta^2} \right)$$

$$E[L'(z|\theta_0, \hat{\phi}) - L'(z|\theta_1, \hat{\phi})] = (\theta_0 - \theta_1) \left(\frac{\theta_0 + \theta_1}{2} - \hat{\theta} \right) \cap I_{\theta_0}$$

esto indica que la prueba puede estar basada en:

$$T_n = n(\theta_0 - \theta_1) \left(\frac{\theta_0 + \theta_1}{2} - \hat{\theta} \right)$$

con $E(T_n) = n(\theta_0 - \theta_1) \left(\frac{\theta_0 + \theta_1}{2} - \theta \right)$

además Cox demostró que:

$$\text{Var}(T_n) = \frac{n(\theta_0 - \theta_1)^2}{\frac{I_{\theta\phi}^2}{I_{\phi\phi}} - I_{\theta\theta}} = n(\theta_0 - \theta_1)^2 \sigma^2(\theta)$$

donde

$$I_{\theta\phi} = E \left(\frac{1}{n} \frac{\partial^2 L^0(z|\theta, \phi)}{\partial \theta \partial \phi} \right)$$

$$I_{\phi\phi} = E \left(\frac{1}{n} \frac{\partial^2 L^0(z|\theta, \phi)}{\partial \phi^2} \right)$$

De igual manera que en el método de Wald, dados α y β se calculan

$$A = \frac{\alpha}{1-\beta} \quad \text{y} \quad B = \frac{1-\alpha}{\beta}$$

y la regla de decisión está dada por:

$$\begin{aligned} \text{Si } T_n > \widehat{c^2(\theta)} \ln B & \text{ se acepta } H_0 \\ T_n < \widehat{c^2(\theta)} \ln A & \text{ se rechaza } H_0 \\ \widehat{c^2(\theta)} \ln A < T_n < \widehat{c^2(\theta)} \ln B & \text{ se necesita más informa-} \\ & \text{ción.} \end{aligned}$$

La función característica operante y el tamaño esperado de muestra se construyen como en el método de Wald.

Prueba de Bartlett.

Igual que para la prueba de Cox, supongamos que X_1, X_2, \dots son v.a.i.i.d. con función de densidad $f(x/\theta, \phi)$, con θ parámetro(s) de interés y ϕ parámetro(s) de estorbo. Se desea probar:

$$H_0: \theta = \theta_0 \quad \text{vs.} \quad H_1: \theta = \theta_1$$

Esta prueba se basa en:

$$L^0(\underline{x} / \theta_0, \hat{\phi}_0) - L^0(\underline{x} / \theta_1, \hat{\phi}_1)$$

donde $\hat{\phi}_i$ es el estimador máximo-verosímil de ϕ dado que $\theta = \theta_i$ ($i=0,1$).

En la prueba de Cox, se llegó a:

$$L^0(\underline{x} / \theta_0, \hat{\phi}) - L^0(\underline{x} / \theta_1, \hat{\phi}) = (\theta_0 - \theta_1) \left(\frac{\theta_0 + \theta_1}{2} - \hat{\theta} \right) \frac{\partial^2 L^0(\underline{x} / \hat{\theta}, \hat{\phi})}{\partial \theta^2}$$

pero esta estadística puede expresarse en términos de $\frac{\partial L^0(\underline{x} / \theta, \hat{\phi})}{\partial \theta}$

y $\frac{\partial \mathcal{L}^0(\underline{x}|\theta, \phi)}{\partial \phi}$ al expandir $\frac{\partial \mathcal{L}^0(\underline{x}|\hat{\theta}, \hat{\phi})}{\partial \theta}$ y $\frac{\partial \mathcal{L}^0(\underline{x}|\hat{\theta}, \hat{\phi})}{\partial \phi}$

alrededor del punto (θ, ϕ) , obteniéndose:

$$\begin{aligned} \mathcal{L}(\underline{x}|\theta_0, \hat{\phi}) - \mathcal{L}(\underline{x}|\theta_1, \hat{\phi}) &= \frac{n^{1/2}(\theta_0 - \theta_1) y_{\theta\theta} (y_{\phi} y_{\theta\phi} / y_{\phi\phi} - y_{\theta})}{(y_{\theta\theta}^2 / y_{\phi\phi} - y_{\theta\theta})} \\ &+ n y_{\theta\theta} (\theta_0 - \theta_1) \left(\frac{\theta_0 + \theta_1}{2} - \theta \right) \end{aligned}$$

donde

$$y_{\theta} = n^{-1/2} \frac{\partial \mathcal{L}^0(\underline{x}|\theta, \phi)}{\partial \theta}$$

$$y_{\phi} = n^{-1/2} \frac{\partial \mathcal{L}^0(\underline{x}|\theta, \phi)}{\partial \phi}$$

$$y_{\theta\theta} = \frac{1}{n} \frac{\partial^2 \mathcal{L}^0(\underline{x}|\theta, \phi)}{\partial \theta^2}$$

$$y_{\phi\phi} = \frac{1}{n} \frac{\partial^2 \mathcal{L}^0(\underline{x}|\theta, \phi)}{\partial \phi^2}$$

$$y_{\theta\phi} = \frac{1}{n} \frac{\partial^2 \mathcal{L}^0(\underline{x}|\theta, \phi)}{\partial \theta \partial \phi}$$

En cambio para la prueba de Bartlett, la expansión es tal - que:

$$\begin{aligned} \mathcal{L}^0(\underline{x}|\theta_0, \hat{\phi}_0) - \mathcal{L}^0(\underline{x}|\theta_1, \hat{\phi}_1) &= n^{1/2} (\theta_0 - \theta_1) (y_{\phi} y_{\theta\phi} / y_{\phi\phi} - y_{\theta}) + \\ &+ n (\theta_0 - \theta_1) (y_{\theta\theta}^2 / y_{\phi\phi} - y_{\theta\theta}) \left(\frac{\theta_0 + \theta_1}{2} - \hat{\theta} \right) \end{aligned}$$

La esperanza y la varianza de la estadística de Bartlett, - están dadas (asintóticamente) por:

$$E[\mathcal{L}^0(\underline{x} | \theta_0, \hat{\phi}_0) - \mathcal{L}^0(\underline{x} | \theta_1, \hat{\phi}_1)] = n(\theta_0 - \theta_1) \left(\frac{I_{\theta_0 \phi}^2}{I_{\phi \phi}} - I_{\theta \theta} \right) \left(\frac{\theta_0 + \theta_1}{2} - \theta \right)$$

$$\text{Var}[\mathcal{L}^0(\underline{x} | \theta_0, \hat{\phi}_0) - \mathcal{L}^0(\underline{x} | \theta_1, \hat{\phi}_1)] = n(\theta_0 - \theta_1)^2 \left(\frac{I_{\theta \theta \phi}^2}{I_{\phi \phi}} - I_{\theta \theta} \right)$$

Si el criterio de decisión está dado por:

$$\text{si } \mathcal{L}^0(\underline{x} | \theta_0, \hat{\phi}_0) - \mathcal{L}^0(\underline{x} | \theta_1, \hat{\phi}_1) \geq \ln B \text{ se acepta } H_0$$

$$\mathcal{L}^0(\underline{x} | \theta_0, \hat{\phi}_0) - \mathcal{L}^0(\underline{x} | \theta_1, \hat{\phi}_1) \leq \ln A \text{ se rechaza } H_0$$

$$\ln A < \mathcal{L}^0(\underline{x} | \theta_0, \hat{\phi}_0) - \mathcal{L}^0(\underline{x} | \theta_1, \hat{\phi}_1) < \ln B \text{ se necesita más información.}$$

donde A y B se calculan como siempre, entonces Joanes probó que la probabilidad de error no excede a la ya especificada ($\alpha + \beta$).

Prueba de Nanak Chand.

Este método sirve para obtener pruebas secuenciales de hipótesis compuestas para una cierta clase de distribuciones. Esta prueba es para cuando se tiene más de un parámetro desconocido, y puede ser usada en una situación más general en la cual el logaritmo del cociente de verosimilitud en el n-ésimo paso puede ser factorizado en una función del (los) parámetro(s) de estorbo exclusivamente y una función de la muestra.

Sea X_1, X_2, \dots v.a.i.i.d. con función de densidad $f(\underline{x} | \theta_1, \theta_2)$
 $\theta_1 \in \mathcal{Q}_1$, $\theta_2 \in \mathcal{Q}_2$; con θ_1 y θ_2 funcionando independientemente. Se desea probar: $H_0: \theta_1 = \theta_{10}$ vs. $H_1: \theta_1 = \theta_{11}$ $\theta_2 \in \mathcal{Q}_2$

Sea $\hat{\theta}_2$ un estimador de θ_2 obtenido de una muestra preliminar independiente de las observaciones secuenciales.

Para $j=1,2,\dots$ sea

$$Y_j(\theta_2) = \frac{f(x_j; \theta_{10}, \theta_2)}{f(x_j; \theta_{11}, \theta_2)}$$

$$y \quad Z_j(\theta_2) = \ln Y_j(\theta_2)$$

Supongamos que para cada $m \geq 1$ se tiene que:

$$\sum_{j=1}^m Z_j(\theta_2) = g_1(\theta_2) g_2(\underline{x})$$

$$y \quad t = \frac{g_1(\theta_2)}{g_1(\hat{\theta}_2)} \quad \text{se distribuye independientemente de } \theta_1, \text{ y } \theta_2$$

Sea $\Delta_m = \sum_{j=1}^m Z_j(\hat{\theta}_2)$, dadas las probabilidades de error α y β , las cotas de terminación A y B quedarán determinadas por las siguientes ecuaciones:

$$\int_{\{t \leq 0\}} \frac{B^t - 1}{B^t - A^t} dG(t) + \int_{\{t > 0\}} \frac{1 - A^t}{B^t - A^t} dG(t) = \alpha$$

$$\int_{\{t \leq 0\}} \frac{B^t(1 - A^t)}{B^t - A^t} dG(t) + \int_{\{t > 0\}} \frac{A^t(B^t - 1)}{B^t - A^t} dG(t) = \beta$$

donde $G(t)$ es la función de distribución de t .

