73

2 ej

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

GENERALIZACIONES DE LOS TEOREMAS DE GLIVENKO-CANTELLI Y DONSKER.

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I O

P R E S E N T A :

A N D R E S P E D R O Z A

DIRECTOR DE TESIS:

DRA. MARIA ASUNCION BEGOÑA FERNANDEZ FERNANDEZ.

1998

VNIVERSDAD NACIONAL
AVPNºMA DE
MEXICO

M. en C. Virginia Abrín Batule
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
P r e s e n t e

Comunicamos a usted que hemos revisado el trabajo de Tesis:
Generalizaciones de los Teoremas de Glivenko-Cantelli y Donsker

realizado por Andrés Pedroza

con número de cuenta 9455495-5 , pasante de la carrera de actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	Dra. María Asunción Begoña Fernández Fernández	Begoña Fernández
Propietario	Dra. María Emilia Caballero Acosta	M E Caballero
Propietario	Dr. José María González Barrios	José Ma. González Barrios
Suplente	M. en C. José Antonio Gómez Ortega	
Suplente	Mat. Luis Alberto B[illegible]rre	

Consejo Departamental de Matemáticas
M. en A. P. Ma[illegible]ar Alonso Reyes

FACULTAD DE CIENCIAS
CONSEJO DEPARTAMENTAL
DE
MATEMATICAS

a mi madre

Agradezco a la Profa. María Emilia Caballero, al Prof. José María González, al Prof. José Antonio Gómez y al Prof. Luis Albero Briseño por las atenciones que me brindaron al revisar el presente trabajo así como las valiosas aportaciones y criticas hechas a éste. De una manera muy especial agradezco a mi asesora la Profa. Begoña Fernández, por la paciencia que mostró durante varios meses y sobre todo a los *jalones de orejas* que bien me merecía. Pido disculpas por las prisas a las que sometí a todo ellos y por los errores que se puedan encontrar en esta tesis.

Contenido

Introducción

Del enorme edificio de la Matemática podemos encontrar en varios de sus pisos la Teoría de la Probabilidad, que entre los diversos pilares con los que cuenta es posible localizar la Ley Fuerte de los Grandes Números, y desde luego, el Teorema del Límite Central. En el devenir del tiempo se han expuesto numerosas generalizaciones y extensiones de estos dos pilares. En este trabajo no nos vamos a quedar atras presentando generalizaciones de estos resultados, por medio de las funciones de *distribución empírica* que han dado pie a una nueva teoría, conocida como *Procesos Empíricos,* cuyos pioneros constructores son: R. Dudley y D. Pollard. Esta teoría se ha venido desarrollando fuertemente en los últimos treinta años, llegando a despertar interés con los vecinos de piso, por las múltiples aplicaciones que han aportado al área de Estadística.

Un factor importante en estas generalizaciones es la *distribución empírica* de n variables aleatorias $X_1, \ldots, X_n$ que H. Cramér definió en el año de 1928

$$F_n(t,\omega) = \frac{1}{n}\sum_{i=1}^{n} 1_{[0,t]}(X_i(\omega)).$$

La Ley Fuerte de los Grandes Números sostiene que si $X_1, \ldots, X_n$ son variables aleatorias independientes e idénticamente distribuídas tal que $EX_1 < \infty$, entonces

$$\frac{X_1 + \cdots + X_n}{n} \to EX_1$$

casi seguramente. Es en el año de 1933 cuando se presenta una generalización de este resultado, dada a conocer por V. Glivenko y F. P. Cantelli en trabajos independientes en la que intervienen las distribuciones empíricas, al establecer que para $X_1, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuídas

$$\sup_t |F_n(t,\omega) - F(t)| \to 0$$

con probabilidad 1, donde F es la distribución común de las variables aleatorias. Este resultado es conocido como el Teorema de Gliveko-Cantelli, y no

es otra cosa que una versión de la Ley Fuerte de los Grandes Núr
espacio de funciones.

Recientemente se han estudiado generalizaciones del Teorema d
Cantelli en las que se considera la *medida empírica*

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

donde δ_{X_i} es la *medida de Dirac* en X_i, y una familia de funcion
$\mathscr{F}$ definidas en $\mathbb{R}$, para preguntarse cuándo

$$\sup_{f\in\mathscr{F}} |\mathbb{P}_n f - E[f(X_1)]| \to 0$$

donde $\mathbb{P}_n f = n^{-1}\sum_{i=1}^{n} f(X_i)$, es decir, cuándo es posible obtener
uniforme de la Ley de los Grandes Números sobre la familia $\mathscr{F}$. Ot
esto es una generalización del Teorema de Glivenko-Cantelli si $\mathscr{F}$
de las funciones indicadoras del intervalo $[0,1]$. La convergencia ant
sujeta al *tamaño* de $\mathscr{F}$, en el sentido de cómo están dispersos lo
de $\mathscr{F}$, como un subconjunto de un espacio normado de funciones
el número de elementos con los que cuenta $\mathscr{F}$. Por tal motivo se
números de entropía que ayudarán a medir la clase $\mathscr{F}$.

El Teorema del Límite Central también habla sobre converge
riables aleatorias, pero de convergencia en *distribución*, al estab
$X_1, \ldots X_n$ son variables aleatorias independientes e idénticamente
con media cero y varianza 1 –lo referente a media y varianza es
ficar notación, lo indispensable es que las variables aleatorias teng
momento finito– entonces

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \to Y$$

en *distribución*, donde Y es una variable aleatoria con distribución
media cero y varianza 1.

El responsable de introducir las distribuciones empíricas en el T
Límite Central es M. D. Donsker, ya que en 1952 prueba la conv
distribución de

$$\sqrt{n}(F_n(t,\omega) - F(t)) \to Y$$

para variables aleatorias independientes con distribución común
es un elemento *Gaussiano*, llamado *el puente Browniano* o *Brown*

Las generalizaciones que se llevaran al cabo considerarán nuevamente una familia $\mathscr{F}$ de funciones, para poder determinar cuándo

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(X_i) - Ef(X_1))$$

converge a un elemento Gaussiano, esta convergencia dependera del tamaño de $\mathscr{F}$, por lo que sera necesario de nuevo el uso de los números de entropía para determinar condiciones necesarias sobre $\mathscr{F}$.

En el Capítulo 1 se trabajará con medidas de probabilidad con valores en la σ-álgebra de Borel de un espacio métrico arbitrario, para después introducir el término de *convergencia débil* de medidas, que nos dara la pauta para establecer la convergencia en distribución.

El Capítulo 2 es el estudio de dos casos particulares de espacios métricos, que son: el espacio de las funciones reales continuas con valores en $[0,1]$ y el espacio de las funciones reales continuas por la derecha con límite por la izquierda con valores en el intervalo $[0,1]$, en este capítulo se desarrollarán resultados para establecer convergencia débil y caracterizar a los conjuntos compactos. Se prueba la existencia de la Medida de Wiener en C, y el Principio de Invarianza, que cosnsiste en la aproximación a una variable aleatoria cuya distribución es la medida de Wiener. Para el caso del espacio D, sólo se extiende la Medida de Wiener, y se presenta el principio para este espacio.

Lo referente al Teorema de Glivenko-Cantelli se estudia en el Capítulo 3, empezando con el Teorema de Glivenko-Cantelli, para seguir con las desigualdades maximales, definiendo los números de entropía y la técnica de Simetrización, herramientas que serán de provecho al momento de probar las generalizaciones de los teoremas de interés.

