



58  
J. Gen

**UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO**

---

---

FACULTAD DE CIENCIAS

**APLICACIONES DE CONVERGENCIA DE  
FUNCIONALES A BONDAD DE AJUSTE**

**T E S I S**

QUE PARA OBTENER EL TITULO DE  
ACTUARIO

**P R E S E N T A**

**ANGELICA PRIETO MARIN**

**MEXICO D.F., AGOSTO 1994**

**TESIS CON  
FALLA DE ORIGEN**



Universidad Nacional  
Autónoma de México



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

**A MIS PADRES**

# Contenido

INTRODUCCIONiii

<b>1</b>	<b>CONVERGENCIA EN DISTRIBUCION</b>	<b>1</b>
1.1	EL ESPACIO EUCLIDIANO $R^k$ .	1
1.2	CONVERGENCIA DEBIL EN ESPACIOS METRICOS.	4
1.3	EL ESPACIO $C$ DE FUNCIONES CONTINUAS.	8
1.4	EL ESPACIO $D$ .	13
<b>2</b>	<b>EL PROBLEMA DE BONDAD DE AJUSTE</b>	<b>20</b>
2.1	EL PROBLEMA DE BONDAD DE AJUSTE	20
2.2	LA FUNCION DE DISTRIBUCION EMPIRICA.	21
2.3	ESTADISTICAS BASADAS EN LA FDE.	21
2.4	PRUEBAS FDE PARA UNA DISTRIBUCION TOTALMENTE ESPECIFICADA.	25
2.5	LA POTENCIA DE LAS ESTADISTICAS.	26
2.6	DESCRIPCION DE ALGUNOS CASOS DE INTERES.	26
2.6.1	EJEMPLO NORMAL.	27
2.6.2	EJEMPLO EXPONENCIAL.	28
2.6.3	EJEMPLO GAMMA.	30
2.7	CONCLUSIONES DEL SEGUNDO CAPITULO.	34
<b>3</b>	<b>CONVERGENCIA DEBIL DE FUNCIONALES INTEGRALES DE PROCESOS ESTOCASTICOS</b>	<b>35</b>
3.1	EL RESULTADO PRINCIPAL.	36
3.2	CONDICIONES NECESARIAS Y SUFICIENTES PARA TENSION $\Phi$ -DEBIL.	40

<b>4</b>	<b>SOLUCION AL PROBLEMA DE BONDAD DE AJUSTE POR MEDIO DE CONVERGENCIA DE FUNCIONALES: UN NUEVO ENFOQUE</b>	<b>44</b>
4.1	SOLUCION PARA LA DIFERENCIA CUADRATICA CRAMER-VON MISES. . . . .	44
4.2	SOLUCION PARA LA DIFERENCIA CUADRATICA ANDERSON-DARLING. . . . .	47
4.3	CONCLUSIONES DEL CUARTO CAPITULO . . . . .	49

# INTRODUCCION

Uno de los problemas que se plantea en la estadística es el de encontrar la distribución de los estimadores, que son funciones de la muestra. Así, por ejemplo, en pruebas de hipótesis esto se convierte en parte fundamental en la resolución del problema de aceptar o rechazar una hipótesis.

Más precisamente si  $X_1, \dots, X_n$  es una muestra aleatoria,  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^k$  es una función medible y  $Z_n = \phi(X_1, \dots, X_n)$  es una estadística, el problema consiste en encontrar la distribución de  $Z_n$ .

La distribución de estas estadísticas depende de la dimensión del vector aleatorio  $(X_1, \dots, X_n)$  y aún cuando existen técnicas para encontrar estas distribuciones, en general resulta un problema difícil que requiere de una gran cantidad de cálculos aún para una computadora.

Una de las técnicas utilizadas para el estudio de las distribuciones de estas estadísticas es el análisis asintótico de la distribución de  $Z_n$  cuando  $n \rightarrow \infty$ . Un resultado muy importante en este sentido lo da el Teorema Central del Límite, que se refiere a la convergencia de la distribución de sumas normalizadas de variables aleatorias, independientes e idénticamente distribuidas. La importancia de este teorema radica en que independientemente de cuál sea la distribución común de estas variables aleatorias, la distribución de las sumas normalizadas converge a distribución una normal estándar. Los primeros trabajos sobre este teorema se remontan al siglo XVIII con De Moivre en el caso de variables aleatorias Bernoulli con parámetro  $p = \frac{1}{2}$  y Laplace en el siglo XIX que considera  $p$  general. Posteriormente se generaliza para variables aleatorias independientes e idénticamente distribuidas con varianza finita.

El Teorema Central del Límite no es sino un caso particular de lo que se llama convergencia en distribución de variables aleatorias: Decimos que una sucesión de variables aleatorias  $\{X_n\}$  converge en distribución a  $X_0$  si  $\lim_{n \rightarrow \infty} F_n(x) = F_0(x)$  para todo  $x$  punto de continuidad de  $F_0$ , función de distribución de  $X_0$ .

Una de las implicaciones más importantes de esta distribución, como se verá más adelante es la siguiente:

Si  $\{X_n\}_{n \geq 1}$  es una sucesión de variables aleatorias tales que  $X_n$  converge en distribución a  $X_0$  y  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  es una función continua, entonces  $\phi(X_n)$  converge en distribución a  $\phi(X_0)$ .

En el caso del problema de bondad de ajuste, para distribuciones continuas del que se ocupa este trabajo, la situación que se presenta es similar a la descrita en el párrafo anterior, como se verá con esta breve explicación.

La hipótesis general puede escribirse de la siguiente manera:

$$H_0 : \text{la muestra proviene de } F_0(x; \theta)$$

donde  $F_0(x; \theta)$  es una función de distribución continua y  $\theta$  es el vector de parámetros el cual puede ser totalmente conocido, parcialmente conocido o desconocido. El caso que se trata en este trabajo, es el de  $\theta$  conocido, y en este caso  $F_0(x; \theta)$  se denota  $F_0(x)$ . A continuación se hace uso de la Función de Distribución Empírica,

$$F_n(x) = \frac{\text{número de observaciones } \leq x}{n},$$

la cual se sabe que estima a la verdadera función de distribución que produjo a la muestra y es además, por el lema de Glivenko Cantelli, un estimador consistente. El problema consiste en medir la discrepancia entre  $F_n$  y  $F_0$ , y para ello se utilizan distintas estadísticas. Para los fines de este trabajo, las estadísticas que más interesan son las de la familia de Cramér-von Mises

$$Q = n \int_{-\infty}^{\infty} \{F_n(x) - F_0(x)\}^2 \psi(x) dF_0(x)$$

donde  $\psi(x)$  es una función de la que sólo diremos por ahora que pondera las diferencias cuadráticas  $\{F_n(x) - F_0(x)\}^2$ . Ahora bien, ya elegida la estadística, es indispensable conocer su distribución bajo  $H_0$  y encontrar así la probabilidad de que la estadística se encuentre en la región de rechazo  $\mathcal{C}$  dado que  $H_0$  es cierta. En este trabajo, se comenta la solución a este problema, haciendo uso de la teoría asintótica considerando a estas estadísticas como funcionales del proceso estocástico

$$\sqrt{n}\{F_n(x) - F_0(x)\}_{x \in \mathbb{R}}.$$

A este proceso, se le aplica el teorema de la Transformación Integral de Probabilidad, el cual convierte cualquier distribución continua en una Continua Uniforme, tomando la siguiente forma:

$$\{X_n(t)\} = \sqrt{n}\{(G_n(t) - t)\}_{t \in [0,1]}$$

Donde  $G_n(t)$  es la *FDE* de la muestra transformada.

De esta manera, las estadísticas estudiadas aquí, la de Cramer-von Mises

$$n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 dF_0(x)$$

y la de Anderson Darling

$$n \int_{-\infty}^{\infty} \frac{[F_n(x) - F_0(x)]^2}{F_0(x)(1 - F_0(x))} dF_0(x)$$

son equivalentes a:

$$\int_0^1 X_n^2(t) dt$$

y

$$\int_{\epsilon}^{1-\epsilon} \frac{X_n^2(t)}{t(1-t)} dt, \quad \epsilon > 0$$

respectivamente.

De manera general, se puede afirmar que las funcionales  $\psi(x)$  aplicadas al proceso  $X_n(t)$  en las estadísticas Anderson Darling y Cramer-von Mises, son continuas para toda  $t \in (0, 1)$ .

En el capítulo 1 se probará, haciendo uso del concepto de *tensión*, la convergencia de proceso  $X_n(t)$  al Browniano Atado en el espacio  $D$  de las funciones continuas por la derecha con límite por la izquierda. Así mismo se mencionará que cuando se tiene la convergencia de un proceso a otro, ésta se sigue manteniendo cuando se aplican funciones continuas. Sin embargo, el concepto de tensión antes mencionado, resulta, en la práctica complicado y de difícil aplicación. Es por eso que esta tesis pretende presentar una nueva opción para probar la convergencia de las funcionales del proceso estocástico al Browniano Atado, dado que lo que interesa en el problema de Bondad de Ajuste es encontrar la convergencia de las funcionales del proceso y no la del proceso en sí.

El capítulo 1 toma su contenido de [1], el capítulo 2 de [3] y el capítulo 3 de [2].



# Capítulo 1

## CONVERGENCIA EN DISTRIBUCION

En este capítulo se presentan las definiciones y resultados, sin demostraciones, sobre convergencia en distribución que serán utilizados en el capítulo 3, las definiciones del movimiento Browniano (o proceso de Wiener), el Browniano Atado (o Puente Browniano) que serán los procesos de interés, así como la convergencia del proceso empírico a estos procesos. En la sección 1 se considera la convergencia en distribución de vectores aleatorios en  $\mathbb{R}^k$ . En la sección 2 se dan las definiciones y resultados generales para variables aleatorias con valores en un espacio métrico. En la sección 3 se considera el espacio de las funciones continuas con la norma del supremo y se mencionan el movimiento Browniano y el Browniano atado y por último en la sección 4 se comenta el espacio de las funciones continuas por la derecha con límite por la izquierda.

### 1.1 EL ESPACIO EUCLIDIANO $\mathbb{R}^k$ .

Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y  $X_n, n \geq k$  vectores aleatorios de dimensión  $k, k = 1, \dots, n$  definidos sobre  $(\Omega, \mathcal{F}, P)$ . Recordemos que sus funciones de distribución  $F_n, n \geq 0$  están definidas por

$$\begin{aligned} F_0(x) &= P\{X_0 \leq x\} \\ F_n(x) &= P\{X_n \leq x\} \\ &= P\{(X_{1_n}, \dots, X_{k_n}) \in (-\infty, x_1) \times \dots \times (-\infty, x_k)\}; x \in (x_1, \dots, x_k). \end{aligned}$$

**Definición 1.1.1** Sean  $X_n$ ,  $n \geq 0$ , vectores aleatorios definidos sobre  $(\Omega, \mathcal{F}, P)$ . Decimos que  $X_n$  converge en distribución a  $X_0$  si

$$F_n(x) \rightarrow F_0(x)$$

para toda  $x \in \mathbb{R}^k$ , punto de continuidad de  $F_0$ , y lo denotaremos por  $F_n \Rightarrow F_0$  ó  $X_n \xrightarrow{d} X_0$ .

### EL TEOREMA CENTRAL DEL LIMITE PARA ENSAYOS MULTINOMIALES.

Un caso particular del Teorema Central del Límite para el espacio Euclidiano  $\mathbb{R}^k$ , es el Teorema Central del Límite para ensayos multinomiales, ver [5] pag. 89, el cual se utilizará más adelante para probar la convergencia del proceso estocástico  $\sqrt{n}(F_n(x, \omega) - F_0(x))$  al proceso de Wiener.

Sea  $\mu_i$ ,  $i = 1, 2, \dots, k$  el número de ocurrencias del evento  $A_i$  en  $n$  ensayos repetidos.

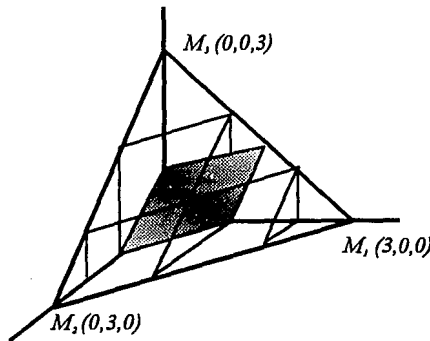


Fig. 1  
 $n=3, k=3$

Los valores  $\mu_i$  pueden tomar sólo valores  $0, 1, 2, \dots, n$ . Más aún, como sólo hay  $k$  posibles resultados en cada ensayo y como estos resultados son mutuamente excluyentes, se sabe que

$$\mu_1 + \mu_2 + \dots + \mu_k = n. \quad (1.1.1)$$

Refiéranse ahora las cantidades  $\mu_1, \mu_2, \dots, \mu_k$  como las coordenadas rectangulares de un punto en el espacio Euclidiano  $k$  dimensional. Por la ecuación 1.1.1, los resultados de los  $n$  ensayos no están representados por el hiper cubo  $0 \leq \mu_i \leq n$   $i = 1, \dots, k$ , sino

por el hiperplano 1.1.1. La figura 1 ilustra la posición de los posibles resultados dentro del hiperplano 1.1.1 para el caso  $n = 3, k = 3$ .

Tómese la transformación

$$x_i = \frac{\mu_i - np_i}{\sqrt{np_i q_i}}, \quad i = 1, 2, \dots, k; \quad q_i = 1 - p_i$$

y  $p_i$  = probabilidad de que el evento  $k$  ocurra.

Ahora, sea  $P_n(R)$  la probabilidad de que como resultado de  $n$  ensayos, el número de ocurrencias  $\mu_i, i = 1, 2, \dots, k$  de cada posible resultado será tal que

$$x_i = \frac{\mu_i - np_i}{\sqrt{np_i q_i}}$$

caiga dentro de alguna región  $R$ .

**Teorema 1.1.1** *Si en un esquema de ensayos independientes cada uno teniendo  $k$  posibles resultados en cada uno de los ensayos, la probabilidad de cada resultado no depende del número del ensayo y es distinta de 0 y de 1, entonces para una región  $R$  para la cual el volumen  $(k-1)$ -dimensional de su frontera es igual a cero, la relación*

$$P_n(R) \longrightarrow \sqrt{\frac{q_1 q_2 \cdots q_k}{(2\pi)^{k-1} \prod_{i=1}^k p_i q_i}} \int_R e^{-\frac{1}{2} \sum v_i^2} dv$$

se cumple uniformemente en  $R$  cuando  $n \rightarrow \infty$ , donde  $dv$  denota el elemento de volumen de la región  $R$  y la integración se toma sobre la región  $R$ .

### UNA DEFINICION MAS GENERAL PARA FUNCIONES DE DISTRIBUCION Y CONVERGENCIA EN DISTRIBUCION

Considérese nuevamente  $X_n, n \geq 0$  vectores aleatorios de dimensión  $k$ , definidos sobre  $(\Omega, \mathcal{F}, P)$  y denótese por  $\mathbf{P}_n, n \geq 0$  las distribuciones de  $X_n, n \geq 0$  i.e., las inducidas por los vectores aleatorios,

$$\mathbf{P}_n(A) = P(X_n^{-1}A) = P\{\omega : X_n(\omega) \in A\} = P(X_n \in A), \quad A \in B(\mathbb{R}^k)$$

y por  $\delta A$  a la frontera del conjunto  $A$ . El siguiente resultado es una definición equivalente de convergencia en distribución.

**Definición 1.1.2** Sean  $\mathbf{P}_n$ ,  $n \geq 0$  medidas de probabilidad sobre  $(\mathbb{R}^k, B(\mathbb{R}^k))$ , decimos que  $\mathbf{P}_n$  converge débilmente a  $\mathbf{P}_0$  (se denotará  $\mathbf{P}_n \Rightarrow \mathbf{P}_0$ ) si

$$\mathbf{P}_n(A) \longrightarrow \mathbf{P}_0(A) \quad \forall A \in B(\mathbb{R}^k), \quad \mathbf{P}_0(\delta A) = 0$$

**Teorema 1.1.2** Sea  $X_n$ ,  $n \geq 0$  una sucesión de vectores aleatorios definidos sobre  $(\Omega, \mathcal{F}, P)$ , entonces  $X_n \xrightarrow{D} X_0$  si y sólo si  $\mathbf{P}_n \Rightarrow \mathbf{P}_0$ , donde  $\mathbf{P}_n$ ,  $n \geq 0$  son las distribuciones de  $X_n$ ,  $n \geq 0$ .

Si bien el concepto de convergencia en distribución de vectores aleatorios está asociado a  $\mathbb{R}^k$ , el teorema 1.1.2 sugiere que este concepto puede ser formulado en términos de convergencia débil de medidas de probabilidad definidas sobre un espacio  $(S, \mathcal{S})$  donde  $S$  es un espacio métrico y  $\mathcal{S}$  es la mínima  $\sigma$ -álgebra generada por los conjuntos abiertos y es lo que se tratará en la siguiente sección.

## 1.2 CONVERGENCIA DEBIL EN ESPACIOS METRICOS.

En esta sección se presentarán nuevamente las definiciones de convergencia débil y en distribución en forma generalizada para espacios métricos cualesquiera, así como nuevas herramientas que ayudarán a concluir la convergencia de los procesos estocásticos al Browniano atado. Recuérdese que los resultados se presentan sin demostraciones, para mayor información véase [1].