La regla de decisión está dada por:

$$\text{Si } \Delta_m \geq B \quad \text{se acepta } H_0$$

$$\Delta_m \leq A$$

se acepta H_1

$$A < \Delta_m < B$$

se necesita más información.

En este caso si es posible obtener expresiones para la función característica operante y el tamaño esperado de muestra, bajo la suposición de la existencia de la función $h=h(\theta_1, \theta_2)$ no nula que es solución de:

$$E \left[\frac{f(x; \theta_{10}, \theta_2)}{f(x; \theta_{11}, \theta_2)} \right]^h = 1$$

Dichas expresiones son:

$$L(\theta_1, \theta_2) = \begin{cases} \int_{t \leq 0} \frac{B^{th} - 1}{B^{th} - A^{th}} dG(t) + \int_{t > 0} \frac{1 - A^{th}}{B^{th} - A^{th}} dG(t) & \text{si } E(X(\theta_2)) \neq 0 \\ \left[\frac{\ln B \Pr(t > 0) + \ln A \Pr(t \leq 0)}{\ln B + \ln A} \right] & \text{si } E(X(\theta_2)) = 0 \end{cases}$$

y

$$E(N|\theta_1, \theta_2) = \begin{cases} \frac{\int (\ln A (1 - B^{th})t + \ln B (1 - A^{th})t) dG(t)}{(B^{th} - A^{th}) E(X(\theta_1))} & \text{si } E(X(\theta_2)) \neq 0 \\ \int \frac{\ln B \ln A t^2}{E(X^2(\theta_1))} dG(t) & \text{si } E(X(\theta_2)) = 0 \end{cases}$$

Pruebas secuenciales entre tres hipótesis.

En las secciones precedentes se ha expuesto la teoría de -- pruebas secuenciales que se refiere a 2 hipótesis ya sean simples o -- compuestas. Sin embargo hay muchas situaciones prácticas en las que -- se debe escoger entre tres o más un curso de acción. Para el caso en -- que se tengan tres hipótesis se han desarrollado 2 métodos diferentes -- basados en la prueba de Wald, los cuales pueden generalizarse para -- cuando se tienen k hipótesis.

Sean X_1, X_2, \dots, X_n v.a.i.i.d. con función de densidad -----
 $f(x/\theta)$. Se desea probar:

$$H_j: \theta \in \omega_j \quad j=0,1,2$$

donde $\omega_j \subset \Omega$ y $\omega_i \cap \omega_j = \emptyset$ si $i \neq j$, con probabilidades de error $\alpha_0, \alpha_1, \alpha_2$, es decir

$$P(\text{Rechazar } H_j \text{ al final de la prueba} \mid \theta \in \omega_j) \leq \alpha_j \\ j=0,1,2.$$

Para cada n , el espacio Ω_n (conjunto de todas las mues--
 tras de tamaño n) se particiona en 4 regiones, 3 de las cuales son de
 preferencia por aceptación y otra una región de continuación o indife--
 rencia.

Método de Sobel-Wald.

Para poder tomar una decisión acerca de la prueba de hipó--
 tesis propuesta, Sobel y Wald la dividen en 2 pruebas, a saber:

$$H_0: \theta \in \omega_0 \quad \text{vs.} \quad H_1: \theta \in \omega_1 \quad (3.3)$$

y

$$H_0: \theta \in \omega_0 \quad \text{vs.} \quad H_2: \theta \in \omega_2 \quad (3.4)$$

Tanto (3.3) como (3.4) se prueban utilizando el método de funciones ponderadas de Wald con probabilidades de error:

$$P(\text{rechazar } H_0 \text{ al final de la prueba (3.3) / } H_0 \text{ cierta}) \leq \frac{\alpha_1}{2}$$

$$P(\text{aceptar } H_0 \text{ al final de la prueba (3.3) / } H_1 \text{ cierta}) \leq \alpha_1$$

y

$$P(\text{rechazar } H_0 \text{ al final de la prueba (3.4) / } H_0 \text{ cierta}) \leq \frac{\alpha_2}{2}$$

$$P(\text{aceptar } H_0 \text{ al final de la prueba (3.4) / } H_1 \text{ cierta}) \leq \alpha_2$$

El criterio de decisión está basado en las decisiones de (3.3) y (3.4) de la siguiente forma:

Si en la prueba (3.3) se acepta:	y	Si en la prueba (3.4) se acepta:	Decisión final
H_1		H_0	H_1
H_0		H_0	H_0
H_0		H_2	H_2

en otro caso se toma una observación más.

La función característica operante en este caso queda determinada por la siguiente expresión:

$$\begin{aligned} L(\theta) &= P(\text{aceptar } H_0 \text{ al final de la prueba / } \theta) \\ &= P(\text{aceptar } H_0 \text{ en (3.3) y en (3.4) / } \theta) \\ &= 1 - P(\text{rechazar } H_0 \text{ en (3.3) o rechazar } H_0 \text{ en (3.4) / } \theta) \\ &= 1 - P(\text{rechazar } H_0 \text{ en (3.3) / } \theta) - P(\text{rechazar } H_0 \text{ en (3.4) / } \theta) \end{aligned}$$

$$= 1 - L_1(\theta) - L_2(\theta)$$

donde L_1 y L_2 son las funciones características para las pruebas (3.3) y (3.4) respectivamente.

Sin embargo no se ha podido encontrar una expresión para el tamaño esperado de muestra, solo se obtuvo una cota para el mismo en función de los tamaños esperados de muestra de cada una de las pruebas auxiliares, es decir:

$$E(N/\theta) \geq \max\{E(N \text{ en (3.3)/}\theta), E(N \text{ en (3.4)/}\theta)\}$$

Método de Armitage.

Armitage sugiere que para que el método de Sobel-Wald sea más eficiente, en vez de hacer solo 2 pruebas de hipótesis, se hagan las 3 pruebas posibles.

El plan de muestreo y la regla de decisión están dados de la siguiente manera: Muestrear hasta obtener un conjunto de resultados tales que para alguna H_i , H_i se prefiere a H_j para $j \neq i$ en base a cada prueba:

$$H_i: \theta \in \omega_i \quad \text{vs.} \quad H_j: \theta \in \omega_j$$

En este procedimiento no se conoce casi nada de la función tamaño esperado de muestra pues no se ha podido ni siquiera encontrar la función característica $L_j(\theta)$ igual a la probabilidad de aceptar H_j al final de la prueba dado θ , en una forma simple.

CAPITULO IV

OTRAS ALTERNATIVAS

En este último capítulo se tratará de dar una idea de algunos otros temas estadísticos que han sido analizados desde un punto de vista secuencial. Dichos temas, así como los tratados en los capítulos anteriores y algunos que no fueron mencionados, pueden ser ampliados y desarrollados más a fondo, por lo que al final del presente capítulo se dará una amplia bibliografía para aquellas personas interesadas en los mismos.

Región de confianza secuencial para la media de una población normal - multivariada.

Este problema fue trabajado por Khan quien consideró una población normal p -variada con vector de medias $\underline{\mu}$ y matriz de varianzas-covarianzas Σ diagonal, es decir

$$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_p^2 \end{pmatrix}$$

Se desea una región de confianza para $\underline{\mu}$ al $(1-\alpha) \cdot 100\%$ tal que la longitud del eje μ_i de la elipsoide no sea mayor que $2l_i$ ($i=1, 2, \dots, p$).

Sea n fijo y a un real tal que:

$$P(\chi_p^2 \leq a) = 1 - \alpha$$

entonces

$$P(\underline{\mu} \in S_p^\alpha) = 1 - \alpha \quad \text{con}$$

$$S_p^\alpha = \{ \underline{\mu} \mid n(\bar{X}_n - \underline{\mu})' \Sigma^{-1} (\bar{X}_n - \underline{\mu}) \leq a \}$$

donde

$$\bar{X}_n = \begin{pmatrix} \bar{x}_{1n} \\ \bar{x}_{2n} \\ \vdots \\ \bar{x}_{pn} \end{pmatrix}$$

$$\bar{x}_{in} = \frac{\sum_{j=1}^n X_{ij}}{n}$$

Considerar la elipsoide:

$$\sum_{i=1}^p \frac{n(\bar{x}_{in} - \mu_i)^2}{a\sigma_i^2} = 1$$

Como se desea que la longitud del eje μ_i sea menor o igual que $2l_i$ ($i=1, 2, \dots, p$) entonces

$$\frac{a\sigma_i^2}{n} \leq l_i^2 \quad \text{es decir} \quad n \geq \frac{a\sigma_i^2}{l_i^2}$$

Si las σ_i^2 's son conocidas, sea

$$n^0 = \max_i \left(\frac{a\sigma_i^2}{l_i^2} \right)$$

entonces el procedimiento secuencial es equivalente a un procedimiento de tamaño fijo n^0 (que concuerda con lo que se obtuvo anteriormente si $p=1$).

Si las σ_i^2 's son desconocidas, entonces la regla de terminación está dada por: Muestrear hasta el primer $n \geq p$ tal que:

$$n \geq \frac{a s_{in}^2}{l_i^2} \quad \text{para toda } i=1, \dots, p$$

Región de confianza para los parámetros de una regresión lineal.