Para dar fin a este trabajo, el Capítulo 4 corresponde al Teorema de Donsker, se tratan las versiones de dicho teorema para el caso de los dos espacios métricos estudiados en el capítulo 2, y después se presentan dos generalizaciones de este teorema en la que se involucran los números de entropía, para dar condiciones necesarias sobre la familia de funciones $\mathscr{F}$.

1 Convergencia de Medidas

Parte fundamental en el desarrollo de esta tesis es la convergencia débil de medidas de probabilidad, definidas en los conjuntos Boreleanos $\mathscr{B}(S)$ de un espacio métrico arbitrario (S, d). Se presenta la métrica de Prohorov, que es una métrica sobre el espacio de las medidas de probabilidad en $\mathscr{B}(S)$, y resulta que convergencia bajo la métrica de Prohorov equivale a convergencia débil de medidas; y se exponen equivalencias y condiciones necesarias para obtener convergencia débil de medidas de probabilidad. También aparece el concepto de medida tensa que desempeña un papel importante para determinar convergencia débil, y parte de ello se resume en el Teorema de Prohorov. Un resultado importante que también se trata es el Teorema de Representación de Skorohod, que da el enlace entre convergencia débil de medidas y convergencia casi segura de variables aleatorias.

A partir de la convergencia débil de medidas se define la convergencia en distribucón de variables aleatorias, término importante que aparece más adelante en los Principios de Invarianza como en los Teoremas de Donsker.

Con el concepto de convergecia débil de medidas, se tiene una topología en el espacio de medidas de probabilidad $\mathscr{P}(S)$, es por eso que se estudian las caracteristicas de completitud, separabilidad y compacidad, que no serán necesarias para el resto de este trabajo, pero no por ello dejan de ser importantes.

El contenido de este capítulo fue extraído de los libros Billingsley [1968], Ethier y Kurtz [1986], excepto por la ultima sección que es del libro Parthasaraty [1967].

1.1 La Métrica Prohorov

Denotaremos por (S, d) un espacio métrico, por $\mathscr{B}(S)$ la *σ- álgebra de Borel* de S, esto es, la σ-álgebra generada por los conjuntos abiertos de S; y por $\mathscr{P}(S)$ la familia de medidas de probabilidad en $\mathscr{B}(S)$. Para $F \in \mathscr{B}(S)$ se define:

$$F^{\varepsilon} = \{x \in S : \inf d(x, F) < \varepsilon\}$$

que también es un elemento de $\mathscr{B}(S)$ por ser un conjunto abierto.

Definición. *Una medida de probabilidad P en $\mathscr{B}(S)$ es regular, si dado $A \in \mathscr{B}(S)$ y $\varepsilon > 0$, existen conjuntos F cerrado y G abierto en S, tal que $F \subset A \subset G$ y $P(G \setminus F) < \varepsilon$.*

Esto es, si una medida P es regular, entonces P está totalmente derminada por los valores que toma en los conjuntos cerrados solamente, o bién en los conjuntos abiertos.

Teorema 1.1 *Toda medida de probabilidad en $\mathscr{B}(S)$ es regular.*

DEMOSTRACIÓN: Si A es cerrado, entonces tomamos $F = A$ y $G = \{x : d(x, A) < \delta\}$ con $\delta > 0$, de esta forma $G \to A$ cuando δ se aproxima a cero. Basta probar que la colección $\mathscr{G}$ de conjuntos que cumplen la condición de el teorema es la σálgebra de Borel de S.

Sean $A_n \in \mathscr{G}$, entonces existen conjuntos cerrados F_n y abiertos G_n tal que $F_n \subset A_n \subset G_n$ y $P(G_n \setminus F_n) < \varepsilon/2^{n+1}$. Sea n_0 de manera que $P(\cup_n F_n \setminus F) < \varepsilon/2$ donde $F = \cup_{n \le n_0} F_n$, sí $G = \cup_n G_n$, entonces $F \subset \cup_n A_n \subset G$ y $P(G \setminus F) \le P(G \setminus \cup_n F_n) + P(\cup_n F_n \setminus F) < \varepsilon$, por tanto $\mathscr{G}$ es cerrada bajo uniones numerables. Además $\mathscr{G}$ es cerrada bajo complementos, por lo cual, $\mathscr{G} = \mathscr{B}(S)$. □

Consideremos la función $\rho : \mathscr{P}(S) \times \mathscr{P}(S) \to \mathbb{R}$ definida como:

$$\rho(P, Q) = \inf\{\varepsilon > 0 : P(F) \le Q(F^{\varepsilon}) + \varepsilon \quad \text{para todo} \quad F \in \mathscr{C}\}$$

donde $\mathscr{C}$ es la colección de los subconjuntos cerrados de S, el Teorema 1.1 justifica el por qué es suficiente en considerar sólo la colección de los conjuntos cerrados de S en la definición de ρ.

Lema 1.1 *Sean P, $Q \in \mathscr{P}(S)$ y α, $\beta > 0$. Si*

$$P(F) \le Q(F^{\alpha}) + \beta$$

para todo $F \in \mathscr{C}$, entonces

$$Q(F) \le P(F^{\alpha}) + \beta.$$

Demostración: Sea $F_1 \in \mathscr{C}$ y $\alpha > 0$, consideremos a $F_2 = S - F_1^{\alpha}$, entonces $F_1 \subset S - F_2^{\alpha}$, de aquí se tiene que

$$P(F_1^{\alpha}) = 1 - P(F_2) \geq 1 - Q(F_2^{\alpha}) - \beta \geq Q(F_1) - \beta$$

por tanto $Q(F_1) \leq P(F_1^{\alpha}) + \beta$. □

Teorema 1.2 *$(\mathscr{P}(S), \rho)$ es un espacio métrico.*

Demostración: Por definición ρ es una función no negativa. Sean $P, Q \in \mathscr{P}(S)$ tal que $\rho(P, Q) = 0$, esto equivale a $P(F) = Q(F)$ para todo subconjunto F cerrado de S, entonces $P = Q$ por la regularidad de las medidas. Para demostrar la desigualdad del triángulo consideremos P, Q, R en $\mathscr{P}(S)$ de manera que $\rho(P, Q) < \varepsilon$ y $\rho(Q, R) < \delta$, para F cerrado se tiene

$$\begin{aligned} P(F) &\leq Q(F^{\varepsilon}) + \varepsilon \leq Q(\overline{F^{\varepsilon}}) + \varepsilon \\ &\leq R((\overline{F^{\varepsilon}})^{\delta}) + \varepsilon + \delta \leq R(F^{\varepsilon+\delta}) + \varepsilon + \delta \end{aligned}$$

es decir, $\rho(P, R) \leq \varepsilon + \delta$. La propiedad de simetría se obtiene del Lema 1.1. Por tanto ρ es una métrica en $\mathscr{P}(S)$. □

La métrica ρ es llamada *métrica de Prohorov*. Nos encaminamos a ver bajo que condiciones $\mathscr{P}(S)$ hereda las propiedades de completitud y separabilidad de S, así como también el Teorema de Representación de Skorohod, para ello necesitaremos los siguientes lemas. Diremos que μ es una *medida de Borel en S* cuando μ este definida en la σ-álgebra de Borel de S, esto es en $\mathscr{B}(S)$.

Lema 1.2 *Sea μ una medida positiva finita de Borel en S, $p_i \geq 0$ y $A_i \in \mathscr{B}(S)$ para $i = 1, \ldots, n$. Supongamos que*

$$\sum_{i \in I} p_i \leq \mu\left(\bigcup_{i \in I} A_i\right) \qquad \textit{para todo } I \subset \{1, \ldots, n\} \tag{1.1}$$

Entonces existen medidas positivas de Borel $\lambda_1, \ldots, \lambda_n$ en S tales que $\lambda_i(A_i) = \lambda_i(S) = p_i$ con $i = 1, \ldots, n$ y $\sum_{i=1}^{n} \lambda_i(A) \leq \mu(A)$ para toda $A \in \mathscr{B}(S)$.