### CONVERGENCIA DEBIL

Sea  $S$  un espacio métrico,  $\mathcal{S}$  la familia de subconjuntos de Borel de  $S$  ( $\mathcal{S}$  es la mínima  $\sigma$ -álgebra generada por los conjuntos abiertos).

**Definición 1.2.1** Sean  $\mathbf{P}_n$ ,  $n \geq 0$ , medidas de probabilidad sobre  $(S, \mathcal{S})$ , decimos que  $\mathbf{P}_n$  converge débilmente a  $\mathbf{P}_0$  si

$$\mathbf{P}_n(A) \longrightarrow \mathbf{P}_0(A), \quad \forall A \in \mathcal{S}, \quad \mathbf{P}_0(\delta A) = 0$$

### EL TEOREMA DE PORTMANTEAU.

El teorema de Portmanteau ver [1] pag. 11, aporta condiciones útiles y equivalentes para la convergencia débil. Cada una de estas condiciones sirve además como definición.

**Teorema 1.2.1** Sean  $P_n$ ,  $n \geq 0$ , medidas de probabilidad en  $(S, \mathcal{S})$ . Entonces estas dos condiciones son equivalentes:

i)  $P_n \Rightarrow P_0$

ii)  $\lim_n \int f dP_n = \int f dP_0$  para toda  $f$  con valores en  $\mathbb{R}$ , acotada y uniforme continua.

### UN TEOREMA PARA CONVERGENCIA DEBIL.

El siguiente teorema nos dará una condición útil para determinar la convergencia débil de una sucesión  $\{P_n\}$  de medidas de probabilidad a otra que denotaremos  $P_0$  ver [1] pag. 16:

**Teorema 1.2.2** Sean  $P_n$ ,  $n \geq 0$  medidas de probabilidad en  $(S, \mathcal{S})$ , entonces  $P_n \Rightarrow P_0$  si y sólo si cada subsucesión  $\{P_{n'}\}$  contiene otra subsucesión  $\{P_{n''}\}$  tal que  $P_{n''} \Rightarrow P_0$ .

### MEDIDAS DE PROBABILIDAD.

Sea  $X$  una función del espacio de probabilidad  $(\Omega, \mathcal{B}, \mathcal{P})$  en un espacio métrico  $S$ . Si  $X$  es medible se llamará elemento aleatorio. La distribución de  $X$  es la medida de probabilidad  $P = PX^{-1}$  en  $(S, \mathcal{S})$  dada por:

$$P(A) = P(X^{-1}A) = P\{\omega : X(\omega) \in A\} = P\{X \in A\}, A \in \mathcal{S}$$

### CONVERGENCIA EN DISTRIBUCION.

Se dice que la sucesión  $\{X_n\}$  de elementos aleatorios converge en distribución al elemento aleatorio  $X_0$ , y se escribe

$$X_n \xrightarrow{D} X_0$$

si las distribuciones  $P_n$  de las  $X_n$  convergen débilmente a la distribución  $P_0$  de  $X_0$ , es decir,

$$P_n \Rightarrow P_0$$

### CONVERGENCIA DEBIL Y MAPEOS.

Como se verá en el capítulo 2, se requieren condiciones bajo las cuales, al aplicar un mapeo a una sucesión de medidas de probabilidad que converge, se mantenga la convergencia. Estas condiciones se presentan a continuación. Si  $h$  es un mapeo medible de  $S$  en otro espacio métrico  $S'$  (con métrica  $\rho'$  y  $\sigma$ -álgebra  $\mathcal{S}'$  de conjuntos borelianos), entonces cada medida de probabilidad en  $(S, \mathcal{S})$  induce en  $(S', \mathcal{S}')$  una única medida

$P_h^{-1}$ , definida por  $P_h^{-1}(A) = P(h^{-1}A)$  para  $A \in S'$ . Se necesitan condiciones para las cuales  $P_n \Rightarrow P_0$  implique  $P_n h^{-1} \Rightarrow P_0 h^{-1}$ . Una de estas condiciones (ver [1] pag. 29) es que  $h$  sea continua. Tómesese ahora,  $\phi(y)$  otra función acotada y continua en  $S'$ . Como entonces,  $\phi(h(x))$  es acotada y continua en  $S$ ,  $P_n \Rightarrow P_0$  implica  $\int \phi(h(x))P_n(dx) \rightarrow \int \phi(h(x))P_0(dx)$ , una relación que, bajo la transformación de las integrales es

$$\int \phi(y)P_n h^{-1}(dy) \rightarrow \int \phi(y)P_0 h^{-1}(dy).$$

Por ejemplo, la proyección natural  $\pi_k$  de  $R^\infty$  en  $R^k$  es continua, se tiene que  $P_n \Rightarrow P_0$  implica  $P_n \pi_k^{-1} \Rightarrow P_0 \pi_k^{-1}$  para cada  $k$ . Recíprocamente si  $P_n \pi_k^{-1} \Rightarrow P_0 \pi_k^{-1}$  para cada  $k$  entonces  $P_n \Rightarrow P_0$ .

### EL TEOREMA PRINCIPAL.

Se ha visto que  $P_n \Rightarrow P_0$  implica  $P_n h^{-1} \Rightarrow P_0 h^{-1}$  si  $h$  es un mapeo continuo de  $S$  en  $S'$ , pero esta suposición se puede hacer más débil. Suponga sólo que  $h$  es medible y sea  $D_h$  el conjunto de discontinuidades de  $h$ , entonces  $D_h \in S$  y tenemos el siguiente resultado (ver [1] pag. 30):

**Teorema 1.2.3** Si  $P_n \Rightarrow P_0$  y  $P_0(D_h) = 0$ , entonces  $P_n h^{-1} \Rightarrow P_0 h^{-1}$ .

### UNA EXTENSION DEL TEOREMA.

Sean  $h_n$  y  $h$  mapeos medibles de  $S$  a  $S'$ . Podría preguntarse si  $P_n \Rightarrow P_0$  implica  $P_n h_n^{-1} \Rightarrow P_0 h^{-1}$  cuando  $h_n$  converge a  $h$  de algún modo. Sea  $E$  el conjunto de  $x$  tal que  $h_n(x_n) \rightarrow h(x)$  falla para alguna sucesión  $\{x_n\}$  aproximación de  $x$ . Entonces se tiene que ([1] pag. 34):

**Teorema 1.2.4** Si  $P_n \Rightarrow P_0$  y  $P_0(E) = 0$  entonces  $P_n h_n^{-1} \Rightarrow P_0 h^{-1}$ .

### OTROS TEOREMAS UTILES PARA VERIFICAR CONVERGENCIA DEBIL.

#### FAMILIAS DE MEDIDAS DE PROBABILIDAD RELATIVAMENTE COMPACTAS.

Sea  $\Pi$  una familia de medidas de probabilidad en  $(S, S)$ .

**Definición 1.2.2** Se dice que  $\Pi$  es relativamente compacta si cada sucesión de elementos de  $\Pi$  contiene una subsucesión que converge débilmente, esto es, si para cada sucesión  $\{P_n\}$  existe una subsucesión  $\{P_{n'}\}$  y una medida de probabilidad  $Q$  (definida en  $(S, S)$ ), pero no necesariamente un elemento de  $\Pi$ ), tal que  $P_{n'} \Rightarrow Q$ .

El concepto de Medidas de Probabilidad Relativamente Compactas aportará una técnica muy importante para probar convergencia débil de medidas de probabilidad sobre  $C$ , el espacio de las funciones continuas con la norma del supremo y en otros espacios de funciones. Suponga que  $\mathbf{P}_n$  y  $\mathbf{P}_0$  son medidas de probabilidad en  $(C, \mathcal{C})$ , donde  $\mathcal{C}$  es la mínima  $\sigma$ -álgebra de subconjuntos de  $C$  (se especifica más ampliamente posteriormente), y que las distribuciones finito-dimensionales de  $\mathbf{P}_n$  convergen débilmente a la distribución  $\mathbf{P}_0$  restringida a las correspondientes conjuntos finito-dimensionales. Como se verá más adelante,  $\mathbf{P}_n$  no necesariamente converge débilmente a  $\mathbf{P}_0$ . Suponga, sin embargo, que  $\{\mathbf{P}_n\}$  es relativamente compacta. Entonces cada subsucesión  $\{\mathbf{P}_{n'}\}$  contiene otra subsucesión  $\{\mathbf{P}_{n''}\}$  que converge débilmente a algún límite  $Q$ . Las distribuciones finito-dimensionales de  $Q$  deben ser los límites débiles de aquellos de  $\{\mathbf{P}_{n''}\}$  y por lo tanto deben de coincidir con las distribuciones finito-dimensionales de  $\mathbf{P}_0$  (para cada  $(t_1, \dots, t_k)$ ,  $\mathbf{P}_{n''}\pi_{t_1, \dots, t_k}^{-1} \Rightarrow Q\pi_{t_1, \dots, t_k}^{-1}$  y  $\mathbf{P}_{n''}\pi_{t_1, \dots, t_k}^{-1} \Rightarrow \mathbf{P}_0\pi_{t_1, \dots, t_k}^{-1}$  así que  $Q\pi_{t_1, \dots, t_k}^{-1} = \mathbf{P}_0\pi_{t_1, \dots, t_k}^{-1}$ ). Pero entonces, como una medida de probabilidad en  $C$  está completamente determinada por sus distribuciones finito-dimensionales  $Q$  y  $\mathbf{P}_0$  deben coincidir ver [1] pag. 35. De esta manera, cada subsucesión  $\{\mathbf{P}_{n'}\}$  contiene una subsucesión que converge débilmente a  $\mathbf{P}_0$ , y se sigue del Teorema 1.1.2, que la sucesión  $\{\mathbf{P}_n\}$  converge débilmente a  $\mathbf{P}_0$ . Supóngase ahora que sabemos que  $\{\mathbf{P}_n\}$  es relativamente compacta y que, para cada  $(t_1, \dots, t_k)$ ,  $\mathbf{P}_n\pi_{(t_1, \dots, t_k)}^{-1}$  converge débilmente a alguna medida de probabilidad  $\mu_{(t_1, \dots, t_k)}$  en  $(\mathbb{R}^k, B(\mathbb{R}^k))$ . El punto que omitimos ahora en nuestra suposición es que  $\mu_{(t_1, \dots, t_k)}$  son las distribuciones finito-dimensionales de una medida de probabilidad en  $(C, \mathcal{C})$ . Aún así se cumple que cada subsucesión  $\{\mathbf{P}_{n'}\}$  contiene otra subsucesión  $\{\mathbf{P}_{n''}\}$  que converge débilmente a algún límite. Como este límite debe tener a las  $\mu_{t_1, \dots, t_k}$  como sus distribuciones finito-dimensionales, este límite es único. Por lo tanto  $\{\mathbf{P}_n\}$  converge débilmente a algún  $\mathbf{P}_0$ .

En resumen, si  $\{\mathbf{P}_n\}$  es una sucesión de medidas de probabilidad tal que las distribuciones finito-dimensionales de  $\{\mathbf{P}_n\}$  convergen débilmente y además dicha sucesión es relativamente compacta, entonces podremos probar su convergencia débil a una medida de probabilidad  $\mathbf{P}_0$ . Por ello, es necesario tener un criterio para determinar que una sucesión  $\{\mathbf{P}_n\}$  de medidas de probabilidad es relativamente compacta, para lo cual haremos uso de la siguiente definición:

**Definición 1.2.3** En una familia  $\Pi$  de medidas de probabilidad en el espacio métrico general  $S$ , se dice que es *tenso* si para cada  $\epsilon$  positiva existe un conjunto compacto  $K$  tal que  $\mathbf{P}(K) > 1 - \epsilon$  para toda  $\mathbf{P}$  en  $\Pi$ .

**Teorema 1.2.5** Si  $\Pi$  es una familia de medidas de probabilidad en  $(S, S)$ , un espacio métrico, que es *tenso*, entonces  $\Pi$  es *relativamente compacta*.

Ver [1] pag. 37.

**Teorema 1.2.6** Suponga que  $S$  es separable y completo. Si  $\Pi$  es *relativamente compacta* entonces es *tenso*.

Ver [1] pag. 37.

A continuación se estudiará la convergencia débil de los procesos estocásticos en dos espacios de interés:  $S = C[0, 1]$  el espacio de las funciones continuas y  $S = D[0, 1]$  el de las continuas por la derecha con límite por la izquierda.

### 1.3 EL ESPACIO $C$ DE FUNCIONES CONTINUAS.

Sea  $C = C[0, 1]$ , el espacio de todas las funciones continuas en  $[0, 1]$ . Para los puntos  $t_1, \dots, t_k$ , sea  $\pi_{t_1, \dots, t_k}$  el mapeo que lleva el punto  $x$  de  $C$  al punto  $(x(t_1), \dots, x(t_k))$  de  $\mathbb{R}^k$ . Los conjuntos finito- dimensionales son ahora definidos como conjuntos de la forma  $\pi_{t_1, \dots, t_k}^{-1} H$  donde  $H \in \mathbb{R}^k$ .

Definimos también la distancia entre dos puntos  $x$  y  $y$  (dos funciones  $x$  y  $y$  de  $t \in [0, 1]$ ) como

$$\rho(x, y) = \sup_{t \in [0, 1]} |x(t) - y(t)|$$

Se puede demostrar con un contra-ejemplo que si se tiene convergencia de distribuciones finito-dimensionales de  $\mathbf{P}_n$ , es decir, convergencia débil de el pedir la convergencia débil de  $\mathbf{P}_n$  en la clase de los conjuntos finito-dimensionales, lo anterior no es suficiente para tener que  $\mathbf{P}_n \Rightarrow \mathbf{P}_0$  en el espacio  $C$ , ver [1] pag. 20.

Sea  $\mathbf{P}_0$  la masa unitaria en el 0, y sea  $\mathbf{P}_n$  la masa unitaria en la función  $x_n$ , donde

$$x_n(t) = \begin{cases} nt & \text{si } 0 \leq t \leq \frac{1}{n} \\ 2 - nt & \text{si } \frac{1}{n} \leq t \leq \frac{2}{n} \\ 0 & \text{si } \frac{2}{n} \leq t \leq 1 \end{cases}$$



La figura 2 representa dicha función. Como  $x_n$  no converge a 0 uniformemente, (esto es, en la topología de  $C$ ),  $P_n$  no puede converger débilmente a  $P_0$ . (Por ejemplo, si  $A$  es el intervalo abierto  $A = S(0, \frac{1}{2})$ , entonces  $P_0(\delta A) = 0$ , mientras que  $P_n(A) = 0$  no converge a  $P_0(A) = 1$ ). Pero  $P_n(A) \rightarrow P_0(A)$  se cumple para conjuntos finito dimensionales y  $P_0$ -continuos (es decir que  $P_0(\delta A) = 0$ ). De hecho, si  $A = \pi_{t_1, \dots, t_k}^{-1} H$  y si  $\frac{2}{n}$  es más pequeño que la mínima  $t_i$  diferente de cero, entonces  $P_n(A) = P_0(A)$ .

Por lo tanto, es necesario buscar condiciones de convergencia para  $C$  más allá de la

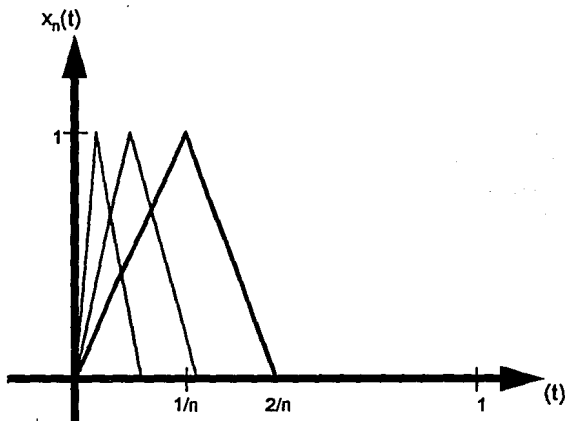


Figura 2

convergencia sobre la clase de los conjuntos finito-dimensionales. Esta nueva condición es la de tensión, como ya se vio en la sección anterior de espacios métricos.

**Teorema 1.3.1** Sean  $P_n, P_0$  medidas de probabilidad en  $(C, C)$ . Si las distribuciones finito-dimensionales de  $P_n$  convergen débilmente a los de  $P_0$  y si  $\{P_n\}$  es tensa, entonces  $P_n \Rightarrow P_0$ .

Ver [1] pag. 54.

Se hace aquí una pausa para comentar las características de la medida de Wiener y del Browniano atado ya que, como se verá más adelante, el proceso estocástico (definido después), cuya distribución asintótica es necesaria en el problema de Bondad de Ajuste, converge débilmente a un Browniano atado.

## LA MEDIDA DE WIENER.

La medida de Wiener, denotada por  $W$ , es la medida de probabilidad en  $(C, C)$  que tiene las dos propiedades siguientes:

1) Para cada  $t$ , la variable aleatoria  $X_t$  se distribuye normalmente bajo  $W$  con media 0 y varianza  $t$ :

$$W\{X_t \leq \alpha\} = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\alpha} e^{-\frac{u^2}{2t}} du$$

Si  $t=0$ , se interpreta como  $W\{X_0 = 0\} = 1$ .