Este es un caso parecido al anterior, el cual fue estudiado por Gleser. Sea y_1, y_2, \dots una sucesión de observaciones independientes con:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + u_i$$

con $u_i \sim N(0, \sigma^2)$.

Se desea encontrar una región R p -dimensional tal que
 $P(\beta \in R) = 1 - \alpha$ (donde β es el vector de parámetros (β_i) y
 la longitud del intervalo del eje β_i delimitado por R es menor o igual que $2d$, $i=1, \dots, p$)

Para cada n ($n \geq p$), se sabe que el estimador por máxima --
 verosimilitud de β tiene componentes insesgados con varianza mínima
 dentro de la clase de estimadores lineales e insesgados de β , por lo
 cual este hecho será usado para la construcción de la región deseada,

Sea n fijo, $n \geq p$, entonces

$$\hat{\beta}_{(n)} = (\mathbf{X}_n' \mathbf{X}_n)^{-1} \mathbf{X}_n' \mathbf{Y}_n$$

donde

$$\mathbf{Y}_n = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X}_n = \begin{pmatrix} x_{11} & x_{21} & \dots & x_{p1} \\ x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & & x_{pn} \end{pmatrix}$$

por lo tanto $E(\hat{\beta}_{(n)}) = \beta$, $\text{Var}(\hat{\beta}_{(n)}) = \sigma^2 (\mathbf{X}_n' \mathbf{X}_n)^{-1}$

además

$$\frac{(\hat{\beta}_{(n)} - \beta)' (\mathbf{X}_n' \mathbf{X}_n)^{-1} (\hat{\beta}_{(n)} - \beta)}{\hat{\sigma}^2} \sim F_{(p, n-p)}$$

La regla de decisión determina muestrear hasta encontrar el
 el primer $n \geq p$ tal que:

$$(\hat{\beta}_{(n)} - \beta)' (\mathbf{X}_n' \mathbf{X}_n)^{-1} (\hat{\beta}_{(n)} - \beta) \leq d^2 \hat{\sigma}^2$$

$\rightarrow P[F \leq d^2] \leq 1 - \alpha$ donde $F \sim F_{(p, n-p)}$

Intervalo secuencial para la mediana.

Sea X_1, X_2, \dots v.a.i.i.d. con una única mediana γ . Se desea probar la hipótesis:

$$H_0: \gamma = 0 \quad \text{vs.} \quad H_1: \gamma \neq 0$$

con probabilidades de error $\alpha = \beta$.

La decisión estará basada en un intervalo de confianza al $(1 - \alpha) \cdot 100\%$ para γ .

Sea n fijo, la muestra aleatoria X_1, X_2, \dots, X_n obtenida secuencialmente se ordena, es decir, se obtienen:

$$X_{n,1} \leq X_{n,2} \leq \dots \leq X_{n,n} \quad \text{las estadísticas de orden}$$

Sean $\{b_n\}$ y $\{a_n\}$ dos sucesiones definidas como:

$$b_n = \max \left(1, \left[\frac{n}{2} - K_{\alpha} \frac{n^{1/2}}{2} \right] \right)$$

$$a_n = n - b_n + 1$$

donde K_{α} es tal que $\bar{\Phi}(K_{\alpha}) = 1 - \alpha/2$

El intervalo de confianza tiene la forma:

$$I_n = (X_{n,b_n}, X_{n,a_n})$$

Se van a obtener observaciones hasta que $n \geq n_0$ (mínimo de observaciones que se toman) tal que:

$$X_{n,a_n} - X_{n,b_n} \leq 2d$$

finalmente si $0 \in I_N$ se acepta H_0

$0 \notin I_N$ se rechaza H_0

Prueba del rango ordenado.

Consideremos 2 poblaciones X y Y con funciones de distribución F y G respectivamente. Se desea probar:

$$H_0: F = G \quad \text{vs.} \quad H_1: G = F^k \quad 0 < k < 1$$

con probabilidades de error α y β .

Sean X_1, X_2, \dots, X_{n_1} una muestra aleatoria de la pob. X

Y_1, Y_2, \dots, Y_{n_2} una muestra aleatoria de la pob. Y

$s_1 < s_2 < \dots < s_{n_2}$ los rangos de Y_1, Y_2, \dots, Y_{n_2} en

la muestra combinada de tamaño $n = n_1 + n_2$.

Supongamos que las observaciones de X y Y son tomadas en grupos de m y n respectivamente por lo que en el t -ésimo paso se tendrán mt observaciones de X y nt observaciones de Y .

Se calculan $s_j^{(t)}$ para $j = 1, 2, \dots, nt$ y

$$\Delta_m = \frac{n^t [(m+n)t]!}{(s_1^{(t)} - 1)!} \prod_{j=1}^{nt} \frac{\Gamma(s_j^{(t)} + j(m-1))}{\Gamma(s_{j+1}^{(t)} + j(n-1))}$$

Dados α y β se calculan las cotas A y B de la manera usual. La regla de decisión tiene la misma forma de la del método de Wald para pruebas de hipótesis.

Planteamiento de un problema secuencial que puede ser resuelto por métodos bayesianos.

Sea X_1, X_2, \dots una sucesión de v.a.i.i.d. con función de densidad $f(x/\theta)$. Sea $r(\delta(x); \theta)$ la pérdida que se tiene cuando se toma una decisión $\delta(x)$ siendo θ el verdadero valor del parámetro. Además de considerar la pérdida, se toma en cuenta el costo de la experimentación $(C(N))$, es decir, el costo que se tiene al tomar N observaciones (con N aleatoria). La tarea del estadístico consiste en escoger una regla de terminación y una regla de decisión para estimar θ . Las siguientes son posibles consideraciones que contemplarla:

i) Presupuesto limitado (forzando a tener una cota sobre el costo total esperado de las observaciones). Esto es, sujeto a que $E_{\theta}(C(N)) \leq n_0$, tratarla de minimizar $E_{\theta}(r(\delta(x); \theta))$.

ii) Acuracidad limitada (con lo cual se tiene una cota sobre la pérdida esperada debida a malas decisiones). Esto es, tratarla de minimizar $E_{\theta}(C(N))$ sujeto a que $E_{\theta}(r(\delta(x); \theta)) \leq \rho$.

iii) Las 2 cosas anteriores. Esto es, tratarla de minimizar $E_{\theta}(C(N) + r(\delta(x); \theta))$.

Por lo general, no hay un procedimiento secuencial que satisfaga iii) uniformemente en θ a menos que el criterio sea modificado. Tal modificación puede estar dada por un criterio bayesiano de optimización.

Un plan secuencial se dice que es una solución de Bayes si es obtenido a partir de minimizar algo sobre todos los riesgos con respecto a una distribución apriori.

Como puede verse, este punto es el que posiblemente pueda extenderse mas ampliamente.

Bibliografía de información general.

Como se dijo en un principio, ahora se dará un bibliografía sobre el tema de Análisis secuencial, la cual se trato de que fuera lo más amplia posible.

Aivazian, S.A. (1959). A comparison of the optimal properties of the Neyman-Pearson and the Wald sequential probability ratio tests. Theor. Probability Appl. 4, 86-93.

Albert, G.E. (1947). A note on the fundamental identity of sequential analysis. Annals of Mathematical Statistics, 18', 593-596.

----- (1948). Correction to A note on the fundamental identity of sequential analysis. Annals of Mathematical Statistics, 19', 426-427.

Anastasi, Anne (1953). An empirical study of the applicability of sequential analysis to item selection. Educational and Psychological Measurement, 13, 3-13.

Anderson, T.W. (1960). A modification of the sequential probability ratio test to reduce the sample size. Ann. Math. Statist., 31', 165-197.

Anderson, T.W., y Friedman, M. (1960). A limitation of the optimum property of the sequential probability ratio test. En *Contributions to probability and statistics* (essays in Honor of Harold Hotelling) (I. Olkin et al., eds.), No. 6, 57-69. Stanford Univ. Press, Stanford California.

Anscombe, F.J. (1946). Linear sequential rectifying inspection for control fraction defective. *Journal of the Royal Statistical Society, Series B*, 8, 216-222.

----- (1949). Tables of sequential inspection schemes to control fraction defective. *Journal of the Royal Statistical Society, Series A* 112, 180-206.

----- (1949). Large-sample theory of sequential estimation. *Biometrika*, 36, 455-458.

----- (1952). Large-sample theory of sequential estimation. *Proceedings of the Cambridge Philosophical Society*, 48, 600-607.

----- (1954). Fixed-sample-size analysis of sequential observations. *Biometrics*, 10, 89-100.

----- (1963). Sequential medical trials. *J. Amer. Statist. Ass.*, 58, 365-383.

Anscombe, F.J., y Page, E.S. (1954). Sequential tests for binomial and exponential populations. *Biometrika*, 41, 252-253.

Armitage, P. (1950). Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis. *J. Roy. Statist. Soc. Ser. B*, 12, 137-144

----- (1954). Sequential analysis in prophylactic and therapeutic trials. *The Quarterly Journal of Medicine*, 23, 255-274,

Armitage, P. (1958). Numerical studies in the sequential estimation of a binomial parameter. *Biometrika*, 45, 1-15.

----- (1960). *Sequential medical trials*. Blackwell, Oxford.

Arrow, K. J., Blackwell, D., y Girshick, H. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica*, 17, 213-244.

Bahadur, R. R. (1958). A note on the fundamental identity of sequential functions. *Ann. Math. Statist.*, 29, 534-543.

Bangdiwala, I. S. y Monroe, R. J. (1958). Some sequential procedures for ordering populations according to means, variances and regression coefficients. *Institute of Statistics, Mimeo Series No. 202, Raleigh, North Carolina*.