Demostración: Sin pérdida de generalidad supongamos que p_i es positiva para toda i.

Por inducción sobre n; para el caso $n = 1$, definamos $\lambda_1(A) = p_1 \cdot \mu(A \cap A_1)/\mu(A_1)$ con $A \in \mathscr{B}(S)$, satisface que $\lambda_1(A_1) = \lambda_1(S) = p_1$; como $p_1 \leq \mu(A_1)$ se tiene $\lambda_1(A) \leq \mu(A)$ para toda $A \in \mathscr{B}(S)$, concluyendo el caso $n = 1$.

Supongamos la validez del lema para $1 \leq m \leq n-1$. Definamos $\eta(A) = \mu(A \cap A_n)/\mu(A_n)$, sobre $\mathscr{B}(S)$ y a ε_0 como:

$$\varepsilon_0 = \sup_{I \subset \{1,\dots,n-1\}} \left\{ \varepsilon : \sum_{i \in I} p_i \leq (\mu - \varepsilon\eta)\left(\bigcup_{i \in I} A_i\right)\right\} \tag{1.2}$$

[Caso 1: $\varepsilon_0 \geq p_n$] Sea $\lambda_n = p_n\eta$ y $\mu' = \mu - \lambda_n$, como $p_n \leq \mu(A_n)$, $\mu(A) \geq (p_n\mu(A \cap A_n))/\mu(A_n) = \lambda_n(A)$, esto es, μ' es positiva. Tomando $\varepsilon = p_n$ en (1.2), y la hipótesis de inducción, existen $\lambda_1, \dots, \lambda_{n-1}$ medidas positivas de Borel tales que $\lambda_i(A_i) = \lambda_i(S) = p_i$ con $1 \leq i \leq n-1$ y $\sum_{i=1}^{n-1} \lambda_i(A) \leq (\mu - p_n\eta)(A)$ para todo $A \in \mathscr{B}(S)$; sumado λ_n de ambos lados concluimos.

[Caso 2: $\varepsilon_0 < p_n$] Sea $\mu' = \mu - \varepsilon_0\eta$; por ser ε_0 el supremo, existe $I_0 \subset \{1, \dots, n-1\}$ tal que

$$\sum_{i \in I} p_i \leq \mu'\left(\bigcup_{i \in I} A_i\right) \text{ para todo } I \subset I_0$$

y la igualdad se da en el caso en que $I = I_0$, por la hipótesis de inducción existen λ_i con $i \in I_0$ medidas positivas de Borel sobre S, que satisfacen $\lambda_i(A_i) = \lambda_i(S) = p_i$ para cada $i \in I_0$ y $\sum_{i \in I_0} \lambda_i(A) \leq \mu'(A)$ para toda $A \in \mathscr{B}(S)$. Definamos $p'_i = p_i$ para $i = 1, \dots, n-1$, $p'_n = p_n - \varepsilon_0$, $B_0 = \cup_{i \in I_0} A_i$, y $\mu''(A) = \mu'(A) - \mu'(A \cap B_0)$ para toda $A \in \mathscr{B}(S)$, entonces $\sum_{i \in I_0} p_i = \mu'\left(\bigcup_{i \in I_0} A_i\right) = \mu'(B_0)$; tomando $I_1 = \{1, \dots, n\} - I_0$, para todo $I \subset I_1$ se tiene

$$\sum_{i \in I_0 \cup I} p'_i = \mu'(B_0) + \sum_{i \in I} p'_i$$

entonces por (1.2) $\sum_{i \in I \cup I_0} p'_i \leq \mu'\left(\cup_{i \in I \cup I_0} A_i\right)$ si n no pertenece a I; por otro lado si n pertenece a I

$$\begin{aligned} \sum_{i \in I \cup I_0} p'_i &\leq \mu\left(\bigcup_{i \in I \cup I_0} A_i\right) - \varepsilon_0 \\ &= \mu'\left(\bigcup_{i \in I \cup I_0} A_i\right) \end{aligned}$$

puesto que $p'_n = p_n - \varepsilon_0$ y por (1.1), de esto se sigue

$$\begin{aligned}
\sum_{i\in I} p'_i + \mu'(B_0) &\leq \mu'\left(\bigcup_{i\in I\cup I_0} A_i\right) \\
&= \mu'\left(\bigcup_{i\in I} A_i\right) + \mu'(B_0) - \mu'\left(\bigcup_{i\in I} A_i \cap B_0\right) \\
&= \mu''\left(\bigcup_{i\in I} A_i\right) + \mu'(B_0)
\end{aligned}$$

así pues $\sum_{i\in I} p_i \leq \mu''(\cup_{i\in I} A_i)$ para todo $I \subset I_1$.

Por la hipótesis de inducción existen medidas positivas de Borel λ'_i en S con $i \in I_1$, tales que $\lambda'_i(A_i) = \lambda'_i(S) = p'_i$ para cada $i \in I_1$ y $\sum_{i\in I_1} \lambda'_i(A) \leq \mu''(A)$ para toda $A \in \mathcal{B}(S)$. Tomando $\lambda_i = \lambda'_i$ con $i \in I_1 - \{n\}$ y $\lambda_n = \lambda'_n + \varepsilon_0\eta$, resulta $\lambda_i(A_i) = \lambda_i(S) = p_i$ para toda $i \in I_1$, y

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_i(A) &= \sum_{i\in I_0} \lambda_i(A) + \sum_{i\in I_1} \lambda_i(A) \\
&= \sum_{i\in I_0} \lambda_i(A\cap B_0) + \sum_{i\in I_1} \lambda'_i(A) + \varepsilon_0\eta(A) \\
&\leq \mu'(A\cap B_0) + \mu''(A) + \varepsilon_0\eta(A) \\
&= \mu'(A) + \varepsilon_0\eta(A) \\
&= \mu(A)
\end{aligned}$$

para todo $A \in \mathcal{B}(S)$, y $\lambda_1, \ldots, \lambda_n$ satisfacen lo deseado. $\square$

Corolario 1.1 *Sea μ una medida positiva de Borel en S, $p_i \geq 0$ y $A_i \in \mathcal{B}(S)$ para $i = 1, \ldots, n$. Sea $\varepsilon > 0$ de manera que*

$$\sum_{i\in I} p_i \leq \mu\left(\bigcup_{i\in I} A_i\right) + \varepsilon \qquad \textit{para todo } I \subset \{1, \ldots, n\} \tag{1.3}$$

Entonces existen $\lambda_1, \ldots, \lambda_n$ medidas positivas de Borel en S de manera que para $i = 1, \ldots, n$, $\lambda_i(A_i) = \lambda_i(S) \leq p_i$, $\sum_{i=1}^{n} \lambda_i(S) \geq \sum_{i=1}^{n} p_i - \varepsilon$, y más aún $\sum_{i=1}^{n} \lambda_i(A) \leq \mu(A)$ para todo $A \in \mathcal{B}(S)$.