2) El proceso estocástico  $\{X_t : 0 \leq t \leq 1\}$  tiene incrementos independientes bajo  $W$ :  
Si

$$0 \leq t_0 \leq t_1 \leq \dots \leq t_k \leq 1$$

entonces las variables aleatorias

$$(X_{t_1} - X_{t_0}), (X_{t_2} - X_{t_1}), \dots, (X_{t_k} - X_{t_{k-1}})$$

son independientes bajo  $W$ . Para la prueba de la existencia de tal medida consultar [1] pag. 64.

## EL BROWNIANO ATADO.

El elemento aleatorio  $X$  de  $C$  es Gaussiano si todas sus distribuciones finito dimensionales son gaussianas. La distribución en  $C$  de un elemento aleatorio Gaussiano está completamente especificada por las medidas  $E\{X(t)\}$  y los momentos cruzados  $E\{X(s)X(t)\}$ ,  $0 \leq s, t \leq 1$  porque estas determinan las distribuciones finito dimensionales.

Para  $X$  los momentos son

$$E\{X_t\} = 0$$

y

$$E\{X_s X_t\} = s \text{ si } s \leq t$$

y esto tiene, si los incrementos son independientes, asociada una medida de Wiener.

Para estudiar el comportamiento de funciones de distribución empírica, se necesitarán los elementos aleatorios Gaussianos  $X^0$  de  $C$  cuya distribución es especificada por los requerimientos

$$E\{X_t^0\} = 0$$

y

$$E\{X_s^0 X_t^0\} = s(1-t) \text{ si } s \leq t$$

Aunque se puede demostrar la existencia de una medida  $W^0$ , la medida asociada al proceso  $\{X_t^0\}$ , por diferentes métodos, se construirá definiendo (ver [1] pag. 64):

$$X_t^0 = X_t - tX_1, \quad 0 \leq t \leq 1$$

El elemento aleatorio  $X^0$  se llama Browniano atado. Note que  $X_0^0 = X_1^0 = 0$  con probabilidad 1. De los requerimientos para el Browniano atado tenemos que

$$E\{(X_t^0 - X_s^0)^2\} = (t-s)(1-(t-s)) \text{ si } s \leq t$$

y que

$$E\{(X_{s_2}^0 - X_{s_1}^0)(X_{t_2}^0 - X_{t_1}^0)\} = -(s_2 - s_1)(t_2 - t_1) \text{ si } s_1 \leq s_2 \leq t_1 \leq t_2$$

Se usará  $W^0$  para denotar la distribución en  $C$  del elemento aleatorio  $X^0$ . Si  $h: C \rightarrow C$  lleva la función  $x$  a la función con valor  $x(t) - tx(1)$  en  $t$ , entonces las medidas  $W^0$  y  $W$  están relacionadas por  $W^0 = Wh^{-1}$ .

### LAS FUNCIONES DE DISTRIBUCION EMPIRICAS.

Se introduce ahora, este concepto que lleva al uso de los procesos estocásticos de interés.

Sean  $T_1, T_2, \dots$  variables aleatorias en algún espacio de probabilidad  $(\Omega, \mathcal{B}, \mathcal{P})$ . Se supondrá que

$$0 \leq T_n(\omega) \leq 1,$$

(lo cual puede arreglarse por medio de la Transformación con la Integral de Probabilidad, y de esta manera, la distribución común de las  $T_i$  es uniforme en  $[0, 1]$ ).

**Definición 1.3.1** La función de distribución empírica, la FDE,  $F_n(t, \omega)$ , correspondiente a los puntos  $T_1(\omega), \dots, T_n(\omega)$  se define, para  $0 \leq t \leq 1$  como el número  $i \leq n$  de valores para los cuales  $T_i(\omega) \leq t$  multiplicado por  $\frac{1}{n}$ , i.e.,

$$F_n(t) = \frac{\text{número de observaciones} \leq t}{n}.$$

Considérese la función

$$X_n(t, \omega) = \sqrt{n}(F_n(t, \omega) - t)$$

en  $t$ . Aunque  $X_n(t, \omega)$  es una función en  $[0,1]$  producida aleatoriamente, no es un elemento de  $C$ , siendo, obviamente, discontinua. En la siguiente sección se analizará  $X_n$  como un elemento aleatorio de un espacio métrico de funciones discontinuas. Sin embargo, para resolver el problema en  $C$ , se adoptará una definición diferente para la función de distribución empírica.

Sea  $G_n(t, \omega)$  una función de  $t$  en  $[0,1]$ , la función de distribución correspondiente a

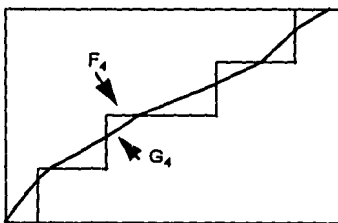


Figura 3

una uniforme que distribuye masa  $(n+1)^{-1}$  sobre cada uno de los  $(n+1)$  intervalos  $[T_{i-1}(\omega), T_i(\omega)]$ , donde  $T_{(0)} = 0, T_{(n+1)} = 1$ , y  $T_{(1)}, \dots, T_{(n)}$  son los valores  $T_1, \dots, T_n$  ordenados en forma ascendente.

Las funciones  $F_n(t, \omega)$  y  $G_n(t, \omega)$  son cercanas como puede verse en la figura 3:

$$|F_n(t, \omega) - G_n(t, \omega)| \leq \frac{1}{n}, \quad 0 \leq t \leq 1$$

Ahora sea  $Z_n(\omega)$  el elemento de  $C$  con valor

$$Z_n(t, \omega) = \sqrt{n}(G_n(t, \omega) - t)$$

Entonces

$$Z_n(t, \omega) - X_n(t, \omega) = \sqrt{n}\{G_n(t, \omega) - F_n(t, \omega)\}$$

y

$$\sup_t |X_n(t, \omega) - Z_n(t, \omega)| \leq \frac{1}{\sqrt{n}}.$$

**Teorema 1.3.2** Si las  $T_i$  son independientes e uniformemente distribuidas en  $[0,1]$ , y si  $Z_n$  se define como se hizo anteriormente, entonces,

$$Z_n \xrightarrow{D} W^0$$

donde  $W^0$  es el Browniano atado.

Si el lector quiere verificar esta demostración, tiene que probar primero que  $Z_n$  es tensa véase [1] pag. 105. El análisis que aquí se hace utiliza a la sucesión  $Z_n$  (que es cercana a  $X_n$ ) sólo para poder permanecer en  $C$ , y probar la convergencia débil en este espacio. En la siguiente sección se analizará al proceso empírico  $X_n$  en su espacio natural para probar  $X_n \xrightarrow{D} W^0$ .

## 1.4 EL ESPACIO $D$ .

Siendo el proceso estocástico de interés con trayectorias en un espacio de una funciones continuas por la derecha con límite por la izquierda, este espacio de funciones es el espacio natural para el estudio del mismo. Sea  $D = D[0,1]$  el espacio de las funciones  $x$  en  $[0,1]$  tales que son continuas por la derecha y tienen límite por la izquierda:

- i) Para  $0 \leq t < 1$ ,  $x(t+) = \lim_{s \downarrow t} x(s)$  existe y  $x(t+) = x(t)$
- ii) Para  $0 < t \leq 1$ ,  $x(t-) = \lim_{s \uparrow t} x(s)$  existe.

### LA TOPOLOGIA DE SKOROHOD.

Dos funciones  $x$  y  $y$  son "cercanas" en la topología uniforme usada para el espacio  $C$ , si la gráfica de  $x(t)$  puede llevarse a la gráfica de  $y(t)$  mediante una pequeña perturbación uniforme de las ordenadas, dejando fijas a las abscisas. En el espacio  $D$ , se permitirá una pequeña deformación uniforme en el eje de las abscisas, es decir, en la variable independiente  $t$  que representa el tiempo. La topología de Skorohod desarrolla esta idea.

Sea  $\Lambda$  la clase de todos los mapeos continuos estrictamente crecientes de  $[0,1]$  en él mismo. Si  $\lambda \in \Lambda$  entonces  $\lambda(0) = 0$  y  $\lambda(1) = 1$ . Para  $x$  y  $y$  en  $D$ , defínase  $d(x, y)$  como

el ínfimo de aquellas  $\epsilon > 0$  para las cuales existe  $\lambda \in \Lambda$  tal que

$$\sup_t |\lambda t - t| \leq \epsilon$$

y

$$\sup_t |x(t) - y(\lambda t)| \leq \epsilon$$

Se puede demostrar que  $d(x, y)$  es una métrica ([1] pag. 111) y define a la topología de Skorohod.

Un elemento  $x_n$  en  $D$  converge a un límite  $x$  en la topología de Skorohod si y sólo si existen funciones  $\lambda_n \in \Lambda$  tales que

$$\lim_n x_n(\lambda_n t) = x(t) \tag{1.4.2}$$

uniformemente en  $t$  y

$$\lim_n \lambda_n t = t \tag{1.4.3}$$

uniformemente en  $t$ .

NOTA: Si  $\lambda_n$  está en  $\Lambda$ , entonces  $\lambda_n^{-1}$  también y además [1] pag. 111

$$\sup_t |\lambda_n^{-1} t - t| = \sup_t |\lambda_n t - t|$$

y

$$\sup_t |x(\lambda_n^{-1} t) - x_n(t)| = \sup_t |x(t) - x_n(\lambda_n t)|.$$

Sin embargo, el espacio  $D$  no es completo bajo la métrica  $d$  (ver [1] pag. 112). Por lo tanto, es necesario introducir otra métrica, denotada  $d_0$ , que es equivalente a  $d$ , es decir, define a la topología de Skorohod, pero bajo la cual  $D$  es completo. Como se sabe, el pedir que un espacio métrico sea completo facilita la caracterización de conjuntos compactos.

Sea  $\lambda$  una función no decreciente en  $[0, 1]$  con  $\lambda(0) = 0$  y  $\lambda(1) = 1$ . Sea

$$\|\lambda\| = \sup_{s \neq t} \left| \log \frac{\lambda t - \lambda s}{t - s} \right|$$

Ahora, sea  $d_0(x, y)$  el ínfimo de aquellas  $\epsilon$  positivas para las cuales  $\Lambda$  contiene alguna  $\lambda$  con

$$\|\lambda\| \leq \epsilon$$

y

$$\sup_t |x(t) - y(\lambda t)| \leq \epsilon.$$

**Teorema 1.4.1** *El espacio  $D$  es completo en la métrica  $d_0$ .*

Ver [1] pag. 115.

### CONJUNTOS FINITO-DIMENSIONALES.

Los conjuntos finito-dimensionales desempeñan en  $D$  el mismo papel que en  $C$ . Para puntos  $t_1, \dots, t_k$  en  $[0,1]$ , define la proyección natural  $\pi_{t_1, \dots, t_k}$  de  $D$  en  $\mathbb{R}^k$  como:

$$\pi_{t_1, \dots, t_k}(x) = (x(t_1), \dots, x(t_k))$$

Ahora,  $\pi_0$  y  $\pi_1$  son continuas en todas partes. Suponga que  $0 < t < 1$ . Si los puntos  $x_n$  convergen a  $x$  en la Topología de Skorohod, y si  $x$  es continua en  $t$ , entonces ([1] pag. 112)  $x_n(t) \rightarrow x(t)$ . Suponga, por otro lado, que  $x$  es discontinuo en  $t$ . Si  $\lambda_n$  es el elemento de  $\Lambda$  que es lineal en  $[0, t]$  y en  $[t, 1]$  y satisface  $\lambda_n t = t - \frac{1}{n}$ , y si  $x_n(s) = x(\lambda_n s)$ , entonces  $x_n$  converge a  $x$  en la topología de Skorohod pero  $x_n(t)$  no converge a  $x(t)$ .

Entonces, si  $0 < t < 1$ , entonces  $\pi_1$  es continua en  $x$  si y sólo si  $x$  es continuo en  $t$ .

Se requiere comprobar que  $\pi_{t_1, \dots, t_k}$  es medible con respecto a la  $\sigma$ -álgebra  $\mathcal{D}$  de conjuntos Borelianos para la topología de Skorohod.

Para probar que  $\pi_{t_1, \dots, t_k}$  es medible con respecto a la  $\sigma$ -álgebra de los conjuntos de Borel en la topología de Skorohod, sólo necesitamos probarlo para un punto  $t$  (en virtud de que una función con valores en  $\mathbb{R}^k$  es medible si cada componente de la función lo es), para el cual asumimos  $t < 1$  ( $\pi_1$  es continua).

Si  $x_n$  converge a  $x$  en la topología de Skorohod, entonces existe una sucesión  $\{\lambda_n\}_n$  en  $\Lambda$  tal que se cumplen 1.4.2 y 1.4.3; pero usando la desigualdad del triángulo y la nota de arriba tenemos

$$|x_n(t) - x(t)| \leq |x_n(t) - x(\lambda_n^{-1}t)| + |x(\lambda_n^{-1}t) - x(t)|$$

de donde  $x_n(t) \rightarrow_{n \rightarrow \infty} x(t)$  para todo punto de continuidad de  $t$  en  $x$ . De lo anterior se sigue que  $x_n(s) \rightarrow_{n \rightarrow \infty} x(s)$  para toda  $s$  afuera de un conjunto de medida de Lebesgue 0 (si  $x \in D[0, 1]$  entonces  $x$  tiene a lo más un número numerable de discontinuidades).

Además como  $x_n$  está en  $D[0, 1]$ , para cada  $n = 1, 2, \dots$  y  $x_n$  converge a  $x$  en la

topología de Skorohod, la sucesión  $\{x_n\}_n$  es uniformemente acotada, luego usando el teorema de convergencia dominada

$$\frac{1}{\epsilon} \int_t^{t+\epsilon} x_n(s) ds \xrightarrow{n \rightarrow \infty} \frac{1}{\epsilon} \int_t^{t+\epsilon} x(s) ds, \quad \forall \epsilon > 0.$$

Lo anterior nos dice que  $\eta_\epsilon(x) := \frac{1}{\epsilon} \int_t^{t+\epsilon} x(s) ds$  es continua en la topología de Skorohod, por tanto,  $\eta_\epsilon : D[0, 1] \rightarrow \mathbb{R}$  es  $D$ -medible.

Por continuidad por la derecha de los elementos de  $D[0, 1]$ ,  $\eta_\epsilon(x) \rightarrow \pi_t(x)$  ( $= x(t)$ ) para cada  $x$ , cuando  $\epsilon \downarrow 0$ . Entonces  $\pi_t(x)$  es límite de funciones  $D$ -medibles y por tanto es  $D$ -medible. Entonces  $\pi_t(x)$  es medible.

Habiendo visto esto, definimos los conjuntos finito-dimensionales, al igual que en  $C$ , a los conjuntos de la forma  $\pi_{t_1, \dots, t_k}^{-1} H$  con  $k \geq 1$  y  $H \in \mathbb{R}^k$ . Cada conjunto finito-dimensional está en  $\mathcal{D}$  por definición de medibilidad.

### DISTRIBUCIONES FINITO-DIMENSIONALES.

Para una medida de probabilidad  $\mathbf{P}_0 \in (D, \mathcal{D})$  sea  $T_{\mathbf{P}_0}$  el conjunto de los puntos  $t \in [0, 1]$  para los cuales la proyección  $\pi_t$  es continua excepto en los puntos que pertenecen a un conjunto de  $\mathbf{P}_0$  - medida 0. Los puntos 0 y 1 pertenecen siempre al conjunto  $T_{\mathbf{P}_0}$ . Si  $0 < t < 1$ , entonces  $t \in T_{\mathbf{P}_0}$  si y sólo si  $\mathbf{P}_0(J_t) = 0$  donde

$$J_t = \{x : x(t) \neq x(t-)\}$$

es el conjunto de  $x$  que son discontinuas en  $t$ . Recordar que para  $0 < t < 1$ ,  $\pi_t$  es continua en  $x$  si y solo si  $x$  es continua en  $t$ .

Un elemento de  $D$  tiene a lo más un número numerable de saltos. Se tiene que  $\mathbf{P}_0(J_t) > 0$  se cumple a lo más para un número numerable de  $t$ .

Supóngase que

$$\mathbf{P}_n \Rightarrow \mathbf{P}_0$$

donde  $\mathbf{P}_n$  y  $\mathbf{P}_0$  son medidas de probabilidad en  $(D, \mathcal{D})$ . Entonces, por el teorema principal 1.2.3

$$\mathbf{P}_n \pi_{t_1, \dots, t_k}^{-1} \Rightarrow \mathbf{P}_0 \pi_{t_1, \dots, t_k}^{-1}$$

se cumple para todo  $t_i$  en  $T_{\mathbf{P}_0}$ .

**Teorema 1.4.2** Si  $\{\mathbf{P}_n\}$  es tenso,  $\mathbf{P}_n \pi_{t_1, \dots, t_k}^{-1} \Rightarrow \mathbf{P}_0 \pi_{t_1, \dots, t_k}^{-1}$  y se cumple siempre que  $t_1, \dots, t_k \in T_{\mathbf{P}_0}$  entonces,  $\mathbf{P}_n \Rightarrow \mathbf{P}_0$ .