Barnard, G. A. (1946). Sequential tests in industrial statistics (with discussion). *J. Roy. Statist. Soc., Ser. B*, 8, 1-26.

----- (1952). The frequency justification of certain sequential tests. *Biometrika*, 39, 144-150.

Barraclough, Elizabeth D. y Page, E. S. (1959). Tables for Wald tests for the mean of a normal distribution. *Biometrika*, 46, 169-177.

Bartholomew, D. J. (1956) A sequential test for randomness of intervals. *J. Roy. Statist. Soc., Ser. B*, 18, 95-103.

Bartlett, M. S. (1946). The large-sample theory of sequential tests. *Proceedings of the Cambridge Philosophical Society*, 42, 237-244.

Bechhofer, R.E. (1958). A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance and its use with various experimental designs. *Biometrics*, 14, 408-429.

----- (1960). A note on the limiting relative efficiency of the Wald sequential probability test. *J. Amer. Statist. Assoc.*, 55, 660-663.

Bechhofer, R., Kiefer, J., y Sobel, M. (1968). *Sequential Identification and Ranking Problems, with special reference to Koopman-Darmois populations*. *Statis. Res. Monographs Ser.*, Vol. III., Univ. of Chicago Press, Chicago, Illinois.

Berk, R.H. (1969). Strong consistency of certain sequential estimators. *Ann. Math. Statist.*, 40, 1492-1495; ver también corrección: *Ann. Math. Statist.*, 42, 1135-1137 (1971).

Berk, R.H., y Savage, I.R. (1968). The information in a rank-order and stopping time of some associated SPRTS. *Ann. Math. Statist.*, 39, 1661-1674.

Bessler, S.A. (1960). *Theory and Application of Sequential Design of Experiments, k-actions and Infinitely Many Experiments*. Tech. Rep. No. 55 Dept. of Statist., Stanford Univ. Stanford, California.

Bessler, S.A., Chernoff, H., y Marshall, A.W. (1962). An optimal sequential accelerated life test. *Technometrics*, 4, 367-379.

Bhate, D.H. (1955). *Some properties of sequential tests*. Ph.D. thesis, Univ. of London.

Billard, D., y Vagholkar, M.K. (1969). A sequential procedure for testing a null hypothesis against a two-sided alternative hypothesis. *J. Roy. Statist. Soc., Ser. B*, 31, 285-294.

Birnbaum, A., y Healy, W.C., Jr. (1960). Estimates with prescribed variance based on two-stage sampling. *Ann. Math. Statist.*, 31, 662-676.

Billewicz, W.Z. (1956). Matched pairs in sequential trials for significance of a difference between proportion. *Biometrics*, 12, 283-300.

Blackwell, D. (1946). On an equation of Wald. *Ann. Math. Statist.*, 17, 84-87.

----- (1947). A lower bound for the variance of some unbiased sequential estimates. *Ann. Math. Statist.*, 18, 277-280.

Blackwell, D. y Girshick, M.A. (1946). On functions of sequences of independent chance vectors with applications for the problem of the Random Walk in k dimensions. *Ann. Math. Statist.*, 17, 310-317.

Blasbalg, H. (1957). Transformation of the fundamental relationships in sequential analysis. *Ann. Math. Statist.*, 28, 1024-1028.

Blom, G. (1949). A generalization of Wald's fundamental identity. *Ann. Math. Statist.*, 20, 439-444.

Box, G.E.P. (1953). Non-normality and tests on variances. *Biometrika*, 40, 318-335.

Bradt, R.N., Johnson, S.M., y Karlin, S. (1956). On sequential designs for maximizing the sum of n observations. *Ann. Math. Statist.*, 27, 1060-1074.

Bross, I. (1952). Sequential medical plans. *Biometrics*, 8, 188-205.

Burgess, G.C. (1955). Use of sequential analysis for determining test item difficulty level. *Educational and Psychological Measurement*, 15, 80-86.

Burkholder, D.L. y Wijsman, R.A. (1963). Optimum properties and admissibility of sequential tests. *Ann. Math. Statist.*, 34, 98-103.

Burman, J.P. (1946). Sequential sampling formulae for a binomial population. *J. Roy. Statist. Soc., Ser. B*, 8, 98-103.

Burr, I.W. (1953). Fundamental principles of sequential analysis. *Industrial Quality Control*, 9, May, 92-98.

Callahan, J.D. (1969). On Sequential Estimation Procedures for the Mean Vector of a Multi-variate Normal Distribution. Tech. Rep. No. 132. Dept. of Statist., Johns Hopkins Univ., Baltimore, Maryland.

----- (1969). Fixed-size Confidence Regions for Regression Parameters. Tech. Rep. No. 129. Dept. of Statist., Johns Hopkins Univ., Baltimore, Maryland.

Castoli, L. (1956). Identità generalizzata di Wald nella teoria probabilistica delle sequenze casuali. *Bolletino della Unione Matematica Italiana* (Bologna), 11, 22-27.

Chanda, K.C. (1971). Asymptotic distribution of the sample size for a sequential probability ratio test. *J. Amer. Statist. Assoc.*, 66, 178-183.

Chapman, D.G. (1952). Inverse, multiple and sequential sample censuses. *Biometrics*, 8, 286-306.

Chernoff, H. (1956). Sequential design of experiments. *Ann. Math. Statist.*, 30, 755-770.

----- (1960). Sequential tests for the mean of a normal distribution. *Proc. Symp. Math. Statist. Probability*, 4th, Berkeley, 1, 79-92.

----- (1967). Sequential models for clinical trials, *Proc. Symp. Math. Statist. Probability*, 5th, Berkeley, 4, 805-812.

Chernoff, H., y Ray, S.N. (1965). A Bayes sequential sampling inspection - plan. *Ann. Math. Statist.*, 36, 1387-1407.

Chernoff, H., y Savage, I.R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.*, 29, 972-994.

Chung, J.H. (1950). Sequential sampling from finite lots when the proportion defective is small. *J. Amer. Statist. Assoc.*, 45, 557-569.

Close, G.C. (1951). Sequential sampling simplifies receiving inspection. *Modern Machine Shop*, 23, 132.

Cowden, D.J. (1946). An application of sequential sampling to testing students. *J. Amer. Statist. Assoc.*, 41, 547-556.

Cox, D.R. (1949). The use of the range in sequential analysis. *J. Roy. Statist. Soc., Ser. B*, 11, 101-114.

----- (1952). Sequential tests for composite hypotheses. *Proceedings of the Cambridge Philosophical Society*, 48, 290-299.

----- (1952). A note on the sequential estimation of means, *Proceedings of the Cambridge Philosophical Society*, 48, 447-450.

Cox, C.P., y Roseberry, T.D. (1966). A note on the variance of the distribution of sample number in sequential probability ratio tests. *Technometrics*, 8, 700-704.

Cox, D.R. (1963). Large-sample sequential tests for composite hypotheses. *Sankhyā Ser. A.*, 25, 5-12.

Darling, D. A., y Robbins, H. (1967). Inequalities for the sequence of sample means. *Proc. Nat. Acad. Sci. U.S.A.*, 57, 1577-1580.

----- (1968). Some nonparametric sequential tests -- with power one. *Proc. Nat. Acad. Sci. U.S.A.*, 61, 804-809.

Darwin, J. H. (1959). Note on a three-decision test for comparing two binomial populations. *Biometrika*, 46, 106-113.

Davis, M., (1963). Comparison of sequential experiments for estimating -- the dosage-response curve. *Biometrics*, 19, 504.

DeBoor, J. (1953). Sequential test with three possible decisions for testing an unknown probability. *Applied Scientific Research B*, 3, 249-259.

DeGroot, M. H. (1959). Unbiased sequential estimation for binomial populations. *Ann. Math. Statist.*, 30, 80-101.

DeGroot, M. H. y Nadler, J. (1958). Some aspects of use of the sequential -- probability ratio test. *J. Amer. Statist. Assoc.*, 53, 187-199.

Divatia, V. V. (1950). A note on sequential exponential tests. *J. Indian -- Soc. of Agricultural Statist.*, 30, 601-606.

Donnelly, T. G. (1957). A Family of Truncated Sequential Tests. Doctoral -- Dissertation. Univ. of North Carolina, Chapel Hill.

Dresselaers, C. y Gillis, P. P. (1951). Un test séquential unilatéral, *Acadé -- mic Royale de Belgique Bulletin de la Classe des Sciences*, 37, 713-727.

Dvoretzky, A., Kiefer, J., y Wolfowitz, J. (1953). Sequential decision pro -- blems for processes with continuous time parameter. Testing hypotheses. *Ann. Math. Statist.*, 24, 254-272.

Dvoretzky, A., Kiefer, J., y Wolfowitz, J. (1953). Sequential decision problems for processes with continuous time parameter. *Problems of estimation*. *Ann. Math. Statist.*, 24, 403-415.

----- (1959). Corrections to sequential decision problems for processes with continuous time parameter. *Testing hypotheses*. *Ann. Math. Statist.*, 30, 1265.

Eicker, F. (1963). Central limit theorems for families of sequences of random variables. *Ann. Math. Statist.*, 34, 447-457.

Epstein, B. (1955). A sequential two-sample life test. *J. Franklin Institute*, 260, 25-29.

Epstein, B., y Sobel, M. (1955). Sequential life tests in the exponential case. *Ann. Math. Statist.*, 26, 82-93.

Fabian, V. (1974). Note on Anderson's sequential procedures with triangular boundary. *Ann. Statist.*, 2, 170-176.

Farrell, R.H. (1962). Bounded length confidence intervals for the zero of a regression function. *Ann. Math. Statist.*, 33, 237-247.