DEMOSTRACIÓN: Sea $S' = S \cup \{\Delta\}$, con Δ un punto ajeno a S, y extendamos μ a S' definiendo $\mu(\{\Delta\}) = \varepsilon$, considerando los conjuntos $A'_i = A_i \cup \{\Delta\}$, por (1.3) tenemos

$$\sum_{i \in I} p_i \leq \mu\left(\bigcup_{i \in I} A'_i\right) \qquad \text{para todo } I \subset \{1, \ldots, n\}$$

entonces por el Lema 1.2, existen $\lambda'_1, \ldots, \lambda'_n$ medias positivas de Borel en S', tales que $\lambda'_i(A'_i) = \lambda'_i(S') = p_i$ y $\sum_{i=1}^n \lambda'_i(A) \leq \mu(A)$ para toda $A \in \mathscr{B}(S')$. Sea λ_i igual a λ'_i restringida a $\mathscr{B}(S)$, entonces $\lambda_i(A_i) = \lambda'_i(A_i) \leq \lambda'_i(A'_i) = p_i$ y $\lambda_i(S \setminus A_i) = \lambda'_i(S' \setminus A'_i) = 0$, se sigue que $\lambda_i(S) \leq p_i$ con $i = 1, \ldots. n$, así mismo

$$\begin{aligned} \sum_{i=1}^n \lambda_i(S) &= \sum_{i=1}^n (p_i - \lambda'_i(\{\Delta\})) \\ &\geq \sum_{i=1}^n p_i - \mu(\{\Delta\}) \\ &= \sum_{i=1}^n p_i - \varepsilon \end{aligned}$$

Por último $\sum_{i=1}^n \lambda_i(A) = \sum_{i=1}^n \lambda'_i(A_i) \leq \mu(A)$ para todo $A \in \mathscr{B}(S)$. □

Proposición 1.1 *Sea S separable y $\varepsilon > 0$, entonces existen $E_1, \ldots, E_n \in \mathscr{B}(S)$, ajenos con diámetro menor que ε y $P(E_0) \leq \varepsilon$ con $E_0 = S \setminus \cup_{i=1}^n E_i$.*

DEMOSTRACIÓN: Sea $\{x_n\}$ un subconjunto denso en S, y $B(x_n, \varepsilon) = B_n$ bolas abiertas de radio ε centradas en x_n, como $S = \cup_n B_n$, existe N de manera que $P(\cup_{i=1}^N B_i) \geq 1 - \varepsilon$. Elegimos $E_i = B_i \setminus \cup_{j=1}^{i-1} B_j$ conjuntos ajenos de diámetro menor que ε, $\cup_{i=1}^N B_i = \cup_{i=1}^N E_i$ y $P(E_0) < \varepsilon$. □

Todo está listo para demostrar el siguente lema, que es el núcleo de la demostración de la completitud de $\mathscr{P}(S)$, así como también de otros resultados posteriores como el Teorema de Representación de Skorohod.

Lema 1.3 *Sea S separable, P, $Q \in \mathscr{P}(S)$ tal que $\rho(P, Q) < \varepsilon$, y $\delta > 0$. Supongamos que $E_1, \ldots, E_N \in \mathscr{B}(S)$ son conjuntos ajenos con diámetro menor que δ y $P(E_0) \leq \delta$, con $E_0 = S \setminus \cup_{i=1}^N E_i$. Entonces existen constantes $c_1, \ldots, c_N \in [0, 1]$, variables aleatorias $X, Y_0, \ldots, Y_N$ S-valuadas, y ξ*

$[0,1]$-valuadas todas ellas indepedientes definidas en un espacio de probabilidad $(\Omega, \mathscr{F}, \nu)$, tal que X tiene distribución P, ξ es uniformemente distribuída en $[0,1]$, la variable aleatoria

$$Y = \begin{cases} Y_i & \text{en } \{X \in E_i, \xi \geq c_i\}, \text{ con } i = 1, \ldots, N \\ Y_0 & \text{en } \{X \in E_0\} \cup \bigcup_{i=1}^{N} \{X \in E_i, \xi < c_i\} \end{cases} \tag{1.4}$$

tiene distribución Q,

$$\{d(X,Y) \geq \delta + \varepsilon\} \subset \{X \in E_0\} \cup \left\{\xi < \max\left[\frac{\varepsilon}{P(E_i)} : P(E_i) > 0\right]\right\}$$

y

$$\nu\{d(X,Y) \geq \delta + \varepsilon\} \leq \delta + \varepsilon.$$

DEMOSTRACIÓN: Definamos $p_i = P(E_i)$ y $A_i = E_i^\varepsilon$ con $i = 1, \ldots, N$, entonces para todo $I \subset \{1, \ldots, N\}$ se tiene

$$\sum_{i \in I} p_i \leq P\left(\overline{\bigcup_{i \in I} E_i}\right) \leq Q\left(\overline{\bigcup_{i \in I} E_i}\right) + \varepsilon \leq Q\left(\bigcup_{i \in I} A_i\right) + \varepsilon$$

por el Corolario 1.1, existen $\lambda_1, \ldots, \lambda_N$ medidas positivas de Borel en S que satisfacen $\lambda_i(A_i) = \lambda_i(S) \leq p_i$, $\sum_{i=1}^{N} \lambda_i(S) \geq \sum_{i=1}^{N} p_i - \varepsilon$ y $\sum_{i=1}^{N} \lambda_i(A) \leq Q(A)$ para toda $A \in \mathscr{B}(S)$. Definamos $c_i = (p_i - \lambda_i(S))/p_i$ y 0 en el caso en que $p_i = 0$; de esta manena $c_i \in [0,1]$, con $i = 1, \ldots, N$; calculando

$$\begin{aligned} P(E_0) + \sum_{i=1}^{N} c_i P(E_i) &= P\left(S \setminus \bigcup_{i=1}^{N} E_i\right) + \sum_{i=1}^{N} (P(E_i) - \lambda_i(S)) \\ &= 1 - \sum_{i=1}^{N} \lambda_i(S) \end{aligned} \tag{1.5}$$

Definamos las medidas $Q_0, \ldots, Q_N$ en $\mathscr{B}(S)$ como

$$\begin{aligned} Q_i(B) &= \frac{\lambda_i(B)}{\lambda_i(S)} \qquad \text{para } i = 1, \ldots, N \text{ y} \\ Q_0(B) &= \frac{Q(B) - \sum_{i=1}^{N} \lambda_i(B)}{1 - \sum_{i=1}^{N} \lambda_i(S)} \end{aligned}$$

con $B \in \mathscr{B}(S)$. Sean $X, Y_0, \dots, Y_N$ variables aleatorias independientes definidas en un mismo espacio de probabilidad $(\Omega, \mathscr{F}, \nu)$ en el cual tienen distribución $P, Q_0, \dots, Q_N$ y ξ esta distribuida uniformemente en $[0,1]$, supongamos que Y_i toma valores en A_i. Procedamos a desentrañar la distribución de Y definida en (1.4), para $B \in \mathscr{B}(S)$ y por (1.5) tenemos

$$\begin{aligned}\nu\{Y \in B\} &= \sum_{i=1}^{N} Q_i(B)(1-c_i)P(E_i) + Q_0(B)\left(P(E_0) + \sum_{i=1}^{N} c_i P(E_i)\right) \\ &= Q(B)\end{aligned}$$

Como $A_i = E_i^{\varepsilon}$, se tiene $\{X \in E_i, \xi \geq c_i\} \subset \{X \in E_i, Y \in A_i\} \subset \{d(X,Y) < \delta + \varepsilon\}$ para $i = 1, \dots, N$, entonces

$$\begin{aligned}\{d(X,Y) \geq \delta + \varepsilon\} &\subset \{X \in E_0\} \cup \bigcup_{i=1}^{N} \{X \in E_i, \xi < c_i\} \\ &\subset \{X \in E_0\} \cup \{\xi < \bigvee_{i=1}^{N} c_i\} \\ &\subset \{X \in E_0\} \cup \left\{\xi < \max\left[\frac{\varepsilon}{P(E_i)} : P(E_i) > 0\right]\right\}\end{aligned}$$

Por último se tiene

$$\begin{aligned}\nu\{d(X,Y) \geq \delta + \varepsilon\} &\leq P(E_0) + \sum_{i=1}^{N} c_i P(E_i) \\ &= P(E_0) + \sum_{i=1}^{N} (p_i - \lambda(S)) \\ &\leq \delta + \varepsilon\end{aligned}$$

□

Con ayuda del lema anterior podemos dar una interpretación probabilista de la métrica de Prohorov cuando el espacio S es separable.