Ver [1] pag. 125.

### UN CRITERIO PARA CONVERGENCIA.

**Teorema 1.4.3** Sean  $X_n, n \geq 0$  elementos aleatorios de  $D$ . Escribáse  $T_{X_0}$  por  $T_{\mathbf{P}_0}$ , donde  $\mathbf{P}_0$  es la distribución de  $X_0$ . Supóngase que

$$(X_n(t_1), \dots, X_n(t_k)) \xrightarrow{\mathcal{D}} (X_0(t_1), \dots, X_0(t_k))$$

Si se tiene que  $t_1, \dots, t_k$  está en  $T_{X_0}$ , que  $\mathbf{P}_0\{X_0(1) \neq X_0(1-)\} = 0$  y que

$$\mathbf{P}_0\{|X_n(t) - X_n(t_1)| \geq \lambda, |X_n(t_2) - X_n(t)| \geq \lambda\} \leq \frac{1}{\lambda^{2\gamma}} [F(t_2) - F(t_1)]^{2\alpha}$$

para  $t_1 \leq t \leq t_2$  y  $n \geq 1$ , donde  $\gamma > 0$ ,  $\alpha > \frac{1}{2}$ , y  $F$  es una función continua no decreciente en  $[0, 1]$ , entonces  $X_n \xrightarrow{\mathcal{D}} X_0$ .

Hay una versión más restrictiva de este teorema, la cual involucra momentos:

$$E\{|X_n(t) - X_n(t_1)|^r |X_n(t_2) - X_n(t)|^r\} \leq [F(t_2) - F(t_1)]^{2\alpha}$$

Ver [1] pag. 128.

Esta última versión es la que se utilizará más adelante para tener la convergencia del proceso de interés.

### FUNCION DE DISTRIBUCION EMPIRICA.

Sean  $T_1, T_2, \dots$  variables aleatorias en algún  $(\Omega, \mathcal{B}, \mathcal{P})$ . Se supondrá que

$$0 \leq T_n(\omega) \leq 1,$$

(lo cual puede arreglarse por medio de la Transformación con la Integral de Probabilidad, y de esta manera, la distribución común de las  $T_i$  es uniforme en  $[0, 1]$ .) Considérese la FDE,  $G_n(t, \omega)$  correspondiente a los puntos  $T_{(1)}(\omega), \dots, T_{(n)}(\omega)$  i.e., a la muestra ordenada. Considérese la función

$$X_n(t, \omega) = \sqrt{n}(G_n(t, \omega) - t)$$

en  $0 \leq t \leq 1$ . Como cada  $X_n(t)$  es una variable aleatoria,  $X_n$  es un elemento aleatorio de  $D$  (extendido a  $\mathbb{R}$ ). Tomemos ahora el proceso  $\{X_n(t) : t \in \mathbb{R}\}$  al que se le conoce como proceso empírico.

**Teorema 1.4.4** *Suponga que las  $T_n$  son independientes y tienen función de distribución común. Entonces*

$$X_n \xrightarrow{D} X^0$$

donde  $X^0$  es el elemento aleatorio Gaussiano de  $D$  especificado por

$$\begin{cases} E\{X(t_1)\} = 0 \\ E\{X(t_1)X(t_2)\} = t_1(1 - t_2) \quad t_1 \leq t_2 \end{cases}$$

*Demostración:* La medida  $W^0$  se extiende de  $(C, C)$  a  $(D, D)$ . Sea  $X^0$  un elemento de  $D$  con esta medida extendida como su distribución. Se demostrará primero que, si las  $T_n$  son uniformemente distribuidas en  $[0,1]$ , entonces  $X_n \xrightarrow{D} X^0$ .

Sea  $U_n(t, \omega)$  el número de puntos entre  $T_1(\omega), \dots, T_n(\omega)$  que no exceden a  $t$ . Para  $0 = t_0 < t_1 < \dots < t_k = 1$ , las variables aleatorias  $U_n(t_i) - U_n(t_{i-1})$ ,  $i = 1, \dots, k$ , tienen distribución conjunta multinomial con parámetros  $n$  y  $p_i = t_i - t_{i-1}$ , y por el Teorema Central del Límite para ensayos multinomiales, teorema 1.1.1, que establece que las distribuciones finito-dimensionales de  $X_n$  convergen débilmente a aquellas de  $X^0$ . Por el teorema 1.4.3 es suficiente probar que

$$E\{|X_n(t) - X_n(t_1)|^2 \cdot |X_n(t_2) - X_n(t)|^2\} \leq 6(t - t_1)(t_2 - t) \quad (1.4.4)$$

para  $t_1 \leq t \leq t_2$ , lo cual se hace a continuación.

Para  $1 \leq i \leq n$  sea  $\alpha_i$  igual a  $1 - (t - t_1)$  ó  $-(t - t_1)$  dependiendo si  $T_i$  cae dentro del intervalo  $(t_1, t]$  o no, y sea  $\beta_i$  igual a  $1 - (t_2 - t)$  ó  $-(t_2 - t)$  dependiendo si  $T_i$  se encuentra en el intervalo  $(t, t_2]$  o no. Entonces 1.4.4 es equivalente a

$$E\left\{\left(\sum_{i=1}^n \alpha_i\right)^2 \left(\sum_{i=1}^n \beta_i\right)^2\right\} \leq 6n^2 p_1 p_2. \quad (1.4.5)$$

Como las  $T_i$  son independientes, también lo son los vectores aleatorios  $(\alpha_i, \beta_i)$ . Como las  $T_i$  están uniformemente distribuidas,  $(\alpha_i, \beta_i)$  toman sus valores en los intervalos  $(1 - (t - t_1), -(t_2 - t))$ ,  $(-(t - t_1), 1 - (t_2 - t))$  y  $(-(t - t_1), -(t_2 - t))$  con probabilidades  $p_1 = t - t_1$ ,  $p_2 = t_2 - t$  y  $p_3 = 1 - p_1 - p_2$  respectivamente. Ahora, como  $E\{\alpha_i\} = E\{\beta_i\} = 0$ , y por consideraciones de simetría se tiene que

$$E\left\{\left(\sum_{i=1}^n \alpha_i\right)^2 \left(\sum_{i=1}^n \beta_i\right)^2\right\} = nE\{\alpha_1^2 \beta_1^2\} + n(n-1)E\{\alpha_1^2\}E\{\beta_2^2\} + 2n(n-1)E\{\alpha_1 \beta_1\}E\{\alpha_2 \beta_2\}.$$

Ahora, 1.4.5 se sigue de

$$\begin{aligned}E\{\alpha_1^2\beta_1^2\} &= p_1(1-p_1)^2p_2^2 + p_2p_1^2(1-p_2)^2 + p_3p_1^2p_2^2 \leq 3p_1p_2, \\E\{\alpha_1^2\}E\{\beta_2^2\} &= p_1(1-p_1)p_2(1-p_2) \leq p_1p_2, \\E\{\alpha_1\beta_1\}E\{\alpha_2\beta_2\} &= p_1^2p_2^2 \leq p_1p_2.\end{aligned}$$

De esta manera  $X_n \xrightarrow{D} X^0$ .

Ahora, el pedir que las variables aleatorias sean uniformemente distribuidas no es una pérdida de generalidad, ya que para funciones de distribución continuas, siempre podemos hacer uso del Teorema de la Transformación con la Integral de Probabilidad, el cual hace válido el teorema 1.4.3 independientemente de la distribución de las  $T$ 's.

## Capítulo 2

# EL PROBLEMA DE BONDAD DE AJUSTE

En el presente capítulo se introduce el problema clásico de bondad de ajuste y se describen algunas de las estadísticas más utilizadas. De entre ellas se destacan las denominadas integrales cuadráticas por haber resultado, en la práctica, en procedimientos que llevan a pruebas de hipótesis muy potentes.

Debido a que las inferencias estadísticas en el problema de bondad de ajuste clásico se plantean en un contexto no-paramétrico, estas deben basarse en las llamadas estadísticas de orden; y debido a que las mencionadas estadísticas de orden son una función uno a uno de la función de distribución empírica, se introduce esta última como una representación muestral para el problema.

### 2.1 EL PROBLEMA DE BONDAD DE AJUSTE

Considere una muestra aleatoria  $X_1, \dots, X_n$  de la cual sólo se supone que proviene de una función de distribución  $s_0 \in \mathcal{F}$ .

Sobre la clase  $\mathcal{F}$  se sabe muy poco, típicamente que  $\mathcal{F}$  =clase de todas las continuas (caso continuo), o bien,  $\mathcal{F}$  es la clase de todas las discretas (caso discreto).

El problema de inferencia que se plantea en el contexto anterior es que con base en la muestra, se "pruebe" la hipótesis nula

$$H_0 : F_0 \in \mathcal{F}$$

siendo  $\mathcal{F}$  una familia paramétrica  $\mathcal{F} = \{F(\cdot; \theta); \theta \in \Theta\}$  o bien  $\mathcal{F}$  es una familia con un solo elemento  $\mathcal{F} = \{F_0(\cdot)\}$

Cuando  $\mathcal{F}$  es una familia paramétrica (por ejemplo las normales con media y varianza desconocida), el problema de bondad de ajuste se conoce como compuesto. Si  $\mathcal{F}$  consta de un solo elemento, el problema se conoce como simple.

## 2.2 LA FUNCION DE DISTRIBUCION EMPIRICA.

Considérese nuevamente la FDE definida en el capítulo anterior. La función  $F_n(x)$  es escalonada; cuando  $x$  se incrementa y rebaza cada una de las  $X_{(i)}$ , la función da un salto de tamaño  $\frac{1}{n}$ . Para cualquier  $x$ ,  $F_n(x)$  calcula la proporción de observaciones menores o iguales a  $x$ . La función  $F_n(x)$  estima a la verdadera  $F_0(x)$  que produjo a la muestra, y es de hecho un estimador consistente de  $F_0(x)$  pues cuando  $n \rightarrow \infty$ ,  $|F_n(x) - F_0(x)|$  decrece a cero con probabilidad uno.

Un resultado más fuerte es cierto, el lema de Glivenko Cantelli que muestra que la diferencia en la norma del supremo converge a cero con probabilidad uno, i.e.,  $\sup_x |F_n(x) - F_0(x)|$  converge a 0 con probabilidad 1.

Es claro que si se tuviera en mente una hipótesis simple  $H_0 : F = F_0$ , una discrepancia "grande" entre  $F_n$  y  $F_0$  hará dudar sobre la veracidad de  $H_0$  en el problema de bondad de ajuste correspondiente.

Más adelante se intentará dar una descripción de las distintas medidas de discrepancia y de ciertas estadísticas para decidir qué tan grande es "grande".

## 2.3 ESTADISTICAS BASADAS EN LA FDE.

Para este trabajo se analizará exclusivamente el caso continuo, esto es, cuando  $\mathcal{F}$  es la clase de todas las distribuciones continuas.

Una estadística que mide la diferencia entre  $F_n(x)$  y  $F_0(x)$  será llamada una estadística basada en la FDE. Las más importantes están basadas en las diferencias verticales entre  $F_n(x)$  y  $F_0(x)$  y se dividen en dos clases: la clase del supremo y la clase de integrales cuadráticas.

Las estadísticas del supremo son  $D^+$ ,  $D^-$  (y  $D$ ) y son respectivamente la mayor diferencia vertical cuando  $F_n(x)$  es mayor que  $F_0(x)$  y la mayor diferencia vertical cuando  $F_n(x)$  es menor que  $F_0(x)$ . Formalmente,  $D^+ = \sup_x \{F_n(x) - F_0(x)\}$  y

$D^- = \sup_x \{F_0(x) - F_n(x)\}$ . La estadística más conocida es sin embargo,  $D$  introducida por Kolmogorov:

$$D = \sup_x |F_0(x) - F_n(x)| = \max(D^+, D^-)$$

Las estadísticas cuya forma es de integral cuadrática constituyen una amplia clase de medidas de discrepancia y están dadas por la familia de Cramér-von Mises

$$Q = n \int_{-\infty}^{\infty} \{F_n(x) - F_0(x)\}^2 \psi(x) dF_0(x)$$

Donde  $\psi(x)$  es una función "adecuada" que da los pesos a las diferencias cuadráticas  $\{F_n(x) - F_0(x)\}^2$ . Cuando  $\psi(x) = 1$  es la estadística conocida como de Cramér-von Mises, denotada por  $W^2$ , y cuando  $\psi(x) = \frac{F_0(x)}{1-F_0(x)}$  es la estadística de Anderson-Darling denotada por  $A^2$ .

De las definiciones básicas de las estadísticas del supremo y las de integrales cuadráticas descritas arriba se debe contar con una fórmula explícita. Esto se hace usando el llamado Teorema de la Transformación con la Integral de Probabilidad, PIT (de Probability Integral Transformation):

Cuando  $F_0(x)$  es la verdadera distribución de  $X$ , la variable aleatoria  $T = F_0(X)$  tiene distribución uniforme en el intervalo  $[0,1]$ .

Se denota a la función de distribución uniforme en el  $[0,1]$  por

$$G_0(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ t & \text{si } t \in [0, 1] \\ 1 & \text{si } t \geq 1 \end{cases}$$

Obsérvese que con las variables aleatorias transformadas  $T_i = F_0(X_i)$ , se puede construir su FDE, denotada por  $G_n(t)$ . De las estadísticas FDE algunas, (todas las aquí mencionadas), pueden ser calculadas a partir de  $G_n(t)$  y  $G_0(t)$  o a partir de  $F_n$  y  $F_0$ , indistintamente. Es fácil probar que para valores de  $t$  y  $x$  relacionados por  $t = F_0(x)$ , las diferencias verticales correspondientes en la FDE para las  $X_i$  y para las  $T_i$  son iguales, esto es

$$F_n(x) - F_0(x) = G_n(z) - G_0(t) = G_n(t) - t$$

En consecuencia las estadísticas calculadas con la FDE de las  $T_i$  tomarán los mismos valores como si hubieran sido calculadas de la FDE de las  $X_i$ . Esto lleva a las siguientes fórmulas de cálculo, las cuales incluyen a los valores de  $T$  arreglados en orden

ascendente,  $T_{(1)} < \dots < T_{(n)}$ .

$$\bar{T} = \sum_i \frac{T_i}{n}$$

$$D^+ = \max_i \left\{ \frac{i}{n} - T_{(i)} \right\} \quad (2.3.1)$$

$$D^- = \max_i \left\{ T_{(i)} - \frac{i-1}{n} \right\} \quad (2.3.2)$$

$$D = \max(D^+, D^-) \quad (2.3.3)$$

$$W^2 = \sum_i \left\{ T_{(i)} - \frac{2i-1}{2n} \right\}^2 + \frac{1}{12n} \quad (2.3.4)$$

$$A^2 = -n - \frac{1}{n} \sum_i (2i-1) \left\{ \log T_{(i)} + \log [1 - T_{(n+1-i)}] \right\} \quad (2.3.5)$$

donde  $\log$  significa logaritmo natural y las sumas y los máximos son sobre  $1 \leq i \leq n$ . Todas estas fórmulas son fácilmente reconstruidas con cálculo elemental ver [3] pag. 101. La hipótesis general para el problema de bondad de ajuste puede reescribirse de la siguiente manera:

$$H_0 : \text{la muestra aleatoria proviene de } F_0(x; \theta)$$

donde  $F_0(x; \theta)$  es una función de distribución continua y  $\theta$  es el vector de parámetros que si es conocido, equivale al caso simple; este caso también es referido en literatura como el Caso 0. Entonces  $T_{(i)} = F_0(X_{(i)}; \theta)$  da un conjunto  $T_{(i)}$  que bajo  $H_0$  tiene la distribución de una muestra adecuada de variables uniformes e independientes y las expresiones 2.3.1 a 2.3.5 son usadas como estadísticas de prueba. Por otro lado cuando  $\theta$  es desconocido (total o parcialmente),  $F_0(x, \theta)$  puede ser considerada sólo como un miembro de una familia de distribuciones. Este es el caso compuesto, referido en literatura como caso 1, 2 ó 3, dependiendo del grado de desconocimiento sobre  $\theta$  (usualmente  $\theta$  es un vector y el grado de desconocimiento se refiere al número de componentes desconocidas). Una vez elegida una estadística FDE para probar, digamos una hipótesis nula  $H_0$  simple (caso 0), es crucial el conocer su distribución bajo  $H_0$ . Esto con el objeto de poder controlar el tamaño de la probabilidad de error tipo I, o sea, la probabilidad de rechazar  $H_0$  siendo ésta cierta.

Los valores "grandes" de cualquiera de estas estadísticas (denótelas por  $Z$ ), por su construcción (son discrepancias entre la evidencia " $F_n$ " y la hipótesis " $F_0$ "), debieran conducir al rechazo de  $H_0$  y es indispensable el conocer para cada valor  $z$ ,  $P\{Z > z | H_0\}$ . La necesidad de las distribuciones asintóticas bajo  $H_0$  para estas estadísticas es obvia. Una vez obtenidas sería aún más conveniente conocer la distribución de  $Z$  para  $n$  finito o tener al menos una idea de la aproximación que se tiene al utilizar las distribuciones asintóticas.