----- (1964). Asymptotic behavior of expected sample size in certain one-sided tests. *Ann. Math. Statist.*, 35, 36-72.

Fisher, R.A. (1952). Sequential experimentation. *Biometrics*, 8, 183-187.

Fiske, D.W. y Jones, L.V. (1954). Sequential analysis in psychological research. *Psychological Bulletin*, 51, 264-275.

Fraser, D.A.S. (1951). Sequentially determined statistically equivalent blocks. *Ann. Math. Statist.*, 22, 372-381.

Freeman, H.A. (1944). *Sequential analysis of statistical data; applications*. Statistical Research Group, Columbia University, New York.

Garner, N.R. (1956). The operating characteristic curve for sequential sampling by variables when the producer's and consumer's risks are equal. *J. Amer. Statist. Assoc.*, 51, 108-110.

Geertsema, J.C. (1970). Sequential confidence intervals based on rank tests. *Ann. Math. Statist.*, 41, 1016-1026.

----- (1970). A Monte Carlo study of sequential confidence intervals based on rank tests. *South African Statist.*, 4.

Ghosh, B.K. (1969). Moments of the distribution of sample size in a SPRT. *J. Amer. Statist. Assoc.*, 64, 1560-1574.

----- (1970). *Sequential Tests of Statistical Hypotheses*. Addison-Wesley, Reading, Massachusetts.

Ghosh, J.K. (1960). On some properties of sequential t -test. *Calcutta Statist. Assoc. Bull.*, 9, 77-86.

----- (1960). On the monotonicity of the OC of a class of sequential probability ratio test. *Calcutta Statist. Assoc. Bull.*, 9, 139-144.

Ghosh, M.N. (1960). Bounds for the expected sample size in a sequential probability ratio test. *J. Roy. Statist. Soc., Ser. B*, 22, 360-367.

Ghurye, S.G., y Robbins, H. (1954). Two-stage procedures for estimating the difference between means. *Biometrika*, 41, 146-152.

Gilchrist, W.G., (1961). Some sequential tests using range. *J. Roy. Statist. Soc., Ser. B*, 23, 335.

Gleser, L.J. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.*, 37, 1053-1055.

----- (1966). Correction note. *Ann. Math. Statist.*, 37, 1053-1055.

Gleser, L.J., Robbins, H., y Starr, N. (1964). Some Asymptotic Properties of Fixed-width Sequential Confidence Intervals for the Mean of a Normal Population with Unknown Variance. Tech. Rep. Dept. of Math. Statist., Columbia Univ., New York.

Govindarajulu, Z. (1968). Asymptotic normality of two-sample rank order sequential probability ratio test based on Lehmann alternatives. No publicado.

----- (1969). Stopping time for a c-sample rank order SPRT. No publicado.

Groenveld, R.A. (1971). A note on the sequential sign test. *Amer. Statist.*, 25, 2, 15-16.

Gulati, R.L. (1956). Sequentielle Tests für den Korrelations Koeffizienten. *Mitteilungsblatt für Mathematische Statistik*, 8, 202-233.

Hajnal, J. (1960). Sequential trials of analgesics in rheumatoid arthritis. *Quantitative methods in Pharmacology*, North-Holland, Amsterdam.

Hajnal, J., (1961). A two-sample sequential t -test. *Biometrika*, 48, 65-75.

Hajnal, J., Sharp, J., y Popert, A.J., (1959). A method for testing analgesics in rheumatoid arthritis using a sequential procedure. *Ann. Rheum. Dis.*, 18. 189-206.

Hall, W.J. (1959). On sufficiency and invariance with applications to sequential analysis. Institute of Statistics, Memo Series, No. 228, Chapel Hill, North Carolina.

Hall, W.J. (1962). Some sequential analogs of Stein's two-stage test. *Biometrika*, 49, 367-378.

----- (1965). Methods of Sequentially Testing Composite Hypotheses - with special Reference to the Two-Sample Problem. Mimeo, series, No. - 441, 40- 41.

Hall, W.J., Wijsman, R.A., y Ghosh, J.K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Statist.*, 36, 575-614.

Hamaker, H.C. (1953). The efficiency of sequential sampling for attributes, Parts I and II. *Philips Research Reports*, 8, 35-46 and 427-433.

Harris, T.E., (1947). Note on differentiation under the expectation sign in the fundamental identity of sequential analysis. *Ann. Math. Statist.* 18, 294-295.

Herbach, L.H. (1948). Bounds for some functions used in sequentially testing the mean of a Poisson distribution. *Ann. Math. Statist.*, 19, 400-405.

Hoeffding, W. (1953). A lower bound for the average sample number for a sequential test. *Ann. Math. Statist.*, 24, 127-130.

------(1960). Lower bounds for the expected sample size and the average risk of a sequential procedure. *Ann. Math. Statist.*, 31, 352-368.

Hoel, D.G. (1970). On the monotonicity of the OC of a SPRT. *Ann. Math. Statist.*, 41, 310-314.

Hoel, P.G. (1954). On a property of the sequential t -test. *Skandinavisk Aktuarietidskrift*, 37, 19-22.

------(1955). On a sequential test for the general linear hypothesis. *Ann. Math. Statist.*, 26, 136-139.

Ifram, A.F. (1965). On the asymptotic behavior of densities with applications to sequential analysis. *Ann. Math. Statist.*, 36, 615-637.

Jackson, J.E. y Bradley, R.A. (1959). Multivariate sequential procedures for testing means. *The Development of Statistical Methods for Experimental Designs in Quality Control and Surveillance Testing*, Technical Report No. 10, Virginia Agricultural Experiment Station, Blacksburg, Virginia.

------(1961). Sequential χ^2 and T^2 tests. *Ann. Math. Statist.*, 32, 1063-1077.

------(1961). Sequential χ^2 and T^2 tests and their application to an acceptance sampling problem. *Technometrics*, 3, 519-534.

Johson, N.K. (1953). Some notes on the application of sequential methods in the analysis of variance. *Ann. Math. Statist.*, 24, 614-623.

Johnson, N.L. (1953). Sequential procedures in certain component of variance problems. *Ann. Math. Statist.*, 25, 357-366.

----- (1957). Sequentially determined confidence intervals. *Biometrika*, 44, 279-281.

----- (1959). A proof of Wald's theorem on cumulative sums. *Ann. Math. Statist.*, 30, 1245-1247.

_____ (1960). On the choice of a sequential procedure. *Proc. Biometric Soc. Symp.*, 27-40.

----- (1961). Sequential analysis: a survey. *J. Roy. Statist. Soc.*, Ser. A, 124, 372-411.

Johnson, N.L. y Moore, P.G. (1957). Applications of sequential methods to mortality data. *J. Inst. Actuaries Students Soc.*, 14, 84-93.

Jones, H.L. (1952). Formulas for the group sequential sampling of attributes. *Ann. Math. Statist.*, 23, 72-87.

Kaysor, K.H. (1956). Statistical sequential inspection procedure for the control of the defect percentage. *Rationalisierung*, 1, 114-115.

Kemp, K.W. (1958). Formulae for calculating the operating characteristic and the average sample number of some sequential tests. *J. Roy. Statist. Soc.*, Ser. B, 20, 379-386.

Kemperman, J.H.B. (1949). Some methods from sequential analysis, *Math. Centrum, Amsterdam, Rapport ZW-1949-009*.

----- (1950). Some methods from sequential analysis II. *Math. Centrum, Amsterdam, Rapport ZW-1950-003*.

Kerperman, J.H.B. (1950). The general one-dimensional random walk with absorbing barriers with applications to sequential analysis. Excelsiors Foto-Offset.

Khan, R.A. (1968). Sequential estimation of the mean vector of a multivariate normal distribution. *Sankhyā Ser. A*, 30, 331-334.

----- (1969). A general method of determining fixed-width confidence intervals. *Ann. Math. Statist.*, 40, 704-709.

Kiefer, J. (1952). Sequential minimax estimation for the rectangular distribution with unknown range. *Ann. Math. Statist.*, 23, 586-593.

----- (1953). Sequential minimax search for a maximum. *Proceedings of the American Math. Soc.*, 4, 502-506.

----- (1957). Optimum sequential search and approximation methods under minimum regularity assumptions. *J. Soc. Industrial and Applied Math.* 5, 105-136.

----- (1957). Invariance, minimax sequential estimation and continuous time processes. *Ann. Math. Statist.*, 28, 573-601.

Kiefer, J., y Sacks, J. (1963). Asymptotically optimum sequential inference and design. *Ann. Math. Statist.*, 34, 705-750.

Kiefer, J., y Weiss, L. (1957). Some properties of generalized sequential probability ratio tests. *Ann. Math. Statist.*, 28, 57-74.

Kiefer, J., y Wolfowitz, J. (1952). Sequential tests of hypotheses about mean occurrence time of a continuous parameter Poisson process. *Naval Research Logistics Quarterly*, 3, 205-219.

Kilpatrick, G.S., y Oldham, P.D., (1954). Calcium chloride and adrenaline - as bronchial dilators compared by sequential analysis. *Brit. Med. J.*, II, 1388-1391.

Kimball, A.W. (1950). Sequential sampling plans for use in psychological test work. *Psychometrika*, 15, 1-15.

King, E.P. (1964). Optimal replication in sequential drug screening. *Biometrika*, 51, 1-10.

Knight, W.A. (1965). A method of sequential estimation applicable to the hypergeometric, binomial, Poisson, and exponential distributions. *Ann. Math. Statist.*, 36, 1494-1503.

Lasman, L.L., y Williams, E.J. (1958). Asymptotic distribution of sample size for certain methods of sequential sampling. *Inst. Statist. Technical Report No. 21*, Raleigh, North Carolina.