Teorema 1.3 *Sea (S,d) separable y $P, Q \in \mathscr{P}(S)$. Se define $\mathscr{M}(P,Q)$ como el conjunto de medidas de probabilidad $\mu \in \mathscr{P}(S \times S)$ con marginales P y Q,(es decir, $\mu(A \times S) = P(A)$ y $\mu(S \times A) = Q(A)$ para toda A en $\mathscr{B}(S)$).*

Entonces

$$\rho(P,Q) = \inf_{\mu \in \mathscr{M}(P,Q)} \inf \{\varepsilon > 0 : \mu\{(x,y) : d(x,y) \geq \varepsilon\} \leq \varepsilon\} \quad (1.6)$$

DEMOSTRACIÓN: Sean $\varepsilon > 0$ y $\mu \in \mathscr{M}(P,Q)$ de forma que $\mu\{(x,y) : d(x,y) \geq \varepsilon\} \leq \varepsilon$, entonces

$$\begin{aligned} P(F) &= \mu(F \times S) \\ &= \mu\left((F \times S) \cap (\{(x,y) : d(x,y) < \varepsilon\} \cup \{(x,y) : d(x,y) \geq \varepsilon\})\right) \\ &\leq \mu((F \times S) \cap \{(x,y) : d(x,y) < \varepsilon\}) + \varepsilon \\ &\leq \mu(S \times F^{\varepsilon}) + \varepsilon = Q(F^{\varepsilon}) + \varepsilon \end{aligned}$$

es decir, para ε arbitraria $\rho(P,Q) \leq \varepsilon$, entonces se da la desigualdad menor igual en la ecuación (1.6). Centrandonos en el Lema 1.3 tenemos $\mu \in \mathscr{M}(P,Q)$, y si δ tiende a cero, se obtiene la desigualdad restante. □

Corolario 1.2 *Sea S separable, X_n y X variables aleatorias S- valuadas con $n = 1, 2, \ldots$, definidas en un mismo espacio de probabilidad, con distribución P_n y P respectivamente. Si $d(X_n, X) \to 0$ en probabilidad cuando $n \to \infty$, entonces* $\lim_{n\to\infty} \rho(P_n, P) = 0$.

DEMOSTRACIÓN: Para n=1,2, ... , sea μ_n la distribución conjunta del vector (X_n, X). Entonces $\lim_{n\to\infty} \mu_n\{(x,y) : d(x,y) \geq \varepsilon\} = 0$ para todo $\varepsilon > 0$, y por el Teorema 1.3 culminamos la prueba. □

Se ha trabajado duro, pero a cambio se tiene el siguiente teorema importante, que como mencionamos nos refleja dos propiedades que hereda $\mathscr{P}(S)$ de S, una métrica y la otra topológica.

Antes de dar inicio definamos una medida sencilla e importante, que es la *medida de Dirac*, para x en S se define la medida δ_x como:

$$\delta_x(A) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

con A en $\mathscr{B}(S)$.

Teorema 1.4 *Si S es separable, entonces $\mathscr{P}(S)$ es separable. Si además (S,d) es completo, $(\mathscr{P}(S), \rho)$ es completo.*

DEMOSTRACIÓN: Primero demostremos que $\mathscr{P}(S)$ es separable. Sea $\{x_n\}$ un subconjunto numerable denso en S, y δ_x la medida de probabilidad que esta concentrada en x; consideremos el conjunto

$$\mathscr{D} = \left\{ P \in \mathscr{P}(S) : P = \sum_{i=1}^{n} a_i \delta_{x_i} \text{ con } a_i \in \mathbb{Q}, \sum_{i=1}^{n} a_i = 1, \text{ y } n \text{ finito} \right\}$$

que es un conjunto numerable. Sólo resta probar que es un conjunto denso en $\mathscr{P}(S)$.

Sea $\varepsilon > 0$ y $Q \in \mathscr{P}(S)$, elegimos $E_1, \ldots E_n \in \mathscr{B}(S)$ ajenos de diámetro menor que ε, con $x_i \in E_i$ y tal que $Q(E_0) \leq \varepsilon$ con E_0 como de costumbre. Sea $m > n$ entero positivo, llamemos $a_i = k_i/m$ con $k_i = [m \cdot Q(E_i)]$ y $k_0 = 1 - \sum_{i=1}^{n} k_i$. Para $F \subset S$ y $M = \cup_{i=1}^{n} E_i$ se tiene

(*i*) Si $F \cap M = \emptyset$, entonces $Q(F) \leq \varepsilon$

(*ii*) Si $F \cap M \neq \emptyset$, entonces $F = F_1 \cup F_2$ con $F_1 \subset M$ y F_2 contenido en el complemento de M así que $F_1 \subset \cup_{F_1 \cap E_i \neq \emptyset} E_i$ entonces $Q(F_1) \leq Q(\cup_{F_1 \cap E_i \neq \emptyset} E_i)$, y con (*i*) se tiene

$$Q(F) = Q(F_1) + Q(F_2) \leq \sum_{F_1 \cap E_i \neq \emptyset} Q(E_i) + \varepsilon$$

entonces para F conjunto cerrado de S

$$\begin{aligned} Q(F) &\leq \sum_{F \cap E_i \neq \emptyset} Q(E_i) + \varepsilon \\ &\leq \sum_{i=1}^{n} \frac{[m \cdot Q(E_i)]}{m} \delta_{x_i}(F) + \varepsilon \text{con } x_i \in F \\ &= P(F) + \varepsilon \leq P(F^{\varepsilon}) + \varepsilon \end{aligned}$$

donde $P = \sum([m \cdot Q(E_i)]/m)\delta_{x_i}$, por tanto $\rho(P, Q) \leq \varepsilon$.

Para demostrar la completitud de $\mathscr{P}(S)$ es suficiente considerar una sucesión $\{P_n\} \subset \mathscr{P}(S)$ de manera que para $n \geq 2$ $\rho(P_{n-1}, P_n) < 2^{-n}$. Para cada n elegimos $E_1^{(n)}, \ldots, E_{N_n}^{(n)}$ elementos de $\mathscr{B}(S)$ ajenos, con diámetro menor que 2^{-n} tal que que $P_{n-1}(E_0^{(n)}) \leq 2^{-n}$, donde $E_0^{(n)} = S \setminus \cup_{i=1}^{N_n} E_i^{(n)}$. Utilizando el Lema 1.3 en cada elemento de $\{P_n\}$, llegamos a que existe un espacio de probabilidad $(\Omega, \mathscr{F}, \nu)$ en el cual están definidas las variables aleatorias $Y_0^{(n)}, \ldots, Y_{N_n}^{(n)}$ S-valuadas, $\xi^{(n)}$ $[0,1]$-valuadas con $n = 2, 3, \ldots$, X_1 S-valuada con distribución P_1, todas ellas independientes y constantes $c_1^{(n)}, \ldots, c_{(N_n)}^{(n)} \in$