Como se vió en el capítulo anterior, la teoría asintótica se obtuvo, visualizando las estadísticas FDE como funcionales del proceso empírico

$$\sqrt{n}\{F_n(x) - F_0(x)\}_{x \in \mathbb{R}}$$

que, como ya se mencionó, pueden indistintamente escribirse en términos del proceso empírico obtenido después de transformar al  $[0,1]$

$$\sqrt{n}\{G_n(t) - t\}_{t \in [0,1]}$$

más aún, el interés primordial en este trabajo es el de ilustrar la teoría asintótica de las estadísticas FDE que son integrales cuadráticas.

Cuando  $\theta$  contiene uno o más parámetros desconocidos, éstos pueden ser reemplazados por estimadores, y tener  $\hat{\theta}$  como estimador de  $\theta$ . Entonces las fórmulas pueden seguir siendo usadas para calcular las estadísticas FDE con  $T_{(i)} = F_0(X_{(i)}; \hat{\theta})$ . Sin embargo, aún cuando  $H_0$  sea cierta, los  $T_{(i)}$  no serán una muestra ordenada uniforme, y las distribuciones de las estadísticas FDE serán diferentes de aquellas del caso 0; éstas dependerán de la distribución que se esté probando, posiblemente de los parámetros desconocidos y del método de estimación así como del tamaño de la muestra. Las tablas adecuadas se deberán usar para la prueba elegida, aún para muestras grandes, de otro modo se tendría un error muy serio en el nivel de significancia de la prueba. Cuando los componentes desconocidos de  $\theta$  son de localización y/o escala, y si estos son estimados por métodos apropiados, la distribución de las estadísticas FDE no dependerán de los verdaderos valores de los parámetros desconocidos. Entonces, los puntos de rechazo para tales distribuciones, por ejemplo, la normal, la exponencial, dependen solamente de la familia probada y del tamaño de muestra  $n$ . Aún así las distribuciones exactas de las estadísticas son muy difíciles de encontrar y, a excepción de la distribución exponencial, los estudios Monte Carlo han sido usados extensamente para encontrar



puntos de rechazo para  $n$  finita. Afortunadamente, para las estadísticas de integrales cuadráticas  $W^2$  y  $A^2$ , está disponible una teoría asintótica y aún más, los puntos de rechazo para estas estadísticas para  $n$  finita convergen rápidamente a los puntos de rechazo asintóticos (en otras palabras, las distribuciones asintóticas son una excelente aproximación a las exactas). Para una mejor ilustración véase [3] pag. 102.

Para las estadísticas  $D^+$ ,  $D^-$ ,  $D$  no hay una teoría asintótica general (excepto para el caso 0), y aún los puntos de rechazo asintóticos tienen que ser estimados. Esto puede hacerse graficando para una  $\alpha$  fija, (siendo  $\alpha$  la probabilidad de error tipo I), los puntos obtenidos por Monte Carlo para muestras de tamaño  $n$  contra  $m = \frac{1}{n}$  y entonces extrapolando a  $m = 0$ ; alternativamente, como las estadísticas son funciones de un proceso el cual es asintóticamente Gaussiano, los puntos asintóticos pueden ser encontrados analizandolas como funciones del proceso límite. Para las pruebas correspondientes a muchas familias, Stephens (1970, 1974b, 1977, 1979) ha hecho modificaciones a las estadísticas de prueba, para corregir por el hecho de que  $n$  es finito. Si la estadística es, digamos  $Z$ , la modificación es una función de  $n$  y  $Z$  es entonces referida a los puntos de rechazo asintóticos. Para mayores detalles del tipo de estimadores requeridos en cada caso compuesto, vea [3] pag. 103.

## 2.4 PRUEBAS FDE PARA UNA DISTRIBUCION TOTALMENTE ESPECIFICADA.

(CASO SIMPLE ó 0) El siguiente procedimiento es el utilizado para pruebas FDE en el Caso 0, es decir  $X_1, \dots, X_n$ , muestra aleatoria de  $F_0$  continua

$$H_0 : F_0(x) = F_0(x; \theta); \theta \text{ conocido}$$

- a) Obtenga las estadísticas de orden  $X_{(1)} < \dots < X_{(n)}$ .
- b) Calcule sus transformadas  $T_{(i)} = F_0(X_{(i)}; \theta)$ ,  $i = 1, \dots, n$   
(es igual transformar y ordenar que ordenar y transformar)
- c) Calcule la estadística de prueba elegida.
- d) Modifique la estadística de prueba usando las modificaciones para la cola derecha de la distribución y compare con el punto de rechazo apropiado para el porcentaje  $\alpha$  deseado. Si la estadística excede el valor en la cola derecha, al nivel  $\alpha$ ,  $H_0$  se rechaza a un nivel de la prueba  $\alpha$ .

Estadística	Forma Modificada	Nivel de significancia								
		$Z$	$Z^*$	.25	.15	.10	.05	.025	.01	.005
$D^+(D^-)$	$D^+(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}})$		0.828	0.973	1.073	1.224	1.358	1.518	1.628	1.859
$D$	$D(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}})$		1.019	1.138	1.224	1.358	1.480	1.628	1.731	1.950
$W^2$	$(W^2 - \frac{0.4}{n} + \frac{0.6}{n^2})(1 + \frac{1}{n})$		0.209	0.284	0.347	0.461	0.581	0.743	0.869	1.167
$A^2$	$A^2$ Para toda $n \geq 5$		1.248	1.610	1.933	2.492	3.070	3.857	4.500	6.000

Como nota, en la práctica de pruebas de hipótesis es usual que en lugar de trabajar con el nivel de la prueba (probabilidad de error tipo I) se trabaje con el concepto de significancia. Dicho concepto es el siguiente:

Suponga que una estadística de prueba  $Z$  toma el valor  $z$ ; el nivel de significancia o "p-value" de la estadística es el valor  $p = P(Z > z)$ .

## 2.5 LA POTENCIA DE LAS ESTADÍSTICAS.

Siendo el problema de bondad de ajuste un problema de pruebas de hipótesis, y siendo el criterio "de optimalidad" en pruebas de hipótesis el utilizar estadísticas que a un mismo nivel dos tengan potencia alta, esto es, que la probabilidad de rechazar  $H_0$  cuando ésta es falsa sea alta, las estadísticas FDE son usualmente mucho más potentes que la muy utilizada estadística  $\chi^2$  de Pearson; esto se explica del hecho de que para la estadística  $\chi^2$  los datos deben ser agrupados, perdiéndose información, sobre todo en muestras pequeñas.

De entre las FDE, la estadística más conocida es  $D$ , pero en general es mucho menos potente que las estadísticas cuadráticas  $W^2$  y  $A^2$  según se ha mostrado en numerosos estudios.

Es por esto que resultan de singular importancia las pruebas FDE de integrales cuadráticas.

## 2.6 DESCRIPCION DE ALGUNOS CASOS DE INTERES.

A manera de ilustración, se exponen tres casos de interés en bondad de ajuste; el caso normal, el exponencial y el de la distribución gamma. En cada caso se describe el procedimiento cuando uno o varios de los parámetros son desconocidos. Allí se muestra el papel esencial que juegan las distribuciones asintóticas. Sólo se presentan las tablas

asintóticas para  $A^2$  y  $W^2$  y las modificaciones para  $n$  finito, refiriéndose a [3] al lector para mayor información).

### 2.6.1 EJEMPLO NORMAL.

Caso 1: La varianza  $\sigma^2$  es conocida y  $\mu$  es desconocida y estimada por  $\bar{X}$ , la media muestral.

Caso 2: La media  $\mu$  es conocida y  $\sigma^2$  desconocida y estimada por

$$s_1^2 = \sum_i \frac{(X_i - \mu)^2}{n}$$

Caso 3: Tanto la media como la varianza son desconocidos y estimados por  $\bar{X}$  y

$$s^2 = \sum_i \frac{(X_i - \bar{X})^2}{n - 1}$$

respectivamente. De los tres, el caso 3 es el más importante y el de mayor aplicación.

Para los tres casos, los pasos a seguir al hacer la sustitución  $T_{(i)} = F_0(X_{(i)}; \theta)$  son:

a) Calcular  $w_i$  para  $i = 1, \dots, n$  con

$$\begin{cases} w_i = \frac{(X_{(i)} - \bar{X})}{\sigma} & \text{Caso 1} \\ w_i = \frac{(X_{(i)} - \mu)}{\sigma} & \text{Caso 2} \\ w_i = \frac{(X_{(i)} - \bar{X})}{s} & \text{Caso 3} \end{cases}$$

b) Calcular  $T_{(i)} = \Phi(w_i)$  ( $i = 1, \dots, n$ ) donde  $\Phi(x)$  denota la función de distribución evaluada en  $x$  de una normal estándar. La función  $\Phi(x)$  puede ser encontrada en las tablas o en rutinas de computadora.

c) Calcule las estadísticas de prueba.

d) Para los casos 1 ó 2, use las tablas disponibles, ya que para estos casos no se hacen modificaciones. Para el caso 3, calcule la estadística modificada y busque en la tabla correspondiente. Si el valor de la estadística usada excede el punto de rechazo apropiado al nivel  $\alpha$ ,  $H_0$  es rechazada con nivel de significancia  $\alpha$ .

Puntos de rechazo prueba de normalidad Casos 1 y 2

		Nivel de significancia							
Z		.25	.15	.10	.05	.025	.01	.005	.0025
$W^2$	Caso 1	.094	.117	.134	.165	.197	.238	.270	.302
$W^2$	Caso 2	.190	.263	.327	.442	.562	.725	.851	.978
$A^2$	Caso 1	.744	.782	.894	1.087	1.285	1.551	1.756	1.964
$A^2$	Caso 2	1.072	1.430	1.743	2.308	2.898	3.703	4.324	4.954

Puntos de rechazo prueba de normalidad Caso 3

		Nivel de significancia							
Z	Z*	.50	.25	.15	.10	.05	.025	.01	.005
W <sup>2</sup>	W <sup>2</sup> (1.0 + $\frac{0.5}{\sqrt{n}}$ )	.051	.074	.091	.104	.126	.148	.179	.205
A <sup>2</sup>	A <sup>2</sup> (1.0 + $\frac{0.75}{n} + \frac{2.25}{n^2}$ )	.341	.470	.561	.631	.752	.873	1.035	1.159

Los puntos de rechazo dados por las estadísticas  $W^2, A^2$ , son de sus distribuciones asintóticas y fueron encontrados analíticamente. Las modificaciones para todas las estadísticas fueron calculadas de puntos para  $n$  finita obtenidos por métodos de Monte Carlo.

## 2.6.2 EJEMPLO EXPONENCIAL.

La distribución exponencial  $Exp(\eta, \beta)$  es ([3] pag. 133)

$$F_0(x; \eta, \beta) = 1 - e^{-\frac{(x-\eta)}{\beta}} \quad x > \eta; \beta > 0$$

Se considera

$H_0$ : la muestra aleatoria  $X_1, \dots, X_n$  proviene de una distribución  $Exp(\eta, \beta)$

y se distinguen 3 casos:

Caso 1: el parámetro de localización  $\eta$  es desconocido pero  $\beta$  es conocido.

Caso 2: el parámetro de escala  $\beta$  es desconocido pero  $\eta$  es conocido.

Caso 3: ambos parámetros son desconocidos.

### PRUEBAS PARA EL CASO 1.

El primer método que se describirá para el Caso 1 utiliza la propiedad especial de una distribución exponencial como sigue. Sea

$$X'_{(i-1)} = X_{(i)} - X_{(1)}$$

para  $i = 2, \dots, n$ ; bajo  $H_0$ ,  $X'$  será la variable aleatoria proveniente de una  $Exp(0, \beta)$  y como  $\beta$  es conocido, el caso 0 puede ser usado en los  $n - 1$  valores de  $X'$ .

Un segundo método consiste en que  $\eta$  puede ser insesgadamente estimada por  $\hat{\eta} = (X_{(1)} - \frac{1}{n})$ ; este estimador se deriva del máximo verosímil  $X_{(1)}$  y tiene varianza tan pequeña como  $\frac{1}{n^2}$ . Entonces,  $T_{(i)}$  se calcula como  $T_{(i)} = 1 - exp[-(X_{(i)} - \hat{\eta})/\beta]$ ,  $i = 1, \dots, n$  y las estadísticas FDE calculadas de los  $T_{(i)}$  tendrán las distribuciones asintóticas del caso 0, así que se pueden usar las mismas tablas. Sin embargo, en contraste con

el procedimiento de prueba anterior, las modificaciones dadas ahí no pueden ser aplicadas y como los dos procedimientos son parecidos al tener potencia similar, el primer procedimiento es más recomendado para muestras relativamente pequeñas.

### PRUEBAS PARA EL CASO 2.

Para este caso supóngase primero que  $\eta = 0$ . El estimador máximo verosímil para  $\beta$  está dado por  $\hat{\beta} = \bar{X}$  donde  $\bar{X}$  es la media muestral.

Los pasos a seguir son:

- Calcule  $T_{(i)} = 1 - \exp(-X_{(i)}/\bar{X}), i = 1, \dots, n$
- Calcule las estadísticas FDE.
- Modifique y compare con los puntos de rechazo dados en las tablas correspondientes u obtenga los "p-values".

Si el origen  $\eta$  conocido es  $\eta = \eta_0$ , se puede hacer la sustitución

$$X'_i = X_i - \eta \quad i = 1, \dots, n$$

y los  $X'_i$  pueden ser usados para probar  $\text{Exp}(0, \beta)$ .

Los puntos de rechazo dados allí son los puntos asintóticos para las estadísticas  $W^2$  y  $A^2$ .

Puntos de rechazo prueba exponencial Caso 2

		Nivel de significancia							
Z	Z*	.25	.20	.15	.10	.05	.025	.01	.005
$W^2$	$W^2(1.0 + \frac{0.16}{n})$	.116	.130	.148	.175	.222	.271	.338	.390
$A^2$	$A^2(1.0 + \frac{0.6}{n})$	.736	.816	.916	1.062	1.321	1.591	1.959	2.244

### PRUEBAS PARA EL CASO 3.

Relativamente pocas pruebas han sido propuestas para el caso exponencial con  $\eta$  y  $\beta$  desconocidos, probablemente porque puede ser reducido a una prueba con  $\eta = 0$ . Sin embargo, esto no siempre es el mejor procedimiento. Uno alternativo es usar estimadores tanto para  $\eta$  como para  $\beta$ . La hipótesis nula es:

$H_0$ : la muestra aleatoria proviene de una distribución exponencial con  $\eta$  y  $\beta$  desconocidos.

Procedimiento:

- Calcule los estimadores  $\hat{\beta} = \frac{n(\bar{X} - X_{(1)})}{n-1}$  y  $\hat{\eta} = X_{(1)} - \frac{\hat{\beta}}{n}$ .
- Calcule  $W_i = \frac{X_{(i)} - \hat{\eta}}{\hat{\beta}}, i = 1, \dots, n$ .
- Calcule  $T_i = 1 - \exp(-W_i), i = 1, \dots, n$

d) Encuentre las estadísticas de prueba, modifique  $W^2$ ,  $A^2$  usando las tablas disponibles y compare con los puntos de rechazo asintóticos dados.

El estimador de  $\eta$  es supereficiente y la teoría asintótica es la misma que para el Caso 2. Para  $n$  finita, sin embargo, las distribuciones son diferentes. Las modificaciones para  $W^2$ ,  $A^2$  fueron encontrados por métodos de Monte Carlo.

Puntos de rechazo prueba exponencial Caso 3		Nivel de significancia						
$Z$	$Z^*$	.25	.15	.10	.05	.025	.01	
$W^2$	$W^2(1 + \frac{2.8}{n} - \frac{3}{n^2})$	.116	.148	.175	.222	.271	.338	
$A^2$	$A^2(1 + \frac{5.2}{n} - \frac{11}{n^2})$	.736	.916	1.062	1.321	1.591	1.959	

### 2.6.3 EJEMPLO GAMMA.

La hipótesis nula es ([3] pag. 151):

$H_0$ : la muestra aleatoria proviene de una distribución  $G(x; \eta, \beta, m)$  con densidad

$$f_0(x) = \frac{1}{\beta \Gamma(m)} \left( \frac{x - \eta}{\beta} \right)^{m-1} \exp\{-(x - \eta)/\beta\}, \quad x > \eta; \beta > 0$$

El parámetro de localización  $\eta$  se llamará el origen de la distribución y  $\beta$  y  $m$  son respectivamente, los parámetros de escala y de forma, respectivamente.

#### PRUEBAS CON ORIGEN CONOCIDO.

Si  $\eta = 0$  la hipótesis nula es:

$H_0$ : la muestra proviene de una distribución  $G(x; 0, \beta, m)$

Si  $\eta$  no es cero, pero tiene un valor  $\eta_0$ , se hace la transformación  $X'_i = X_i - \eta_0$ ,  $i = 1, \dots, n$ . Entonces la hipótesis nula  $H_0$  para el conjunto  $X$  se reduce a  $H_{0\eta}$  para el conjunto  $X'$ , y  $H_{0\eta}$  se prueba usando el conjunto  $X'$ .