Lechner, J.A. (1964). Optimality and the OC curve for the Wald SPRT. *J. Amer. Statist. Assoc.*, 59, 464-468.

Leimbacher, W.R. (1953). On some classes of sequential procedures for obtaining confidence intervals of given length. *Univ. California Publications in Statist.*, 2, 1-21.

Lieberman, A. (1959). Sequential life testing plans for the exponential distribution. *Industrial Quality Control*, 16, 14-18.

Lieberman, A., y Kulin, Mary R. (1957). Sequential life test plans (exponential distribution). *Bureau of Ships Statistical Engineering Report No. 373-15(63)*. U.S. Navy, Washington, D.C.

Lindley, D.V., y Barnett, B.N. (1965). Sequential sampling, two decision -- problems with linear losses for binomial and normal random variables. *Biometrika*, 52, 507- 532.

Lombardi, G.J. (1951). The sequential selection of judges for organoleptic testing. *Statistical Methods for Sensory Difference Tests of Food Quality No. 2*, Virginia Agricultural Experiment Station, Blacksburg, Vir.

Lorden, G. (1967). Integrated risk of asymptotically Bayes sequential --- tests. *Ann. Math. Statist.*, 37, 1399-1422.

----- (1972). Likelihood ratio tests for sequential k-decision problems. *Ann. Math. Statist.*, 43, 1412-1427.

MacCabe, G.P., Jr. (1970). Sequential Estimation of a Poisson Integer Mean. Mimeo. Ser. No. 232. Dept. of Statist., Purdue Univ., Lafayette, Indiana.

Maguire, C.A. (1953). Sequential decisions involving the choice of experiments. *Applied Math. Statist. Laboratory Technical Report No. 19*, Stanford University.

Mallik, A.K. (1971). Sequential Estimation of the Common Mean of Two Normal Populations. Tech. Rep. No. 42 Dept. of Statist., Stanford Univ., Stanford California.

Mallows, C.K. (1953). Sequential discrimination. *Sankhyā*, 12, 321-338.

Matthes, T.K. (1963). On the optimality of sequential probability ratio - tests. *Ann. Math. Statist.*, 34, 18-21.

Maurice, Rita J. (1957). A minimax procedure for choosing between two populations using sequential sampling. *J. Roy. Statist. Soc., Ser. B*, 19, 255-261.

Medhi, J. (1956). A note on the risks of error involved in the sequential ratio test. *Biometrika*, 43, 231-234.

Mikhalevick, V.S. (1956). Sequential Bayes solutions and optimal methods of statistical acceptance control. *Theor. Probability Appl.*, 1, 395-421.

Miller, R.G. (1961). A generalisation of Wald's identity with applications to random walks. *Ann. Math. Statist.*, 32, 549-560.

----- (1970). Sequential signed rank test. *J. Amer. Statist. Assoc.*, 65, 1554-1561.

Moorman, W.J. (1950). Some empirical aspects of the sequential analysis technique as applied to an achievement examination. *J. Experimental Education*, 18, 195-207.

Moore, P.G. (1953). A sequential test of randomness. *Biometrika*, 40, 111-115.

Morgan, M.E., Macleod, P., Anderson, E.O., & Bliss, C.I. (1951). A sequential procedure for grading milk by microscopic counts. *Conn. Agric. Exp. Sta. Bull.*, 276-331.

Moshman, J. (1949). The sequential analysis of blood pressure scores. *Industrial Medicine and Surgery*, 18, 344-346.

----- (1958). A method for selecting the size of the initial sample in Stein's two-sample procedure. *Ann. Math. Statist.*, 29, 1271-1275.

Nádas, A. (1969). An extension of a theorem of Chow and Robbins on sequential confidence intervals for the mean. *Ann. Math. Statist.*, 40, 667-671.

Nandi, H.K. (1948). Use of well-known statistics in sequential analysis. *Sankhyā*, 8, 339-344.

Narayana, T.V. (1953). Sequential procedures in probit analysis. Institute of Statist. Mimeo Series No. 82, Chapel Hill, North Carolina.

Noether, G. (1954). A sequential test of randomness against linear trend. *Ann. Math. Statist.*, 25, 176.

----- (1956). Two sequential tests against trend. *J. Amer. Statist. Assoc.*, 51, 440-450.

Oakland, G.B. (1950). An application of sequential analysis to whitefish sampling. *Biometrics*, 6, 59-67.

O'Neill, R.T., y Rohatgi, V.K. (1973). A two-stage procedure for estimating the difference between the mean vectors of two multi-variate normal populations. *Trabajos Estadísticos*, 24, 123-130.

Page, E.S. (1954). An improvement to Wald's approximation for some properties of sequential tests. *J. Roy. Statist. Soc.*, --- Ser. B, 16, 136-139.

Parent, Jr., E.A. (1965). Sequential Ranking Procedures. Tech. Rep. No. 80. Dept. of Statist. Stanford Univ., Stanford, Calif.

Paulson, E. (1947). A note on the efficiency of the Wald sequential test. *Ann. Math. Statist.*, 18, 447-450.

----- (1953). A sequential procedure for choosing one of k hypotheses concerning the unknown mean of a normal distribution. *Ann. Math. Statist.*, 34, 549-554.

Paulson, E. (1964). A sequential procedure for selecting the population with the largest mean from normal populations. *Ann. Math. Statist.*, 35, 174-180.

----- (1964). Sequential estimation and closed sequential decision procedures. *Ann. Math. Statist.*, 35, 1048-1058.

----- (1969). Sequential interval estimation for the means of normal populations. *Ann. Math. Statist.*, 40, 509-516.

Perng, S. K. (1969). A comparison of the asymptotic expected sample sizes of two sequential procedures for ranking problems. *Ann. Math. Statist.*, 40, 2198-2202.

Phatarfod, R. M. (1965). Sequential analysis of dependent observations I. *Biometrika* 52, 157-165.

Pólya, G. (1948). Exact formulas in the sequential analysis of attributes. *University of California Publications in Mathematics, New Series*, 1, 229-239.

Procassini, A. A. y Clark, R. L. (1957). Argus sequential sampling. *Industrial Quality Control*, 9, 15-16.

Radkins, A. P. (1958). Sequential analysis in organoleptic research. *Food Research*, 23, 225-234.

Rao, C. R. (1951). The discriminant function approach in the classification of time series. *Sankhyā*, 11, 257-272.

Rao, M. M. (1958). Note on a remark of Wald. *Amer. Math. Monthly*, 65, 277-278.

Ray, W.D. (1956). Sequential analysis applied to certain experimental designs in the analysis of variance. *Biometrika*, 43, 388-403.

----- (1957). A proof that the sequential probability ratio test of the general linear hypothesis terminates with probability unity. *Ann. Math. Statist.*, 28, 521-523.

Rivett, B.H. (1951). Sequential analysis of machine performance. *Statist. Method in Industrial Production*, Roy. Statist. 86-89.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the Amer. Math. Soc.*, 58, 527-535.

----- (1956). Sequential decision problem with a finite memory. *Proceedings of the Nat. Acad. Sc., U.S.A.*, 42, 920-923.

----- (1959). Sequential estimation of the mean of a normal population. *Probability and Statist.* 235-245.

Robbins, H., y Starr, N. (1965). Remarks on Sequential Hypothesis Testing. *Tech. Rep.*, 68, Univ. of Minnesota, Minneapolis.

Robbins, H., Sobel, M., y Starr, N. (1968). A sequential procedure for selecting the largest of means. *Ann. Math. Statist.*, 39, 88-92.

Robinson, Julia (1948). A note on exact sequential analysis. *University of California Publications in Math.*, 1, 241-246.

Rohatgi, V.K., y O'Neill, R.T. (1973). On the sequential estimation of the mean vector of a multinormal population. *Trabajos Estadist.*, 24, part 3.

Rohatgi, V. K., y Rastogi, S. C. (1973). On sequential estimation of a certain estimable function of the mean vector of a multivariate normal distribution. *J. Austral. Math. Soc.*, 15, 291-295.

Romaní, J. (1956). Non-parametric tests in sequential form. *Tra bajos de Estadística*, 7, 43-96.

Roseberry, T. D., y Gehan, E. A. (1964). Operating characteristic curves and accept-reject rules for two and three stage screening procedures. *Biometrics*, 20, 73-84.

Rushton, S. (1950). On a sequential t -test. *Biometrika*, 37, 326-333.

----- (1952). On sequential tests of the equality of variances of two normal populations with known means. *Sankhyā* 12, 63-78.

----- (1952). On a two-sided sequential t -test. *Biometrika* 39, 302-308.

Sacks, J. (1965). A note on the sequential t -test. *Ann. Math. Statist.*, 36, 1867-1869.

Sainsbury, P., y Lucas, C. J. (1959). Sequential methods applied to the study of prochlorperazine. *Brit. Med. J.*, 2, 737-740.

Sakaguchi, M. (1961). Dynamic programming of some sequential sampling designs. *J. Math. Anal. Appl.*, 2, 446-466.

----- (1967). Relative efficiency of the Wald SPRT and the Chernoff information number. *Kōdai Math. Sem. Rep.*, 19, 138-145

Samuel, E. (1966). Estimators with prescribed bound on the variance for the parameters in the binomial and Poisson distributions, based on two-stage sampling. *J. Amer. Statist. Assoc.*, 61, 220-227.

Samuelson, P. A. (1948). Exact distribution of continuous variables in sequential analysis. *Econometrica*, 16, 191-198.

Savage, L. J. (1947). A uniqueness theorem for unbiased sequential binomial estimation. *Ann. Math. Statist.*, 18, 295-297.