$[0,1]$ con $n = 2, 3, \ldots$, tal que la variable aleatoria

$$X_n = \begin{cases} Y_i^{(n)} & \text{en } \{X_{n-1} \in E_i^{(n)}, \xi^{(n)} \geq c_i^{(n)}\}, \text{con } i = 1, \ldots, N_n \\ Y_0^{(n)} & \text{en } \{X_{n-1} \in E_0^{(n)}\} \cup \bigcup_{i=1}^{N_n} \{X_{n-1} \in E_i^{(n)}, \xi^{(n)} < c_i^{(n)}\} \end{cases}$$

tiene distribución P_n, y

$$\nu\{d(X_n - 1, X_n) \geq 2^{-n+1}\} \leq 2^{-n+1}$$

Por el Lema de Borel-Cantelli se obtiene

$$\nu\left\{\sum_{n=2}^{\infty} d(X_{n-1}, X_n) < \infty\right\} = 1$$

por ser (S, d) un espacio completo $\lim_{n\to\infty} X_n = X$ c.s. y el Corolario 1.2 nos da como resultado $\lim_{n\to\infty} \rho(P_n, P) = 0$, por lo cual $(\mathscr{P}(S), \rho)$ es un espacio métrico completo. □

Una aplicación más del Lema 1.3, es para la prueba del Teorema de Representación de Skorohod, que se expone en el próximo teorema.

Teorema 1.5 (Representación de Skorohod) *Sea (S, d) un espacio separable, P_n, $P \in \mathscr{P}(S)$ con $n = 1, 2, \ldots$ de manera que $\lim_{n\to\infty} \rho(P_n, P) = 0$. Entonces existe un espacio de probabilidad $(\Omega, \mathscr{F}, \nu)$ en el cual están definidas X_n, X variables aleatorias S-valuadas con distribución P_n, P respectivamente con $n = 1, 2 \ldots$, tal que $\lim_{n\to\infty} X_n = X$ c.s.*

DEMOSTRACIÓN: Para $k = 1, 2, \ldots$ tomemos los conjuntos $E_1^{(k)}, \ldots, E_{N_k}^{(k)} \in \mathscr{B}(S)$ ajenos, con diámetro menor que 2^{-k} y $P(E_0^{(k)}) \leq 2^{-k}$, donde $E_0^{(k)} = S \setminus \cup_{i=1}^{N_k} E_i^{(k)}$; supongamos que $P(E_i^{(k)}) > 0$, y defimanos ε_k como el mínimo de $P(E_i^{(k)})$ con $1 \leq i \leq N_k$ y la sucesión $\{k_n\}$ como

$$k_n = \max\left\{1, \left\{k \geq 1 : \rho(P_n, P) < \frac{\varepsilon_k}{k}\right\}\right\}$$

Empleando el Lema 1.3 con $Q = P$, $\varepsilon = \varepsilon_{k_n}/k_n$ si $k_n > 1$ o $\varepsilon = \rho(P_n, P) + 1/n$ si $k = 1$, $\delta = 2^{-k_n}$, $E_i = E_i^{(k_n)}$ y $N = N_{k_n}$ para $n = 1, 2, \ldots$; entonces existe un espacio de probabilidad $(\Omega, \mathscr{F}, \nu)$ en el cual están definidas las variables aleatorias $Y_0^{(n)}, \ldots, Y_{N_{k_n}}^{(n)}$ S-valuadas con $n = 1, 2, \ldots$, la variable aleatoria ξ uniformemente distribuida en $[0, 1]$ y la variable aleatoria X S-valuada con distribución P, todas ellas independientes; además existen constantes $c_1^{(n)}, \dots, c_{N_{k_n}}^{(n)} \in [0,1]$ con $n = 1, 2 \dots$, y la variable aleatoria

$$X_n = \begin{cases} Y_i^{(n)} & \text{en } \{X \in E_i^{(k_n)}, \xi \geq c_i^{(n)}\}, \text{con } i = 1, \dots, N_{k_n} \\ Y_0^{(n)} & \text{en } \{X \in E_0^{(k_n)}\} \cup \bigcup_{i=1}^{N_{k_n}} \{X \in E_i^{(k_n)}, \xi < c_i^{(n)}\} \end{cases}$$

tiene distribución P_n, y

$$\left\{d(X_n, X) \geq 2^{-k_n} + \frac{\varepsilon_{k_n}}{k_n}\right\} \subset \{x \in E_0^{(k_n)}\} \bigcup \left\{\varepsilon < \frac{1}{k_n}\right\}$$

si $k_n > 1$; y para $n = 1, 2, \dots$ definamos $K_n = \min_{m \geq n} k_m > 1$ entonces $K_n > 1$, y

$$\begin{aligned} \nu\left(\bigcup_{m=n}^{\infty} \left\{d(X_m, X) \geq 2^{-k_m} + \frac{\varepsilon_{k_m}}{k_m}\right\}\right) & \\ &\leq \sum_{k=K_n}^{\infty} \nu\{X \in E_0^{(k)}\} + \nu\left\{\xi < \frac{1}{K_n}\right\} \\ &\leq 2^{-K_n+1} + \frac{1}{K_n} \end{aligned}$$

como $\lim_{n \to \infty} K_n = \infty$, jovialmente tenemos $\lim_{n \to \infty} X_n = X$ c.s. □

A continuación se presenta el teorema del mapeo continuo, con el que damos fin a esta primera sección.

Corolario 1.3 *Sean (S, d), (S', d') espacios métricos separables, y $h : S \to S'$ una función Borel medible. Supongamos que P_n, $P \in \mathscr{P}(S)$ con $n = 1, 2, \dots$ satisfacen $\lim_{n \to \infty} \rho(P_n, P) = 0$, sean Q_n, $Q \in \mathscr{P}(S')$ con $n = 1, 2, \dots$ definidas como*

$$Q_n = P_n h^{-1}, \qquad Q = P h^{-1}$$

Sea C_h el conjunto de puntos donde h es continua; si $P(C_h) = 1$, entonces $\lim_{n \to \infty} \rho'(Q_n, Q) = 0$, donde ρ' es la métrica de Prohorov de $\mathscr{P}(S')$.

DEMOSTRACIÓN: Por el Teorema de Representación de Skorohod existe un espacio de probabilidad $(\Omega, \mathscr{F}, \nu)$ en el cual estan definidas las variables aleatorias X_n, X S-valuadas con distribución P_n, P $n = 1, 2, \dots$ respectivamente, tal que $\lim_{n \to \infty} X_n = X$ c.s. Como $\nu(C_h) = 1$, se tiene $\lim_{n \to \infty} h(X_n) = h(X)$ c.s. y por el Corolario 1.2 $\lim_{n \to \infty} \rho'(Q_n, Q) = 0$. □

1.2 El Teorema de Prohorov

Para poder abordar el tema de convergencia de medidas de probabilidad con mayor facilidad, será necesario una caracterización de los conjuntos compactos en $\mathscr{P}(S)$, donde S es un espacio métrico. Para llevar acabo tal objetivo, destacaremos las medias que estan detrminadas por lo valores que toman sobre conujuntos compactos.