Considerando  $H_{0\eta}$ , podemos distinguir 3 casos:

Caso 1:  $m$  conocida y  $\beta$  desconocida.

Caso 2:  $\beta$  conocida y  $m$  desconocida.

Caso 3: ambos parámetros desconocidos.

Para los casos 2 y 3, la teoría de distribución, inclusive teoría asintótica, cuando  $m$  es estimado por máxima verosimilitud u otro método eficiente, dependerá del verdadero valor de  $m$ ; esto es porque no es un parámetro de localización o escala. Sin embargo, pruebas aproximadas pueden hacerse como sigue:

### PRUEBA PARA EL CASO 1.

- a) Ordénese la muestra en forma ascendente  $X_{(1)}, \dots, X_{(n)}$ .  
b) Sea  $\bar{X}$  la media muestral y estime  $\hat{\beta} = \frac{\bar{X}}{m}$  estimador máximo verosímil.  
c) Defina

$$I(X; m, \beta) = \frac{1}{\beta^m \Gamma(m)} \int_0^X x^{m-1} \exp(-x/\beta) dx$$

Calcule  $Z_{(i)} = I(X_{(i)}; m, \hat{\beta})$ , para  $i = 1, \dots, n$ .

- d) Calcule las estadísticas FDE.  
e) Modifique las estadísticas como sigue:  
Para  $m = 1$  calcule

$$W^* = W^2 \left( 1 + \frac{0.16}{n} \right)$$

$$A^* = A^2 \left( 1 + \frac{0.6}{n} \right)$$

Para  $m \geq 2$  calcule

$$W^* = \frac{1.8nW^2 - 0.14}{1.8n - 1}$$

$$A^* = A^2 + \frac{1}{n} \left( 0.2 + \frac{0.3}{n} \right)$$

Las estadísticas modificadas se refieren a los puntos de rechazo en la cola derecha dados en tablas para el valor conocido apropiado de  $m$ . Estos puntos son asintóticos para las diferentes distribuciones.

Las modificaciones dadas arriba, están basadas en estudios de Monte Carlo para  $n$  finita.

### APLICACIONES A LA PRUEBA DE LA DISTRIBUCION $\chi^2$ .

La distribución Gamma con  $m = \frac{r}{2}$  y  $\beta = 2$  es la distribución  $\chi^2$  con  $r$  grados de libertad. Entonces, el caso 1 puede ser usado para probar si las observaciones  $X_i$  multiplicadas por una constante, vienen de una distribución  $\chi^2$  con  $r$  grados de libertad. Por ejemplo, se puede usar  $H_0$ :  $n$  varianzas muestrales independientes  $s_1^2, \dots, s_n^2$ , cada una calculada de una muestra de tamaño  $k$  provienen de poblaciones que son normales con la misma varianza  $\sigma^2$ .

### PRUEBA PARA EL CASO 2.

- a) Ordenar la muestra en forma ascendente.  
 b) Estimar  $m$  resolviendo la ecuación

$$\frac{\sum_i \log X_i}{n} = \psi(m) - \log \beta$$

donde  $\psi(m)$  es la función digamma  $\frac{d}{dm} \log \Gamma(m)$ ;  $\widehat{m}$  es el estimador máximo verosímil de  $m$ .

c) Calcule  $T_{(i)} = I(X_{(i)}; \widehat{m}, \beta)$ ,  $i = 1, \dots, n$

d) Calcule las estadísticas utilizando  $T_{(i)}$ .

e) Rechace  $H_0$ , si el valor de la estadística usada es mayor que el valor en las tablas para el nivel de significancia deseado y para  $\widehat{m}$  apropiada.

Puntos de rechazo prueba Gamma Caso 2								
Nivel de significancia								
Estadística	$m$	.25	.10	.05	.025	.01	.005	
$W^2$	1	.103	.150	.186	.223	.273	.311	
	2	.099	.143	.176	.210	.256	.291	
	3	.097	.140	.172	.205	.250	.283	
	4	.096	.138	.171	.203	.247	.280	
	5	.096	.138	.169	.202	.245	.278	
	6	.095	.137	.169	.201	.244	.276	
	8	.095	.136	.168	.200	.242	.275	
	10	.095	.136	.167	.199	.241	.274	
	12	.095	.136	.167	.199	.241	.273	
	15	.094	.135	.167	.198	.240	.272	
	20	.094	.135	.166	.198	.240	.272	
	$\infty$	.094	.134	.165	.197	.238	.270	
	$A^2$	1	.680	.956	1.170	1.390	1.687	1.916
		2	.661	.926	1.130	1.338	1.619	1.836
3		.655	.915	1.115	1.320	1.596	1.809	
4		.651	.909	1.108	1.310	1.584	1.795	
5		.649	.906	1.103	1.305	1.577	1.787	
6		.648	.904	1.101	1.301	1.572	1.781	
8		.646	.901	1.097	1.297	1.567	1.775	
10		.645	.899	1.095	1.294	1.563	1.771	
12		.644	.898	1.094	1.293	1.561	1.768	
15		.644	.897	1.092	1.291	1.559	1.766	
20		.643	.896	1.091	1.289	1.557	1.763	
$\infty$		.644	.894	1.087	1.285	1.551	1.756	



PRUEBA PARA EL CASO 3.

a) Estime  $m$  resolviendo la ecuación

$$\frac{\sum_i \log X_i}{n} - \log \bar{X} = \psi(m) - \log m$$

donde  $\psi(m)$  es la función digamma descrita arriba y estime  $\hat{\beta} = \frac{\bar{X}}{m}$ .

b) Calcule la estadística de prueba tomando  $T_{(i)} = I(X_{(i)}; \hat{m}, \hat{\beta})$ ,  $i = 1, \dots, n$ .

c) Rechace  $H_0$ , si el valor de la estadística usada es mayor que el valor en la tabla disponible para el nivel de significancia  $\alpha$  deseado y  $\hat{m}$  apropiada.

		Puntos de rechazo prueba Gamma Caso 3					
		Nivel de significancia					
Estadística	$m$	.25	.10	.05	.025	.01	.005
$W^2$	1	.079	.111	.136	.162	.196	.222
	2	.076	.107	.131	.155	.187	.211
	3	.075	.106	.129	.153	.184	.208
	4	.075	.105	.128	.152	.183	.207
	5	.075	.105	.128	.151	.182	.206
	6	.075	.105	.128	.151	.181	.205
	8	.074	.104	.127	.150	.181	.204
	10	.074	.104	.127	.150	.180	.204
	12	.074	.104	.127	.150	.180	.203
	15	.074	.104	.127	.149	.180	.203
	20	.074	.104	.126	.149	.180	.203
	$\infty$	.074	.104	.126	.148	.178	.201
$A^2$	1	.486	.657	.786	.917	1.092	1.227
	2	.477	.643	.768	.894	1.062	1.190
	3	.475	.639	.762	.886	1.052	1.178
	4	.473	.637	.759	.883	1.048	1.173
	5	.472	.635	.758	.881	1.045	1.170
	6	.472	.635	.757	.880	1.043	1.168
	8	.471	.634	.755	.878	1.041	1.165
	10	.471	.633	.754	.877	1.040	1.164
	12	.471	.633	.754	.876	1.039	1.163
	15	.470	.632	.754	.876	1.038	1.162
	20	.470	.632	.753	.875	1.037	1.161
	$\infty$	.670	.631	.752	.873	1.035	1.159

## 2.7 CONCLUSIONES DEL SEGUNDO CAPITULO.

Dados el primer y segundo capítulos, se cuenta ya con todos los elementos necesarios para resolver el problema de encontrar la distribución asintótica de las estadísticas  $A^2$  y  $W^2$ .

Retomando: en el problema de Bondad de Ajuste se tiene una muestra  $T_1, \dots, T_n$ , donde  $T_i = F_0(X_i)$ , independiente e idénticamente distribuidos pues tiene una distribución uniforme en el intervalo  $[0,1]$  y más aún, ésta no depende de la distribución original de la muestra.

A continuación se hace uso de la FDE, obtenida a partir de la muestra ordenada  $T_{(1)}, \dots, T_{(n)}$ , denotada por  $G_n(t, \omega)$ . De aquí, fue posible construir el proceso estocástico

$$X_n(t, \omega) = \sqrt{n}(G_n(t, \omega) - t) \quad (2.7.6)$$

$$0 \leq t \leq 1.$$

Por último, se vio que las distribuciones finito-dimensionales del proceso  $X_n$  convergen a aquellas del Browniano Atado en el espacio de las funciones continuas por la derecha con límite por la izquierda y con la topología de Skorohod. Entonces, por el Teorema 1.4.4 tenemos la convergencia débil del proceso estocástico  $X_n$  al Browniano Atado.

También se estableció en el capítulo anterior que la convergencia se mantenía si se aplica una función continua al proceso de interés. Ahora, las funcionales aplicadas al proceso  $X_n$  para obtener las estadísticas Cramer von-Mises y Anderson Darling son continuas en el intervalo  $[0,1]$  y  $(0,1)$  respectivamente, pues realizando el cambio de variable con la expresión 2.7.6, éstas estadísticas se ven así:

$$W^2 = \int_0^1 X_n^2(t, \omega) dt \quad (2.7.7)$$

$$A^2 = \int_c^{1-c} \frac{X_n^2(t)}{t(1-t)} dt \quad (2.7.8)$$

Por lo cual se concluye la obtención de las distribuciones asintóticas para  $A^2$  y  $W^2$  para el caso simple, ya que el compuesto se considera fuera del alcance de este trabajo.

## Capítulo 3

# CONVERGENCIA DEBIL DE FUNCIONALES INTEGRALES DE PROCESOS ESTOCASTICOS

En el capítulo 1 se ilustró la manera que tradicionalmente se ha utilizado para probar la convergencia del proceso estocástico  $X_n = \sqrt{n}(F_n(x) - F_0(x))$  al Browniano atado. Como pudo verse, la condición requerida es la de tensión (además de la convergencia de las distribuciones finito-dimensionales), y por ende, los espacios en los cuales los procesos  $\{X_n\}$  tienen sus trayectorias, deben ser separables y completos. También se vió que al aplicar una función continua a este proceso, se seguía teniendo la convergencia y de esta manera se consiguió tener una distribución asintótica para las estadísticas cuadráticas.

Sin embargo, las estadísticas que más interesan son la Cramér-Von Mises y la Anderson-Darling, por resultar en la práctica más potentes (ver capítulo 2). Por lo tanto, no interesa tanto la convergencia del proceso  $X_n$  en sí sino la convergencia de las funcionales  $f_\phi(x) = \int_0^1 \phi(t, x(t))\lambda(dt)$ , donde  $\phi(t, \cdot)$  es una función continua, en las que pueden traducirse ambas estadísticas véanse 2.7.8 y 2.7.7.

El objetivo de este trabajo es el de aplicar los resultados de la investigación presentada por Heinz Cremers y Dieter Kadelka del Departamento de Matemáticas de la Universidad de Karlsruhe, Alemania, en su artículo "On Weak Convergence of Integral Functionals of Stochastic Processes with Applications to Processes Taking Paths in  $L_p^{\mathbb{R}}$ ", véase [2], en el cual se estudian las condiciones mucho más débiles que tensión, que deben cumplirse para obtener la convergencia de estas funcionales y compararlas con el enfoque tradicional.

El resultado principal es el siguiente: Si una sucesión de procesos  $\{X_n : n \geq 1\}$  es débilmente tensa (en un sentido definido más adelante), y si además, las distribuciones finito-dimensionales del proceso convergen débilmente, entonces, las funcionales  $(f_{\phi_1}(X_n), \dots, f_{\phi_m}(X_n))$  convergen débilmente a  $(f_{\phi_1}(X_0), \dots, f_{\phi_m}(X_0))$  en donde  $X_0$  es un proceso cuyas distribuciones finito-dimensionales son el límite de las de la sucesión. A continuación se presentan condiciones necesarias y suficientes para tensión débil y se establecen aplicaciones de los resultados para el caso de procesos estocásticos en  $L_p^{\mathbb{R}}$ .

### 3.1 EL RESULTADO PRINCIPAL.

En lo sucesivo, se empleará la siguiente notación:

- $(\Omega, \mathcal{A}, P)$  un espacio de probabilidad;
- $(T, \mathcal{B}, \lambda)$  un espacio medible  $\sigma$ -finito;
- $S$  un espacio de Hausdorff completamente regular con  $\sigma$ -álgebra de Borel  $\mathcal{B}_S = \mathcal{B}(S)$
- $C(S)$   $[C_b(S)]$  para el espacio de funciones continuas [continuas y acotadas] con dominio en  $S$ ;
- $\mathcal{L}_P(\lambda) = \mathcal{L}_P(T, \mathcal{B}, \lambda)$  El espacio de funciones  $P$ -integrables valuadas en  $\mathbb{R}$  con la seminorma  $\|x\|_P = (\int |x(t)|^P \lambda(dt))^{1/P}$ ;
- $L_P(\lambda) = L_P(T, \mathcal{B}, \lambda)$  para el correspondiente espacio de Banach de clases de equivalencia de funciones módulo conjuntos  $\lambda$ -nulos;
- $\mathcal{L}_P^+ = \{x \in \mathcal{L}(\lambda) : x \geq 0\}$ ;
- $a^+ = \max\{0, a\}$  para  $a \in \mathbb{R}$ .

Ahora sea  $\Phi$  un conjunto de funciones  $\mathcal{B} \otimes \mathcal{B}_S$ -medibles  $\phi : T \times S \rightarrow \mathbb{R}$  con  $\phi(t, \cdot) \in C(S) \forall t \in T$ .

Sea también  $M$  un conjunto de funciones  $\mathcal{B} - \mathcal{B}_S$ -medibles  $x : T \rightarrow S$  tales que

$$\int |\phi(t, x(t))| \lambda(dt) < +\infty, \quad \phi \in \Phi. \quad (3.1.1)$$

Entonces defínase  $f_\phi : M \rightarrow \mathbb{R}$ ,  $\phi \in \Phi$  por

$$f_\phi(x) = \int \phi(t, x(t)) \lambda(dt), \quad x \in M. \quad (3.1.2)$$

Finalmente, sea  $\{X_n : n \geq 0\}$  una sucesión de funciones  $\mathcal{A} \otimes \mathcal{B} - \mathcal{B}_S$ -medible,  $X_n : \Omega \times T \rightarrow S$  tales que

$$X_n(\omega, \cdot) \in M, \quad \forall \omega \in \Omega, n \geq 0 \quad (3.1.3)$$

es decir, las  $X_n$  son procesos estocásticos con trayectorias en  $M$ . Note que para  $\phi \in \Phi$  el mapeo  $(\omega, t) \rightarrow X_n^\phi = \phi(t, X_n(\omega, t))$  es  $\mathcal{A} \otimes \mathcal{B}$ -medible y que por 3.1.1 y 3.1.3, para toda  $\omega \in \Omega$ ,  $\widehat{X}_n^\phi(\omega) = X_n^\phi(\omega, \cdot)$  es un elemento de  $L_1(\lambda)$  y  $\widehat{X}_n(\omega) = X_n(\omega, \cdot)$  o la trayectoria.

**Definición 3.1.1** *Las distribuciones finito-dimensionales de un proceso  $X_n$  convergen a aquellas de  $X_0$  casi seguramente (c.s.) si y sólo si existe  $T_0 \in \mathcal{B}$  con  $\lambda(T - T_0) = 0$  tal que  $\forall k \geq 1, t_1, \dots, t_k \in T_0$ :*

$$(X_n(\cdot, t_1), \dots, X_n(\cdot, t_k)) \Rightarrow (X_0(\cdot, t_1), \dots, X_0(\cdot, t_k)).$$

**Definición 3.1.2** *Un subconjunto  $K \subset L_1(\lambda)$  es débilmente relativamente compacto, si y sólo si, es uniformemente integrable, i.e., para cada  $\epsilon > 0, \exists h \in L_1^+(\lambda)$  tal que*

$$\sup_{x \in K} \int (|x| - h)^+ \lambda(dt) \leq \epsilon$$

donde

$$(|x| - h)^+ = |x| - \min\{|x|, h\} \cdot 2$$

**Definición 3.1.3** (a) *Una sucesión  $\{X_n : n \geq 1\}$  de procesos con trayectorias en  $L_1(\lambda)$  es débilmente tensa, si y sólo si, existe un conjunto  $K \subset L_1(\lambda)$  débilmente compacto tal que  $\inf_n P\{\widehat{X}_n(\omega) \in K\} > 1 - \epsilon$ .*

(b) *Una sucesión  $\{X_n, n \geq 1\}$  de procesos medibles con trayectorias en  $M$  es  $\Phi$ -débilmente tensa, si y sólo si para toda  $\phi \in \Phi$  la sucesión  $\{X_n^\phi : n \geq 1\}$  es débilmente tensa.*

El siguiente resultado se utiliza en la demostración del Teorema más importante. Remitimos al lector a [1] pag. 25 para su demostración:

**Lema 3.1.1** *Supóngase que  $Y_n, X_{1n}, X_{2n}, \dots$  tienen dominio común y son elementos aleatorios con valores en  $S$  un espacio métrico con métrica  $\rho$ , separable tal que para*

cada  $u$   $X_{u,n} \xrightarrow{D} X_u$  cuando  $n \rightarrow \infty$  y que  $X_u \xrightarrow{D} X$  cuando  $u \rightarrow \infty$ . Supóngase además que

$$\lim_{u \rightarrow \infty} \limsup_{n \rightarrow \infty} P \{ \rho(X_{u,n}, Y_n) \geq \epsilon \} = 0$$

para cada  $\epsilon > 0$ . Entonces  $Y_n \xrightarrow{D} X$  cuando  $n \rightarrow \infty$ .