Savage, I. R., y Sethuraman, J. (1966). Stopping time of a rank on der sequential probability ratio test based on Lehman alternatives. *Ann. Math. Statist.*, 37, 1154-1160.

Schmetterer, L. (1954). Zum Sequential-verfahren von Robbins -- und Monro. *Monatshefte für Mathematik*, 58, 33-37.

Schmid, J., Jr. (1952). Sequential analysis of test items. *J. Experimental Education*, 20, 261-264.

Schneiderman, M. A., y Armitage, P., (1962). A family of closed sequential procedures. *Biometrika*, 49, 41-56.

----- (1962). Closed sequential tests. *Biometrika*, 49, 359-366.

Schwarz, G. (1960). Asymptotic stopping regions for sequential testing with an indifference region. *Appl. Math. Statist. Lab.*, Technical Report No. 42. Stanford University.

----- (1962). Asymptotic shapes of Bayes sequential testings regions. *Ann. Math. Statist.*, 33, 224-236.

Seelbinder, B.M. (1953). On Stein's two-stage procedure. *Ann. Math. Statist.*, 24, 640-649.

Sen, P.K., y Ghosh, M. (1971). On bounded length sequential confidence intervals based on one-sample rank-order statistics. *Ann. Math. Statist.*, 42, 189-203.

Sethuraman, J. (1970). Stopping time of a rank order sequential probability ratio test based on Lehmann alternatives. *Ann. Math. Statist.*, 41, 1322-1333. ver corrección: *Ann. Math. Statist.*, 41, 1376-1380 (1970).

Simons, G. (1967). Lower bounds for average sample number of sequential multi-hypothesis tests. *Ann. Math. Statist.*, 38, 1343 - 1364.

----- (1967). A sequential three hypothesis test for determining the mean of a normal population with known variance. *Ann. Math. Statist.*, 38, 1365-1375.

----- (1967). A class of sequential procedures for choosing one of k hypotheses concerning the unknown drift parameters of the Wiener process. *Ann. Math. Statist.*, 38, 1376-1383.

----- (1968). On the cost of not knowing the variance when making a fixed-width confidence interval for the mean. *Ann. Math. Statist.*, 39, 1946-1952.

Simons, G., y Zacks, S. (1967). A Sequential Estimation of Tail Probabilities in Exponential Distribution with a Prescribed Proportional Closeness. Tech. Rep. No. 21. Dept. of Statist., Stanford Univ., Stanford, California.

Siskind, V. (1964). On certain formulae applied to the sequential t -test. *Biometrika*, 51, 97-106.

Snell, E. S., y Armitage, P. (1957). Clinical comparison of diamorphine and pholcodine as cough suppressants by a new method of sequential analysis. *The Lancet*, 860-862.

Sobel, M. (1953). An essentially complete class of decision --- functions for certain standard sequential problems. *Ann. Math. Statist.*, 24, 319-337.

----- (1956). Sequential procedures for selecting the best exponential population. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, Vol. 5, 99-110.

Sproule, R. N. (1969). A sequential Fixed-Width Confidence Interval for the Mean of a U -statistic. Doctoral Dissertation. Univ of North Carolina.

Srivastava, M. S. (1966). Some asymptotically efficient sequential procedures for ranking and slippage problems. *J. Roy. Soc. Ser. B.*, 28, 370-380.

----- (1967). The performance of a sequential procedure for a slippage problem. No publicado.

----- (1967). On fixed-width confidence bounds for regression parameters. *Ann. Math. Statist.*, 42, 1403-1411.

Srivastava, M. S., y Ogilvie, J. (1968). The performance of some sequential procedures for a ranking problem. *Ann. Math. Statist.* 39, 1040-1047.

Starr, N. (1966). On the asymptotic efficiency of a sequential procedure for estimating the mean. *Ann. Math. Statist.*, 37, --- 1173-1185.

Starr, N., y Woodroffe, M. B. (1969). Remarks on sequential point estimation. *Proc. Nat. Acad. Sci. U. S. A.*; 63, 285-288.

----- (1971). Further Remarks on Sequential Estimation: The Exponential Case. Tech. Rep. No. 7. Dept. of Statist., Univ. of Michigan, Ann Arbor.

Statistical Research Group, Columbia University (1945). *Sequential analysis of statistical data: applications*. Columbia University Press, New York.

Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.*, 16, 243-258.

Stockman, C. M. (1944). A method of obtaining an approximation for the operating characteristic of a Wald sequential probability test applied to a binomial distribution. Ministry of Supply Advisory Service on Statistical Method and Quality Control, Technical Report Q. C. /R/19.

Stockman, C. M., y Armitage, P. (1946). Some properties of closed sequential schemes. *J. Roy. Statist. Soc., Ser. B*, 8, 104-112.

Storch, J. H. (1953). *Sequential sampling in the inspection of goods*. *Statistica*, Rijswijk, 7, 3 - 14

Taguchi, G. (1957). How to decide the group size in a group sequential sampling inspection plan. *Statistical Quality Control*, 8, 20-21.

Tallis, G.M., y Vagholkar, M.K. (1965). Formulas to improve Wald's approximation for some properties of sequential tests. *J. Roy. Statist. Soc., Ser. B*, 27, 74-81.

Taylor, J. (1956). Exact linear sequential tests for the mean of a normal distribution. *Biometrika*, 43, 452-455.

Tomlinso, R.C. (1957). A simple sequential procedure to test whether average conditions achieve a certain standard. *Appl. Statist.*, 6, 198-207.

Tsao, C.K. (1954) A simple sequential procedure for testing statistical hypotheses. *Ann. Math. Statist.*, 25, 687-702.

Tweedie, M.C.K. (1952). The estimation of parameters from sequentially sampled data on a discrete distribution. *J. Roy. Statist. Soc., Ser. B*, 14, 238-245.

Vagholkar, M.K., y Wetherill, G.B. (1960). The most economical binomial sequential probability ratio test. *Biometrika* 47, 103-109.

VanEeden, Constance (1955), A sequential test with three possible decisions for comparing two unknown probabilities, based on groups of observations. *Revue de l'Institut International de Statistique*, 23, 20-28.

Ville, J., y Schützenberger, M.P. (1951), Les problèmes de diagnostic séquentiel. *Comptes Rendus*, 232, 206-207.

Wald, A. (1943). *Sequential analysis of statistical data*. Theory. Statistical Research Group, Columbia University,

Wald, A. (1945). Sequential tests of statistical hypothesis. *Ann. Math. Statist.*, 16, 117-186.

----- (1945). Sequential method of sampling for deciding between two courses of action. *J. Amer. Statist. Assoc.*, 40, -- 277-306.

----- (1946). Some improvements in setting limits for the expected number of observations required by a sequential probability ratio test. *Ann. Math. Statist.*, 17, 466-474.

----- (1946). Differentiation under the expectation sign in the fundamental identity of sequential analysis. *Ann. Math. Statist.*, 17, 493-497.

----- (1947). Sequential analysis (with discussion). *Proceedings of the International Statistical Conferences*, 3, 67-80.

----- (1947). Foundations of a general theory of sequential decision functions. *Econometrica*, 15, 279-313.

----- (1950). Asymptotic minimax solutions of sequential point estimation problems. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, 1 - 11.

Wald, A., y Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.*, 19, 326-329.

----- (1949). Bayes solutions of sequential decision problems. *Proceedings of the National Academy of Sciences, U.S.A.*, 35, 99-102.

Wald, A., y Wolfowitz, J. (1950). Bayes solutions of sequential decision problems. *Ann. Math. Statist.*, 21, 82-99.

Walker, A. M. (1950). Note on sequential sampling formulae for a binomial population. *J. Roy. Statist. Soc., Ser. B*, 12, 301-307.

Walker, H. M. (1949). Item selection by sequential sampling. -- *Teachers College Record*, 50, 404-409.

Wasan, M. T. (1964). Sequential optimum procedures for unbiased estimation of a binomial parameter. *Technometrics*, 6, 259-272.

Waters, W. E. (1955). Sequential sampling in forest insect surveys. *Forest Science*, 1, 68-79.

Weed, J. H. D., Jr. (1968). Sequential One-sample Grouped Rank tests for Symmetry. Ph.D. Dissertation, Florida State University Library, Tallahassee, Florida.

Weed, Jr., H. D., y Bradley, R. A. (1971). Sequential one-sample grouped signed rank tests for symmetry. *J. Amer. Statist. Assoc.* 66, 321-326.

Weed, Jr., H. D., Bradley, R. A., y Govindarajulu, Z. (1974). Stopping times of two rank order sequential probability ratio tests for symmetry based on Lehmann alternatives. *Ann. Statist.*, 2, 1314-1322.

Weiss, L. (1954). Sequential procedures that control the individual probabilities of coming to the various decisions. *Ann. Math. Statist.*, 25, 779-784.

----- (1956). On the uniqueness of Wald's sequential tests. *Ann. Math. Statist.*, 27, 1178-1181.

Wetherill, G. B. (1959). The most economical sequential sampling scheme for inspection by variables. *J. Roy. Statist. Soc., Ser. B*, 21, 400-408.

----- (1961). Bayesian sequential analysis. *Biometrika*, 48, 281-292.

----- (1961). Sequential estimation of quantal response curves. *J. Roy. Statist. Soc., Ser. B*, 25, 1-48.

Wetherill, G. B., y Chen, H. (1965). Sequential estimation of quantal response curves II. A new method of estimation. Tech. Memo 65-1215-1, Bell Tel. Labs., N.J.

Whittle, P., y Lane, R. O. D. (1967). A class of situations in which a sequential estimation procedure is nonsequential. *Biometrika*, 54, 229-234.

Wijsman, R. A. (1958). On the existence of Wald's sequential test. *Ann. Math. Statist.*, 29, 938-939.