Una medida $P \in \mathscr{P}(S)$ es *tensa*, si para cada $\varepsilon > 0$ existe K compacto en S, tal que $P(K) \geq 1-\varepsilon$. Una familiad de medidas de probabilidad $\mathscr{M} \in \mathscr{P}(S)$ es *tensa*, si para cada $\varepsilon > 0$ existe K compacto en S tal que

$$\inf_{P \in \mathscr{M}} P(K) \geq 1 - \varepsilon.$$

Lema 1.4 *Si (S, d) es un espacio métrico completo y separable, entonces $P \in \mathscr{P}(S)$ es tensa.*

DEMOSTRACIÓN: Sea $\varepsilon > 0$ y $D = \{x_n\}$ un conjunto denso en S; para $n \geq 1$ y por la regularidad de $P \in \mathscr{P}(S)$ se tiene

$$P\left(\bigcup_{k=1}^{N_n} B\left(x_k, \frac{1}{n}\right)\right) \geq 1 - \frac{\varepsilon}{2^n}$$

Sea K la cerradura de $\cap_{n\geq 1} \cup_{k=1}^{N_n} B(x_k, 1/n)$, K es totalmente acotado y como S es completo, en consecuencia K compacto, entonces

$$\begin{aligned} P(K) &\geq P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=1}^{N_n} B(x_k, 1/n)\right) \\ &\geq 1 - \sum_{n=1}^{\infty} P\left(\bigcap_{k=1}^{N_n} B(x_k, 1/n)^c\right) \\ &\geq 1 - \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = 1 - \varepsilon \end{aligned}$$

□

Teorema 1.6 *[Prohorov] Sea (S, d) un espacio métrico completo y separable, y $\mathscr{M} \subset \mathscr{P}(S)$. Entonces las siguientes afirmaciones son equivalentes:*

(a) $\mathscr{M}$ es tensa.

(b) Para cada $\varepsilon > 0$, existe $K \subset S$ compacto de manera que

$$\inf_{P \in \mathcal{M}} P(K^{\varepsilon}) \geq 1 - \varepsilon.$$

(c) $\mathcal{M}$ es relativamente compacto.

DEMOSTRACIÓN:

[$a \Rightarrow b$] Sea $\varepsilon > 0$, por ser $\mathcal{M}$ una familia de medidas tensa existe K subconjunto de S compacto tal que $\inf_{P \in \mathcal{M}} P(K) \geq 1 - \varepsilon$, y por tanto

$$\inf_{P \in \mathcal{M}} P(K^{\varepsilon}) \geq 1 - \varepsilon$$

[$b \Rightarrow c$] Por el Teorema 1.4, la cerradura de $\mathcal{M}$ es completo, sólo resta probar que es totalmente acotado. Sean $\varepsilon, \delta > 0$, de manera que $\varepsilon < \delta/2$, y $K \subset S$ compacto tal que

$$\inf_{P \in \mathcal{M}} P(K^{\varepsilon}) \geq 1 - \varepsilon$$

Además existen $x_1, \ldots, x_n$ en K tal que $K \subset \cup_{i=1}^{n} B(x_i, \varepsilon)$ y por consiguiente $K^{\varepsilon} \subset \cup_{i=1}^{n} B(x_i, 2\varepsilon)$, tomando $x_0 \in S$ y $m \geq n/\varepsilon$ formamos el conjunto $\mathcal{N}$ de las medidas de probabilidad de la forma:

$$\mathcal{N} = \left\{ P \in \mathscr{P}(S) : P = \sum_{i=0}^{n} \left(\frac{k_i}{m} \right) \delta_{x_i}, \text{ donde } 0 \leq k_i \leq m \text{ y } \sum_{i=0}^{n} k_i = m \right\}$$

Sea $Q \in \mathcal{M}$ y $k_i = [m \cdot Q(E_i)]$ $i = 1, \ldots, n$ donde $E_i = B_i \setminus \cup_{j=1}^{i-1} B_j$ y $k_0 = m - \sum_{i=1}^{n} k_i$, para $F \in \mathscr{C}$ se tiene:

$$\begin{aligned} Q(F) &\leq \sum_{F \cap E_i \neq \emptyset} Q(E_i) + \varepsilon \\ &\leq \sum_{F \cap E_i \neq \emptyset} \frac{[mQ(E_i)]}{m} + \frac{n}{m} + \varepsilon \\ &\leq P\left(F^{2\varepsilon}\right) + 2\varepsilon \end{aligned}$$

entonces $\rho(P, Q) \leq 2\varepsilon \leq \delta$, es decir, $\mathcal{M}$ es totalmente acotado.

[$c \Rightarrow a$] Sea $\varepsilon > 0$, como $\mathcal{M}$ es totalmente acotado existe un subconjunto finito $\mathcal{N}_n$ de $\mathcal{M}$ tal que $\mathcal{M} \subset \{Q : \rho(P, Q) < \varepsilon/2^{n+1}$ con $P \in \mathcal{N}_n\}$

para $n = 1, 2, \ldots$. Por el Lema 1.4 existe $K_n \subset S$ compacto tal que $P(K_n) \geq 1 - \varepsilon/2^{n+1}$ para toda $P \in \mathscr{N}_n$. Sea $Q \in \mathscr{M}$, para cada n existe $P_n \in \mathscr{N}_n$ tal que

$$Q(K_n^{\varepsilon/2^{n+1}}) \geq P_n(K_n) - \frac{\varepsilon}{2^{n+1}} \geq 1 - \frac{\varepsilon}{2^n}$$

Tomando K como la cerradura de $\cap_{n\geq 1} K_n^{\varepsilon/2^{n+1}}$, resulta que K es compacto y además

$$Q(K) \geq 1 - \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = 1 - \varepsilon$$

por tanto $\mathscr{M}$ es tensa. □

Si en el teorema precedente se omiten las condiciones de separabilidad y completitud de S, un conjunto relativamente compacto de medidas no necesariamente es tenso. Sin embargo la otra implicación es cierta, tal y como lo presenta el siguiente teorema.

Teorema 1.7 *Sea (S, d) un espacio métrico arbitrario, $\mathscr{M} \subset \mathscr{P}(S)$. Si $\mathscr{M}$ es tensa entonces $\mathscr{M}$ es relativamente compacto.*

DEMOSTRACIÓN: Para $m \geq 1$, existe un conjunto $K_m \subset S$ compacto de manera que

$$\inf_{P \in \mathscr{M}} P(K_m) \geq 1 - \frac{1}{m}$$

Supongamos que $K_1 \subset K_2 \subset \cdots$. Para cada $P \in \mathscr{M}$ y $m \geq 1$ se define $P^{(m)} \in \mathscr{P}(S)$ como $P^{(m)}(A) = P(A \cap K_m)/P(K_m)$, de esta forma $P^{(m)} \in \mathscr{P}(K_m)$, entonces $\mathscr{M}^{(m)} = \{P^{(m)} : P \in \mathscr{M}\}$ es relativamente compacto en $\mathscr{P}(K_m)$ con $m \geq 1$ debido a que K_m es separable y compacto. Entonces

$$P(A) \leq P(A \cap K_m) + \frac{1}{m} \leq P^{(m)}(A) + \frac{1}{m}$$

esto es $\rho(P, P^{(m)}) \leq 1/m$, además obtenemos las siguientes desigualdades

$$P(A)^{(m')} \leq \frac{P(A \cap K_m) + 1/m}{P(K_m)} \leq P^{(m)}(A) + \frac{2}{m} \quad \text{con } m' > m \tag{1.7}$$

$$P(A) \geq P(K_m)P^{(m)}(A) \geq (1 - 1/m)P^{(m)}(A)$$

$$P(A)^{(m')} \geq \frac{P(K_m)P^{(m)}(A)}{P(K_m)} \geq \left(1 - \frac{1}{m}\right) P^{(m)}(A) \quad \text{con } m' > m \tag{1.8}$$

para toda $P \in \mathcal{M}$, $A \in \mathcal{B}(S)$ y $m \geq 1$. Dados conjuntos $A_1, A_2, \ldots \in \mathcal{B}(S)$ ajenos, por las desigualdades (1.7) y (1.8) resulta