**Teorema 3.1.1** Sea  $\{X_n : n \geq 0\}$  una sucesión de procesos estocásticos que satisfacen la condición 3.1.3. Si las distribuciones finito-dimensionales de  $X_n$  convergen débilmente a aquellas de  $X_0$  c.s., y si  $\{X_n : n \geq 1\}$  es  $\phi$ -débilmente tensa, entonces, para toda  $\phi_1, \dots, \phi_m \in \Phi$ ,

$$(f_{\phi_1}(X_n), \dots, f_{\phi_m}(X_n)) \Rightarrow (f_{\phi_1}(X_0), \dots, f_{\phi_m}(X_0))$$

*Demostración.* Sea  $\phi \in \Phi$  y  $h \in \mathcal{L}_1^+$ . Defínase  $\phi^h(t, s) = \max\{-h(t), \min\{h(t), \phi(t, s)\}\}$  y

$$\eta_n^h(\omega, \phi) = \int \phi^h(t, X_n(\omega, t)) \lambda(dt), \quad n \geq 0 \quad (3.1.4)$$

Claramente  $-||h||_1 \leq \eta_n^h \leq ||h||_1$ . Primero se demostrará que  $\eta_n^h(\cdot, \phi) \Rightarrow \eta_0^h(\cdot, \phi)$ , es decir, para  $g \in C_b(S)$

$$\int g(\eta_n^h(\omega, \phi)) P(d\omega) \longrightarrow \int g(\eta_0^h(\omega, \phi)) P(d\omega) \quad (3.1.5)$$

Como  $|\eta_n^h| \leq ||h||_1$  y los polinomios son densos en  $C([-||h||_1, ||h||_1])$ , sea, sin pérdida de generalidad  $g(u) = u^l$  para alguna  $l \geq 1$ . Por el teorema de Fubini se obtiene

$$\begin{aligned} & \int g(\eta_n^h(\omega, \phi)) P(d\omega) \\ &= \int (\eta_n^h(\omega, \phi))^l P(d\omega) \\ &= \Pi_{i=1}^l \left\{ \int \phi^h(t_i, X_n(\omega, t_i)) \lambda(dt_i) \right\} \\ &= \int \left[ \int \phi^h(t_1, X_n(\omega, t_1)) \lambda(dt_1) \cdots \int \phi^h(t_l, X_n(\omega, t_l)) \lambda(dt_l) \right] P(d\omega) \\ &= \int \cdots \int F_n(t_1, \dots, t_l) \lambda(dt_1) \cdots \lambda(dt_l), \end{aligned}$$

donde

$$F_n(t_1, \dots, t_l) = \int \phi^h(t_1, X_n(\omega, t_1)) \cdots \phi^h(t_l, X_n(\omega, t_l)) P(d\omega), \quad n \geq 0$$

Como la función  $(u_1, \dots, u_l) \rightarrow \phi^h(t_1, u_1) \cdots \phi^h(t_l, u_l)$  está en  $C_b(S^l)$  y es  $\otimes_1^l \mathcal{B}_S$ -medible para  $t_1, \dots, t_l \in T_0$  fijas, se tiene que  $F_n(t_1, \dots, t_l) \rightarrow F_0(t_1, \dots, t_l)$  por convergencia débil de las distribuciones finito-dimensionales de  $X_n$ . Como  $|F_n(t_1, \dots, t_l)| \leq h(t_1) \cdots h(t_l)$ ,

$$\int \cdots \int F_n(t_1, \dots, t_l) \lambda(dt_1) \cdots \lambda(dt_l) \rightarrow \int \cdots \int F_0(t_1, \dots, t_l) \lambda(dt_1) \cdots \lambda(dt_l)$$

por convergencia dominada, se obtiene 3.1.5. De igual manera para  $\phi_1, \dots, \phi_m \in \Phi$ :

$$(\eta_n^h(\cdot, \phi_1), \dots, \eta_n^h(\cdot, \phi_m)) \implies (\eta_0^h(\cdot, \phi_1), \dots, \eta_0^h(\cdot, \phi_m)) \quad (3.1.6)$$

Para  $\phi \in \Phi$  sea  $\eta_n(\omega, \phi) = \int \phi(X_n)(\omega) = \int \phi(t, X_n(\omega, t)) \lambda(dt)$ . Fijense  $\phi_1, \dots, \phi_m \in \Phi$ . Por  $\Phi$ -tensión débil y la definición 3.1.2, para toda  $N \geq 1$  existe un subconjunto de  $L_1(\lambda)$ ,  $K_N$  débilmente compacto tal que

$$\sup_{n \geq 1} P\{\omega \in \Omega : \widehat{X}_n^{\phi_i}(\omega) \in K_N, \quad i = 1, \dots, m\} \geq 1 - \frac{1}{N}; \quad (3.1.7)$$

y de esta manera una  $h_N \in \mathcal{L}_1^+(\lambda)$  tal que  $\sup_{x \in K_N} \int (|x| - h_N)^+ \lambda(dt) \leq \frac{1}{N}$ . Sin pérdida de generalidad,  $h_N$  puede escogerse tal que  $h_N \uparrow \infty$  para  $N \uparrow \infty$ . Por convergencia dominada  $\eta_0^{h_N}(\cdot, \phi_i) \rightarrow \eta_0(\cdot, \phi_i)$  puntualmente, de aquí en particular

$$(\eta_n^h(\cdot, \phi_1), \dots, \eta_n^h(\cdot, \phi_m)) \implies (\eta_0(\cdot, \phi_1), \dots, \eta_0(\cdot, \phi_m)) \quad (3.1.8)$$

Ahora, por el lema 3.1.1 y las condiciones 3.1.6 y 3.1.8 se obtiene

$$(\eta_n(\cdot, \phi_1), \dots, \eta_n(\cdot, \phi_m)) \implies (\eta_0(\cdot, \phi_1), \dots, \eta_0(\cdot, \phi_m))$$

pues se tiene que, para toda  $\epsilon > 0$ ,

$$\lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P\{\max_{1 \leq i \leq m} |\eta_n^{h_N}(\omega, \phi_i) - \eta_n(\omega, \phi_i)| \geq \epsilon\} = 0$$

Pero esta es una simple consecuencia de 3.1.7 y

$$\begin{aligned} |\eta_n^{h_N}(\omega, \phi_i) - \eta_n(\omega, \phi_i)| &\leq \int |\phi^{h_N}(t, X_n(\omega, t)) - \phi(t, X_n(\omega, t))| \lambda(dt) \\ &= \int (|\phi(t, X_n(\omega, t))| - h_N(t))^+ \lambda(dt), \end{aligned}$$

donde, para la última ecuación, se ha usado  $|\max\{-a, \min\{a, b\} - b| = (|b| - a)^+$  para  $b \in \mathbb{R}, a \geq 0$   $\square$ .

### 3.2 CONDICIONES NECESARIAS Y SUFICIENTES PARA TENSION $\Phi$ -DEBIL.

En la parte A de esta sección se presentan condiciones equivalentes para tensión  $\Phi$ -débil; en la parte B, se derivan dos condiciones suficientes las cuales se prueban fácilmente y se estudian sus aplicaciones.

A) Sea  $\phi \in \Phi$  y  $X_n, n \geq 0$  dadas como antes. Recuérdese la notación  $\{\widehat{X}_n^\phi(\omega)\}(t) = X_n(\omega, t) = \phi(t, X_n(\omega, t))$ .

**Lema 3.2.1** *Las siguientes condiciones son equivalentes:*

$$\{X_n : n \geq 1\} \text{ es } \Phi\text{-débilmente tensa}; \quad (3.2.9)$$

$$\forall \epsilon > 0, \phi \in \Phi : \inf_{h \in \mathcal{L}_1^+(\lambda)} \limsup_{n \rightarrow \infty} P \left\{ \int (|\widehat{X}_n^\phi(\omega)| - h)^+ \lambda(dt) \geq \epsilon \right\} = 0 \quad (3.2.10)$$

$$\forall \epsilon > 0, \phi \in \Phi : \inf_{h \in \mathcal{L}_1^+(\lambda)} \limsup_{n \rightarrow \infty} P \left\{ \int_{\{|\widehat{X}_n^\phi(\omega)| \geq h\}} |\widehat{X}_n^\phi(\omega)| \lambda(dt) \geq \epsilon \right\} = 0 \quad (3.2.11)$$

$$\begin{aligned} \forall \epsilon > 0, \phi \in \Phi : \lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left\{ \int (|\widehat{X}_n^\phi(\omega)| - N)^+ \lambda(dt) \geq \epsilon \right\} = 0 \text{ y} \\ \inf_{B \in \mathcal{B}, \lambda(B) < \infty} \limsup_{n \rightarrow \infty} P \left\{ \int_{T-B} |\widehat{X}_n^\phi(\omega)| \lambda(dt) \geq \epsilon \right\} = 0 \end{aligned} \quad (3.2.12)$$

**Demostración.** Recuérdese por las definiciones 3.1.2 y 3.1.3, que  $\{X_n : n \geq 1\}$  es  $\Phi$ -débilmente tensa si y sólo si para toda  $\eta > 0$  y  $\phi \in \Phi$  existe  $K \subset L_1(\lambda)$  tal que para toda  $\epsilon > 0$  existe  $h \in \mathcal{L}_1^+(\lambda)$  tal que  $\sup_{x \in K} \int (|x| - h)^+ \lambda(dt) < \epsilon$  y  $\inf_{n \geq 1} P(\{\widehat{X}_n^\phi(\omega) \in K\}) \geq 1 - \eta$ .

3.2.9  $\Rightarrow$  3.2.10: Supóngase 3.2.9 y sea  $\epsilon > 0, \eta > 0$  y  $\phi \in \Phi$  dadas. Primero escójase  $K$  y luego  $h$  como se describieron arriba. Entonces:

$$\sup_n P \left\{ \int (|\widehat{X}_n^\phi(\omega)| - h)^+ \lambda(dt) \geq \epsilon \right\} \leq \sup_n P \{\widehat{X}_n^\phi(\omega) \notin K\} \leq \eta$$



y 3.2.10 se cumple, aún con sup en lugar de limsup.

3.2.10  $\Rightarrow$  3.2.9: Primero se demostrará que limsup en 3.2.10 puede reemplazarse con sup. Para  $\epsilon > 0$ ,  $\phi \in \Phi$  fijas, sea

$$a_n(h) = P \left\{ \int \left( |\widehat{X}_n^\phi(\omega)| - h \right)^+ \lambda(dt) \geq \epsilon \right\}.$$

Entonces por 3.2.10  $\limsup a_n(h) \downarrow 0$  cuando  $h$  crece a infinito y por la definición 3.1.2,  $a_n(h) \downarrow 0$  para cada  $n \geq 1$ . De aquí que  $\sup_n a_n(h) \downarrow 0$ . Ahora, sean dados  $\epsilon > 0$ ,  $\eta > 0$ ,  $\phi \in \Phi$ . Para cada  $k \geq 1$  escoja  $h_k \in L_1^+(\lambda)$  tal que  $\sup_n [\widehat{X}_n^\phi(P)](B_k) \geq 1 - \frac{\eta}{2k+1}$  donde  $B_k = \{x \in L_1(\lambda) : \| |x| - h_k \|_1 \leq \frac{1}{k}\}$ . 3.2.9 queda demostrada para  $K = \bigcap_1^\infty B_k$  y  $h_{k_0}$  con  $\frac{1}{k_0} \leq \epsilon$ .

3.2.10  $\Leftrightarrow$  3.2.11: Nótese que para toda  $a, b \geq 0$  se tiene que:  $aI_{[a,b]} \leq 2(a - \frac{b}{2})^+$ ; de aquí que

$$\int \left( |\widehat{X}_n^\phi(\omega)| - h \right)^+ \lambda(dt) \leq \int_{\{|\widehat{X}_n^\phi(\omega)| \geq h\}} |\widehat{X}_n^\phi(\omega)| \lambda(dt) \leq 2 \int \left( |\widehat{X}_n^\phi(\omega)| - h/2 \right)^+ \lambda(dt)$$

y entonces se tiene la equivalencia.

3.2.12  $\Rightarrow$  3.2.10: Sea  $h = NI_B$  para una cierta  $N \geq 1$  y  $B \in \mathcal{B}$  con  $\lambda(B) < \infty$ .

3.2.10  $\Rightarrow$  3.2.12: Supóngase 3.2.10. Entonces la primera relación se sigue de

$$\int \left( |\widehat{X}_n^\phi(\omega)| - N \right)^+ \lambda(dt) \leq \int \left( |\widehat{X}_n^\phi(\omega)| - h \right)^+ \lambda(dt) + \int (h - N)^+ \lambda(dt)$$

y la segunda de

$$\int_{T-B} |\widehat{X}_n^\phi(\omega)| \lambda(dt) \leq \int \left( |\widehat{X}_n^\phi(\omega)| - h \right)^+ \lambda(dt) + \int_{T-B} h \lambda(dt) \quad \square$$

Observación. La primera condición de 3.2.12 se satisface si  $(\omega, t) \rightarrow \phi(t, X_n(\omega, t))$ ,  $n \geq 1$  es uniformemente acotada y la segunda siempre se cumple para  $\lambda$  medida acotada.

B) Las siguientes dos condiciones se establecen en términos de la medida producto  $P \otimes \lambda$ .

**Lema 3.2.2**  $\{X_n : n \geq 1\}$  es  $\Phi$ -débilmente tenso si

$$\forall \phi \in \Phi : \inf_{h \in C_1^+(P \otimes \lambda)} \limsup_{n \rightarrow \infty} \int \left( |\widehat{X}_n^\phi| - h \right)^+ dP \otimes \lambda = 0 \quad (3.2.13)$$

Demostración. Sea  $g \in \mathcal{L}_1^+(\lambda)$  y  $h \in \mathcal{L}_1^+(P \otimes \lambda)$ . Sea  $\tilde{g}(\omega, t) = g(t)$ . De la desigualdad de Kolmogorov-Markov se tiene que

$$\begin{aligned} P \left\{ \int (|\widehat{X}_n^\phi(\omega) - g|^+ \lambda(dt) \geq \epsilon \right\} &\leq \frac{1}{\epsilon} \int (|X_n^\phi - \tilde{g}|^+ dP \otimes \lambda \\ &\leq \frac{1}{\epsilon} \left[ \int (|X_n^\phi(\omega) - h|^+ dP \otimes \lambda + \int (h - \tilde{g})^+ dP \otimes \lambda \right] \end{aligned}$$

ahora 3.2.10 se sigue de 3.2.13. Nótese que dada  $h \in \mathcal{L}_1^+(P \otimes \lambda)$ ,  $\int (h - \tilde{g})^+ dP \otimes \lambda$  puede ser tan pequeña como se quiera (escójase cualquier  $g > 0$  en  $\mathcal{L}_1^+(\lambda)$  y entonces aplíquese el teorema de convergencia monótona a  $\int (h - n\tilde{g})^+ dP \otimes \lambda$ ).  $\square$

Observación. La condición 3.2.13 se satisface si para toda  $\phi \in \Phi$  la sucesión  $\{X_n^\phi : n \geq 1\}$  es uniformemente integrable con la medida  $dP \otimes \lambda$ .

**Lema 3.2.3** *Supóngase que las distribuciones finito-dimensionales de  $X_n$  convergen débilmente a aquellas de  $X_0$  c.s.. Entonces  $\{X_n : n \geq 1\}$  es  $\Phi$ -débilmente tenso, sólo si*

$$\forall \phi \in \Phi : \limsup_{n \rightarrow \infty} \int |X_n^\phi| dP \otimes \lambda \leq \int |X_0^\phi| dP \otimes \lambda < +\infty \quad (3.2.14)$$

Demostración. Se verificará primero la condición 3.2.13 del Lema 3.2.2. Nótese que para  $\phi \in \Phi$

$$\inf_{h \in \mathcal{L}_1^+(\lambda)} \int (|X_n^\phi - \tilde{h}|^+ dP \otimes \lambda = 0 \quad (3.2.15)$$

Sea  $h \in \mathcal{L}_1^+(\lambda)$ . Del teorema 3.1.1 tenemos que  $\eta_n^h(\cdot, |\phi|) \Rightarrow \eta_0^h(\cdot, |\phi|)$ . De aquí que ([1] teo 5.4 pag.32)

$$\int \eta_n^h(\omega, |\phi|) P(d\omega) \longrightarrow \int \eta_0^h(\omega, |\phi|) P(d\omega) \quad (3.2.16)$$

pues  $|\eta_0^h(\omega, |\phi|)| \leq \|h\|_1$ . Usando 3.2.14 y 3.2.16 y  $|a| = \min\{|a|, b\} + (|a| - b)^+$  para  $a \in \mathbb{R}, b \geq 0$  se obtiene que

$$\begin{aligned} \limsup_{n \rightarrow \infty} \int (|X_n^\phi - \tilde{h}|^+ dP \otimes \lambda &= \limsup_{n \rightarrow \infty} \int |X_n^\phi| dP \otimes \lambda - \lim_{n \rightarrow \infty} \int \eta_n^h(\cdot, |\phi|) dP \\ &\leq \int |X_0^\phi| dP \otimes \lambda - \int \eta_0^h(\cdot, |\phi|) dP \\ &= \int (|X_0^\phi| - \tilde{h})^+ dP \otimes \lambda \end{aligned}$$

se tiene que 3.2.13 se sigue de 3.2.15.  $\square$

## Capítulo 4

# SOLUCION AL PROBLEMA DE BONDAD DE AJUSTE POR MEDIO DE CONVERGENCIA DE FUNCIONALES: UN NUEVO ENFOQUE

Una vez desarrollada esta nueva manera de probar la convergencia de las funcionales del proceso estocástico, el problema de Bondad de Ajuste puede ahora resolverse de un modo mucho más sencillo y eficiente y esto se muestra en el presente capítulo.