----- (1960). A monotonicity property of the sequential probability ratio test. *Ann. Math. Statist.*, 31, 677-684.

----- (1963). Existence, uniqueness and monotonicity of sequential probability ratio tests. *Ann. Math. Statist.*, 34, 1541-1548.

----- (1967). General proof of termination with probability one of invariance sequential probability ratio tests based on multivariate normal observations. *Ann. Math. Statist.*, 38, 8-24.

Wijsman, R.A. (1968). Bounds on the sample size distribution - for a class of invariant sequential probability ratio tests. *Ann. Math. Statist.*, 39, 1048-1056.

----- (1971). Exponentially bounded stopping time of -- invariant sequential probability ratio tests. *Proc. Symp. on Decision Theory and Related Topics*, Purdue Univ. Lafayette, - Indiana, 217-224. Academic Press, New York.

----- (1972). Examples of exponentially bounded stop-- ping time of invariant sequential probability ratio tests -- when the model may be false. *Proc. Symp. Math. Statist. Probability*, 6th, Berkeley, 1, 19-128. Univ. of California Press, Berkeley.

Wilcoxon, F., Rhodes, L.J., y Bradley, R.A. (1963). Two sequential two-sample grouped rank tests with applications to screening experiments. *Biometrics*, 19, 58-84.

Wirjosudirdjo, S. (1961). Limiting Behavior of a Sequence of -- Density Ratios. Doctoral Dissertation. Univ. of Illinois, Urbana; ver también *Ann. Math. Statist.*, 33, 296-297.

Woinsky, M.N., y Kurtz, L. (1969). Sequential nonparametric two-way classification with prescribed maximum asymptotic error probability. *Ann. Math. Statist.*, 40, 445-455.

Wolfowitz, J. (1946). On sequential binomial estimation. *Ann. - Math. Statist.*, 17, 489-493.

----- (1947). Consistency of sequential binomial esti- mates. *Ann. Math. Statist.*, 18, 131-135.

Wolfowitz, J. (1947). The efficiency of sequential estimates -- and Wald's equation for sequential processes. *Ann. Math. Statistics.*, 18, 215-230.

----- (1950). Minimax estimates of the mean of the normal distribution with known variance. *Ann. Math. Statist.*, 21, 218-230.

Wünsche, G. (1953). Sequential Testverfahren in der Versicherungstechnik. *Bl. Deutsch. Ges. Versicherungsmath.*, 1, no. 4, 19-37.

Wong, S. P. (1968). Asymptotically optimum properties of certain sequential tests. *Ann. Math. Statist.*, 39, 1244-1263.

Woodroffe, M. (1971). The Bounded Modification of the SPRT. Tech. Rep. No. 8, Dept. of Statist. Univ. of Michigan, Ann Arbor.

Yahav, J. A., y Bickel, P. J. (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. *Ann. Math. Statist.*, 39, 442-456.

Yen, E. H. (1964). On two-stage nonparametric estimation. *Ann. Math. Statist.*, 35, 1099-1114.

CONCLUSIONES.

A pesar de que el material presentado en este trabajo está constituido esencialmente por resultados teóricos, es importante mencionar que en el Análisis Secuencial se han encontrado aplicaciones a varios campos, como son la Industria, la Medicina, la Demografía, la Psicología, etc.

En la Industria principalmente en Control de Calidad; en la Medicina en ensayos profilácticos y terapéuticos, pruebas organolépticas, así como en estudios del reumatismo; en Demografía en la parte de mortalidad se ha utilizado para construir tablas de mortalidad; finalmente en el Análisis Probit se ha utilizado de igual manera.

Es importante hacer notar la gran similitud en la forma entre los resultados de Estadística Clásica y los resultados que se han obtenido desde el punto de vista secuencial, tales como propiedades de los estimadores, Cota Inferior de la Varianza, intervalos de confianza, --- etc.

Finalmente, dada la información bibliográfica del último capítulo, así como los problemas que se han dejado abiertos en este trabajo, se puede asegurar que hay material suficiente para desarrollar más a fondo este tema, así como hacer trabajos de investigación sobre el mismo, que puede incluir un Seminario con aquellas personas interesadas.

APENDICE A

Sea X_1, \dots, X_n una muestra aleatoria de una población con función de densidad $f(x|\theta)$; con θ en Θ (espacio parametral). --
 Sea $T = T(X_1, \dots, X_n)$ un estimador insesgado de $t(\theta)$ alguna función de θ . Para poder construir la cota inferior de la varianza para la clase de estimadores insesgados de $t(\theta)$, es necesario que se cumplan las siguientes condiciones, que usualmente reciben el nombre de condiciones de regularidad:

- i) $\frac{\partial}{\partial \theta} \ln f(x|\theta)$ existe para toda x y para toda θ
- ii)
$$\frac{\partial}{\partial \theta} \int \dots \int \prod_{i=1}^n f(x_i|\theta) dx_1 \dots dx_n$$

$$= \int \dots \int \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i|\theta) dx_1 \dots dx_n$$
- iii)
$$\frac{\partial}{\partial \theta} \int \dots \int t(x_1, \dots, x_n) \prod_{i=1}^n f(x_i|\theta) dx_1 \dots dx_n$$

$$= \int \dots \int t(x_1, \dots, x_n) \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i|\theta) dx_1 \dots dx_n$$
- iv) $0 < E \left[\left(\frac{\partial}{\partial \theta} \ln f(x|\theta) \right)^2 \right] < \infty$ para todo θ en Θ

BIBLIOGRAFIA.

LIBROS.

- Ghos, B.K. (1970). *Sequential Test of Statistical Hypotheses*. Addison-Wesley, Reading, Massachusetts.
- Govindarajulu, Z. (1975). *Sequential Statistical Procedures*, Academic Press Inc. London.
- Hogg, R. y Craig, A. *Introduction to Mathematical Statistics*. Third Edition. Macmillan Publishing Co., Inc. New York. 1970.
- Johnson Norman L. & Kotz Samuel. *Continuous univariate distributions- 2*. The Houghton Mifflin Series in Statistics.
- Kendall, S.M. y Stuart, A. *The advanced theory of statistics, Vol II*. Fourth Edition. Charles Griffin & Co. limited London, 1979.
- Mendoza, M. *Importancia del análisis estadístico secuencial*, Comunicación Interna no. 13, Depto. de Matemáticas. Facultad de Ciencias. 1978.
- Mood, A., Graybill, F., Boes, D. *Introduction to the theory of statistics*. Third Edition. McGraw-Hill Kogakusha, Ltd. 1974.
- Wald, Abraham. *Sequential Analysis*. Dover Publications, Inc. New York (1947).
- Wetherill, G. *Sequential Methods in Statistics*. John Wiley & Sons Inc. (1966).

ARTICULOS.

Anscombe, F.J. (1949). Large sample theory of sequential estimation. *Biometrika*, 36, 455-458.

----- (1953). Sequential estimation. *J. Roy. Statist. Soc., Ser. B.* 15, 1-29.

Berk, R.H. (1969). Strong consistency of certain sequential estimators. *Ann. Math. Statist.*, 40, 1492-1495. Ver también la corrección: *Ann. Math. Statist.*, 42, 1135-1137.

Cox, D.R. (1952). Estimation by Double Sampling. *Biometrika*, 39, 217-227.

Chapman, D.G. (1954). The estimation of biological populations. *Ann. Math. Statist.*, 25, 1-15.

Chow, Y.S., y Robbins, H. (1965). On the Asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.*, 36, 457-462.

Darling, D.A., y Robbins, H. (1967). Finding the size of a finite population. *Ann. Math. Statist.*, 38, 1392-1398.

Darroch, J.N. (1958). The multiple recapture census. Estimation of a closed population. *Biometrika*, 45, 343-359.

DeGroot, M.H. (1959). Unbiased binomial sequential estimation. *Ann. Math. Statist.*, 30, 80-101.

Girshick, M.A. (1946). Contributions to the theory of sequential analysis I, II, III. *Ann. Math. Statist.*, 17, 123-143, 282-298.

Goodman, L.A. (1953). Sequential sampling tagging for population size problems. *Ann. Math. Statist.*, 24, 56-69.

Joanes, D.N. (1972). Sequential tests of composite hypotheses. *Biometrika*, 59, 633-637.

Lehman, E.K., y Stein, C. (1950). Completeness in the sequential case. *Ann. Math. Statist.*, 21, 376-385.

Loynes, R.M. (1969). The consistency of certain sequential estimators. *Ann. Math. Statist.*, 40, 568-574.

Ray, W.D.A. (1957). Sequential confidence intervals for the mean of a normal population with unknown variance. *J. Roy. Statist. Soc., Ser. B*, 19, -- 133-143.

Robbins, H., Simons, G., y Starr, N. (1967). A sequential analogue of the Behrens-Fisher problem. *Ann. Math. Statist.*, 38, 1384-1391.

Samuel, E. (1968). Sequential maximum likelihood estimation of the size of a population. *Ann. Math. Statist.*, 39, 1057-1068.

Sobel, M. y Wald, A. (1949). A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *Ann. Math. Statist.*, 20, 502-522.

Starr, N. (1966). The performance of a sequential procedure for the fixed-width interval estimate. *Ann. Math. Statist.*, 36, 36-50.

Stein, C. (1949). Some problems in sequential estimation. *Econometrica*, 17 77-78.

Stein, C., y Wald, A. (1947). Sequential confidence intervals for the mean - for a normal distribution with known variance. *Ann. Math. Statist.*, 18, - 427-433.

Wolfowitz, J. (1947). Efficiency of sequential estimates and Wald's equation for sequential processes. *Ann. Math. Statist.*, 18, 215-230.