### 4.1 SOLUCION PARA LA DIFERENCIA CUADRÁTICA CRAMER-VON MISES.

Como se expuso en el capítulo 2, la diferencia cuadrática Cramer-Von Mises es

$$n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 dF_0(x).$$

Recuérdese que en el capítulo anterior se utilizó:

- $(\Omega, \mathcal{A}, P)$  un espacio de probabilidad. En este caso es el espacio de probabilidad correspondiente a la sucesión de variables aleatorias de interés, independientes e idénticamente distribuidas, definidas en los borelianos de los reales  $(\mathbb{R}, \mathcal{A})$ , con medida de probabilidad común  $P$ .
- $(T, \mathcal{B}, \lambda)$  un espacio medible  $\sigma$ -finito. En este caso sea  $T = [0, 1]$ ,  $\mathcal{B} =$  los borelianos

sobre el segmento  $[0,1]$  y  $\lambda$  la medida de Lebesgue sobre este segmento.

-  $S$  un espacio de Hausdorff completamente regular y  $\Phi$  un conjunto de funciones  $\mathcal{B} \otimes \mathcal{B}(S)$  medibles de la siguiente manera

$$\Phi = \{ \phi : T \times S \rightarrow \mathbb{R}, \text{ t.q. } \phi(t, \cdot) \in C(S) \forall t \in T \}.$$

Sea en nuestro caso particular  $\Phi$  una clase con una sola función  $\phi(t, y) = y^2$ . Entonces,  $\phi(t, \cdot) \in C(S) \forall t \in T$ .

-  $M$  un conjunto de funciones  $\mathcal{B} - \mathcal{B}(S)$  medibles definido de la siguiente manera:

$M = \{ x : T \rightarrow S \text{ t.q. } \forall \phi \in \Phi, \int |\phi(t, x(t))| d\lambda(t) < \infty \}$ . En este caso sea  $M = D[0, 1]$ , definido en el capítulo 1, el espacio de todas las funciones continuas por la derecha con límite por la izquierda definidas sobre el intervalo  $[0, 1]$ .

-  $\{X_n, n \geq 0\}$  una sucesión de funciones medibles  $X_n : \Omega \times T \rightarrow S$  tal que

$$X_n(\omega, \cdot) \in M, \omega \in \Omega, n \geq 0.$$

En este caso sea  $S = \mathbb{R}$  que es en particular un espacio de Hausdorff y completamente regular y  $\{X_n : \Omega \times T \rightarrow \mathbb{R}\}$ , la sucesión de procesos definidos por

$$X_n(t) = \sqrt{n}(G_n(t) - t)$$

donde  $t = F_0(x)$  y  $G_n(t)$  es la Función de Distribución Empírica obtenida después de aplicar esta transformación. De esta manera, por el Teorema de Cambio de Variable

$$\int_0^1 X_n^2(t) dt = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 dF_0(x)$$

Ahora, es necesario verificar que la sucesión propuesta efectivamente cumple que  $X_n(\omega, \cdot) \in M = D[0, 1]$ ,  $\forall \omega \in \Omega$ , i.e., es necesario demostrar que

$$\forall \phi \in \Phi, \int |\phi(t, X_n(\omega, t))| d\lambda(t) < \infty$$

o, en el caso particular presentado, como  $\Phi$  tiene el único elemento  $\phi(t, y) = y^2$ , que

$$\int_0^1 X_n^2(\omega, t) dt < \infty \tag{4.1.1}$$

lo cual se demuestra como sigue: se sabe que  $G_n(t) \in D[0, 1]$ ;  $s(t) = t \in D[0, 1]$  y por lo tanto  $\sqrt{n}(G_n(t) - t) = X_n(\omega, \cdot) \in D[0, 1]$ . Pero entonces  $X_n(\omega, \cdot)$  tiene a lo más un

número numerable de discontinuidades y por lo tanto 4.1.1 existe  $\forall \omega \in \Omega$ .

Se sabe que las distribuciones finito-dimensionales de  $X_n(t)$  convergen a las del puente Browniano, sea éste  $X_0(t)$  que es un proceso Gaussiano sobre  $[0,1]$  con  $E(X_0(t)) = 0$ ,  $E(X_0(t)X_0(s)) = t(1-s)$  para  $t < s$ .

Con lo anterior, ya tenemos algunos de los supuestos necesarios para el cumplimiento del teorema 3.1.1:

- $\Phi = \{\phi(t, x_t) = x_t^2\}; \phi(t, \cdot) \in C(\mathbb{R}) \forall t \in T$ .
- $X_n$  es una sucesión de procesos estocásticos la cual cumple  $X_n(\omega, \cdot) \in D[0,1]$  para esa  $\Phi$ .
- $X_n^\phi(t) = \phi(t, X_n(\omega, t)) = X_n^2(\omega, t)$
- Las distribuciones finito-dimensionales de  $X_n$  convergen débilmente a aquellas del Puente Browniano.

Por lo tanto, sólo resta verificar que la sucesión  $\{X_n\}$  es  $\Phi$ -débilmente tensa con  $\Phi$  dada anteriormente para poder concluir que

$$\int_0^1 X_n^2(t) dt \implies \int_0^1 X_0(t)^2 dt.$$

Dado que  $X_n(\omega, t)$  es una familia de variables aleatorias positivas se tiene que:

$$\begin{aligned} \int |X_n^\phi| dP \otimes \lambda &= \int_0^1 \int_\Omega |X_n^\phi(\omega, t)| P(d\omega) dt \\ &= \int_0^1 E [|X_n^\phi(t)|] dt, \end{aligned}$$

véase [4], p. 15 y 16.

Por lo tanto, la condición equivalente establecida por el lema 3.2.3 se puede reescribir así:

$$\forall \phi \in \Phi \quad \limsup_{n \rightarrow \infty} \int_0^1 E [|X_n^\phi(t)|] dt \leq \int_0^1 E [|X_0^\phi(t)|] dt \quad (4.1.2)$$

o en nuestro caso particular reescribir como:

$$\limsup_{n \rightarrow \infty} \int_0^1 E [X_n^2(t)] dt \leq \int_0^1 E [X_0^2(t)] dt$$

Y esta condición se cumple como se demuestra a continuación.

Recuérdese que  $X_n(t) = \sqrt{n}(G_n(t) - t)$ . Pero  $n \cdot G_n(t)$  tiene distribución binomial con parámetros  $(n, t)$  y por lo tanto, la esperanza de  $G_n(t)$  es  $t$ . De aquí que  $E[X_n(t)] = 0$  y  $E[X_n^2(t)] = t(1-t)$  que es la varianza. Por lo tanto,  $E[X_n^2(t)] = E[X_0^2(t)]$ .  $\square$

## 4.2 SOLUCION PARA LA DIFERENCIA CUADRÁTICA ANDERSON-DARLING.

Encontrar la distribución asintótica de la estadística Anderson-Darling definida por

$$n \int_{-\infty}^{\infty} \frac{[F_n(x) - F_0(x)]^2}{F_0(x)(1 - F_0(x))} dF_0(x),$$

no difiere en mucho de encontrarla para la Cramér-Von Mises. De hecho, sólo es un poco diferente en algunos detalles.

Nuevamente recuérdese que en el capítulo anterior se utilizó:

- $(\Omega, \mathcal{A}, P)$  un espacio de probabilidad. En este caso sea el espacio de probabilidad correspondiente a la sucesión de variables aleatorias de interés, independientes e idénticamente distribuidas, definidas en los borelianos de los reales  $(\mathbb{R}, \mathcal{A})$ , con medida de probabilidad común  $P$ .
- $(T, \mathcal{B}, \lambda)$  un espacio medible  $\sigma$ -finito. En este caso sea  $\epsilon > 0$  y  $T_\epsilon = (\epsilon, 1 - \epsilon)$ ,  $\mathcal{B}$  = los borelianos sobre el segmento  $T_\epsilon$  y  $\lambda$  la medida de Lebesgue sobre este segmento.
- $S$  es un espacio de Hausdorff, completamente regular y  $\Phi$  un conjunto de funciones  $\mathcal{B} \otimes \mathcal{B}(S)$  medibles de la siguiente manera

$$\Phi = \{ \phi : T \times S \rightarrow \mathbb{R}, \text{ t.q. } \phi(t, \cdot) \in C(S) \forall t \in T \}.$$

Sea en este caso  $\Phi$  un conjunto compuesto por una sola función  $\phi_1(t, x) = \frac{x^2}{t(1-t)}$ .

Entonces,  $\phi_1(t, \cdot) \in C(S) \forall t \in T_\epsilon$ .

-  $M$  un conjunto de funciones  $\mathcal{B} - \mathcal{B}(S)$  medibles definido de la siguiente manera:

$M = \{ x : T \rightarrow S \text{ t.q. } \forall \phi \in \Phi, \int |\phi(t, x(t))| d\lambda(t) < \infty \}$ . En este caso sea  $M = D(\epsilon, 1 - \epsilon)$ , el espacio de todas las funciones continuas por la derecha con límite por la izquierda definidas sobre  $(\epsilon, 1 - \epsilon)$ .

-  $\{X_n, n \geq 0\}$  una sucesión de funciones medibles  $X_n : \Omega \times T \rightarrow S$  tal que

$$X_n(\omega, \cdot) \in M, \omega \in \Omega, n \geq 0.$$

En este caso, sea de nueva cuenta  $S = \mathbb{R}$  un espacio de Hausdorff completamente regular y  $\{X_n : \Omega \times T \rightarrow \mathbb{R}\}$ , la sucesión de procesos definidos por

$$X_n(t) = \sqrt{n}(G_n(t) - t)$$

donde  $t = F_0(x)$  y  $G_n(t)$  es la Función de Distribución Empírica obtenida después de aplicar esta transformación. De esta manera, por el Teorema de Cambio de Variable

$$\int_{\epsilon}^{1-\epsilon} \frac{X_n^2(t)}{t(1-t)} dt = n \int_{L(\epsilon)}^{U(\epsilon)} \frac{[F_n(x) - F_0(x)]^2}{F_0(x)(1 - F_0(x))} dF_0(x)$$

donde  $F_0(U(\epsilon)) = 1 - \epsilon$  y  $F_0(L(\epsilon)) = \epsilon$ .

Ahora, nuevamente es necesario verificar que la sucesión propuesta efectivamente cumple que  $\forall \omega \in \Omega$ ,  $X_n(\omega, \cdot) \in M = D(\epsilon, 1 - \epsilon)$ , i.e., es necesario demostrar que

$$\forall \phi \in \Phi, \int |\phi(t, X_n(\omega, t))| d\lambda(t) < \infty$$

o, para el nuevo caso particular anterior, que

$$\int_{\epsilon}^{1-\epsilon} \frac{X_n^2(\omega, t)}{t(1-t)} dt < \infty$$

Sin embargo, esta integral existe, pues  $\frac{X_n^2(\omega, t)}{t(1-t)} \in D[0, 1]$  por las mismas razones que la anterior. Por lo tanto  $X_n \in D(\epsilon, 1 - \epsilon) \forall \omega \in \Omega$ .

Por ser el mismo caso que el anterior, las distribuciones finito-dimensionales de  $X_n(t)$  convergen a las del puente Browniano. Con todo esto, se tiene solamente el problema de verificar que  $X_n$  es  $\Phi$ -débilmente tensa, para nuestra nueva  $\Phi$  para poder hacer uso del teorema 3.1.1.

Dado que nuevamente  $X_n$  es una sucesión que toma valores positivos, por Meyer, Probability and Potentials, p. 15 y 16 y sustituyendo en cada caso, se tiene que la condición 4.1.2 es equivalente a

$$\limsup_{n \rightarrow \infty} \int_{\epsilon}^{1-\epsilon} E \left[ \frac{X_n^2(t)}{t(1-t)} \right] dt \leq \int_{\epsilon}^{1-\epsilon} E \left[ \frac{X_0^2(t)}{t(1-t)} \right] dt \quad (4.2.3)$$

Pero esta condición se cumple pues es fácilmente verificable que  $E \left[ \frac{X_n(t)}{\sqrt{t(1-t)}} \right] = E \left[ \frac{X_0(t)}{\sqrt{t(1-t)}} \right] =$

0 y que  $E \left[ \frac{X_n^2(t)}{t(1-t)} \right] = E \left[ \frac{X_0^2(t)}{t(1-t)} \right] = 1$ . Por lo tanto

$$\int_{\epsilon}^{1-\epsilon} \frac{X_n^2(t)}{t(1-t)} dt \implies \int_{\epsilon}^{1-\epsilon} \frac{X_0^2(t)}{t(1-t)} dt \quad \square.$$



Sin entrar en los aspectos finos de una demostración formal, se espera que si  $\epsilon \rightarrow 0$  y  $\int_0^1 \frac{X_n^2(t)}{t(1-t)} dt$  tiene una distribución bien definida, entonces

$$\int_0^1 \frac{X_n^2(t)}{t(1-t)} dt \Rightarrow \int_0^1 \frac{X_0^2(t)}{t(1-t)} dt.$$

### 4.3 CONCLUSIONES DEL CUARTO CAPITULO

Como pudo comprobarse, obtener la convergencia de funcionales integrales por el método propuesto es mucho más sencillo. Esto se debe a que conceptos aparentemente complicados son en realidad, en nuestro caso particular, reducidos a conceptos simples.

El capítulo 1 demuestra la convergencia del proceso estocástico y exige para ello condiciones que si bien se puede demostrar que se cumplen, se requiere de un proceso complicado y tedioso. Por otro lado, lo que en realidad interesa en el problema de Bondad de Ajuste es la convergencia de las funcionales representadas por las estadísticas (este trabajo comprendió únicamente la Cramer von-Mises y la Anderson Darling), pues es la que se necesita para encontrar la región de rechazo. Por lo tanto, de la manera tradicional es necesario hechar mano del teorema que comprueba que al aplicar una función continua al proceso, la convergencia se mantiene. El método propuesto se aboca al problema de encontrar la convergencia de las funcionales integrales logrando así resolver el problema de una manera más sencilla.

Hagamos aquí un pequeño análisis de los conceptos utilizados en el capítulo 3.

La medida  $(\Omega, \mathcal{A}, P)$  es el espacio de probabilidad en el que está definida la muestra aleatoria original.

$(T, \mathcal{B}, \lambda)$  un espacio medible  $\sigma$ -finito en el intervalo  $[0,1]$ , los borelianos en este segmento y  $\lambda$  la medida de Lebesgue.

El conjunto  $\Phi$  de funciones continuas compuesto por una sola, a saber:  $\phi(t, y) = y^2$  para  $W^2$  y  $\phi(t, y) = \frac{y^2}{t(1-t)}$  para  $A^2$ .

$M = D[0, 1]$  el cual es un espacio bastante conocido.

$X_n$  el proceso del cual hemos hablado a lo largo de este trabajo.

$S$  un espacio de Hausdorff completamente regular que no es sino los reales  $\mathbb{R}$ .

ESTA TESIS NO DEBE  
SALIR DE LA BIBLIOTECA

Por último, la demostración de que la sucesión es  $\Phi$ -débilmente tensa, la cual, dada la sencillez de todo lo anterior no es mas que el cálculo del segundo momento de una variable aleatoria con distribución binomial.

Por lo tanto, creo éste un método mucho más eficiente para encontrar la distribución asintótica de  $A^2$  y  $W^2$  resolviendo así el problema propuesto.

# Bibliografía

- [1] Patrick Billingsley *Convergence of Probability Measures*, Wiley, 1968
- [2] Cremers, Kadelka *On Weak Convergence of Integral Functionals of Stochastic Processes with Applications to Processes Taking Paths in  $L_p^{\mathbb{R}}$*
- [3] D'Agostino y Stephens *Goodness of Fitness Techniques*, Marcel de Ker Inc., 1986
- [4] Meyer *Probability and Potentials*
- [5] Boris Vladimirovich Gnedenko *The Theory of Probability*, Mir, 1969
- [6] William Feller *An Introduction to Probability Theory and its Applications*, Wiley, 1968
- [7] Paul R. Halmos *Measure Theory*, Van Nostrand Reinhold Co., 1950