$$\begin{aligned}\sum_i \mid P^{(m')}(A_i) - P^{(m)}(A_i) \mid & \\ & \leq \sum_i \left(P^{(m')}(A_i) - \left(1 - \frac{1}{m}\right) P^{(m)}(A_i) + \frac{1}{m} P^{(m)}(A_i) \right) \\ & \leq P^{(m')}\left(\bigcup_i A_i\right) - P^{(m)}\left(\bigcup_i A_i\right) + \frac{2}{m} \\ & \leq \frac{2}{m} + \frac{2}{m} = \frac{4}{m}\end{aligned} \tag{1.9}$$

para cada $P \in \mathcal{M}$ y $m' > m \geq 1$. Sea $\{P_n\} \subset \mathcal{M}$ cualquier sucesión, entonces existe una subsucesión $\{P_{n_k}\}$ de $\{P_n\}$ y $Q^{(m)} \in \mathcal{P}(K_m)$ de manera que

$$\lim_{k\to\infty} \rho(P_{n_k}^{(m)}, Q^{(m)}) = 0 \text{ para } m \geq 1$$

por ser $\mathcal{M}^{(m)}$ relativamente compacto en $\mathcal{P}(K_m)$, luego para F cerrado existe ε, con $Q^{(m)}(F) \leq P_{n_k}^{(m)}(F^\varepsilon) + \varepsilon$, de esta manera $Q^{(m)}(F)$ se puede calcular mediante

$$Q^{(m)}(F) = \lim_{\varepsilon\to 0} \limsup_{k\to\infty} P_{n_k}^{(m)}(F^\varepsilon)$$

para todo F subconjunto cerrado de S. De (1.9) se tiene

$$Q(A) = \lim_{m\to\infty} Q^{(m)}(A)$$

para toda $A \in \mathcal{B}(S)$, y así se define una medida de probabilidad $Q \in \mathcal{P}(S)$. Como

$$\begin{aligned}\rho(P_{n_k}, Q) &\leq \rho(P_{n_k}, P_{n_k}^{(m)}) + \rho(P_{n_k}^{(m)}, Q^{(m)}) + \rho(Q^{(m)}, Q) \\ &\leq \frac{1}{m} + \rho(P_{n_k}^{(m)}, Q^{(m)}) + \frac{2}{m}\end{aligned}$$

para cada k y $m \geq 1$, entonces $\lim_{k\to\infty} \rho(P_{n_k}, Q) = 0$; probando que $\mathcal{M}$ es relativamente compacto. □

Proposición 1.2 *Sean (S_k, d_k) $k = 1, 2, \ldots$ espacios, métricos y (S, d) un espacio métrico con $S = \prod_{k=i}^{\infty} S_k$ y $d(x, y) = \sum_{k=1}^{\infty}(d_k(x_k, y_k) \wedge 1)$ para toda*

$x, y \in S$. Sea $\{P_\alpha\} \subset \mathscr{P}(S)$; para cada α se define $P_\alpha^k \in \mathscr{P}(S)$ la k-ésima distribución marginal de P_α con $k = 1, 2, \ldots$ (i.e. $P_\alpha^k = P_\alpha \pi_k^{-1}$ donde la proyección $\pi_k : S \to S_k$ esta dada por $\pi_k(x) = x_k$). Entonces $\{P_\alpha\}$ es tensa si y sólo si para $k = 1, 2, \ldots$ $\{P_\alpha^k\}$ es tensa.

DEMOSTRACIÓN: Supongamos que $\{P_\alpha^k\}$ es tensa para $k = 1, 2, \ldots$ y sea $\varepsilon > 0$, entonces para cada k existe $K_k \subset S_k$ compacto tal que

$$\inf_\alpha P_\alpha^k(K_k) \geq 1 - \frac{\varepsilon}{2^k}$$

sí $K = \prod_{k=1}^\infty K_k = \cap_{k=1}^\infty \pi_k^{-1}(K_k)$, K es compacto en S y

$$\begin{aligned} P_\alpha(K) &= 1 - P_\alpha\left(\bigcup_{k=1}^\infty \{\pi_k^{-1}(K_k)\}^c\right) \\ &\geq 1 - \sum_{k=1}^\infty (1 - P_\alpha^k(K_k)) \\ &\geq 1 - \varepsilon \end{aligned}$$

para toda α, de aquí que $\{P_\alpha\}$ es tensa.

Sea $K \subset S$ compacto, $\pi_k(K)$ es compacto en S_k y

$$\inf_\alpha P_\alpha^k(\pi_k(K)) \geq \inf_\alpha P_\alpha(K) \geq 1 - \varepsilon \quad \text{para } k = 1, 2, \ldots$$

Por lo cual $\{P_\alpha^k\}$ es tensa. □

1.3 Convergencia Débil

Denotemos por $C(S)$ el conjunto de las funciones continuas real-valuadas en (S, d), dotado con la norma $\|f\| = \sup_{x \in S} | f(x) |$, y a $C_0(S)$ un subconjunto de $C(S)$, el de las funciones acotadas. Una sucesión $\{P_n\} \in \mathscr{P}(S)$ *converge débilmente* a $P \in \mathscr{P}(S)$ ($P_n \Rightarrow P$) si

$$\lim_{n \to \infty} \int f dP_n = \int f dP \quad \text{para toda } f \in C_0(S)$$

Como un ejemplo de convergencia débil consideremos una sucesión convergente $\{x_n\}$ en S al punto $x \in S$, entonces se tiene una sucesión de medidas de probabilidad en $\mathscr{B}(S)$, a saber, $\{\delta_{x_n}\}$. Para f en $C_0(S)$, $f(x_n) \to f(x)$ y más aún $\int f\delta_{x_n} = f(x_n)$ de donde $\delta_{x_n} \Rightarrow \delta_x$.

La *distribución* de una variable aleatoria X S-valuada, es una medida de probabilidad en $\mathscr{B}(S)$ definida como $PX^{-1}(B) = P\{X \in B\}$. Una sucesión $\{X_n\}$ de variables aleatorias S-valuadas *converge en distribución* a la variable aleatoria X S-valuada ($X_n \Rightarrow X$), si $PX_n^{-1} \Rightarrow PX^{-1}$. Cuando se presente confución se especificará el espacio métrico que se este considerando, es decir, $P_n \Rightarrow P$ en S.

Si S' es un espacio métrico, $f : S \to S'$ una función continua y $X_n \Rightarrow X$ en S, entonces $f(X_n) \Rightarrow f(X)$ en S', pues si $g \in C_0(S')$ entonces $g \circ f \in C_0(S)$.

Demos principio a esta sección probando que la convergencia débil es equivalente a convergencia en la métrica de Prohorov. Antes de esto, diremos que $A \in \mathscr{B}(S)$ es P-continuo si $P(\partial A) = 0$, donde ∂A es la frontera del conjunto A.

Teorema 1.8 *Sea (S, d) un espacio métrico, $\{P_n\}$ una sucesión en $\mathscr{P}(S)$ y $P \in \mathscr{P}(S)$. Las condiciones de (b) a (f) son equivalentes, y (a) las implica a todas ellas. Si además S es separable, todas ellas son equivalentes:*

(a) $\lim_{n\to\infty} \rho(P_n, P) = 0$.

(b) $P_n \Rightarrow P$.

(c) $\lim_{n\to\infty} \int f dP_n = \int f dP$ *para toda* $f \in C_0(S)$ *uniformemente continua.*

(d) $\limsup_{n\to\infty} P_n(F) \leq P(F)$ *para todo* $F \subset S$ *cerrado.*

(e) $\liminf_{n\to\infty} P_n(G) \geq P(G)$ *para todo* $G \subset S$ *abierto.*

(f) $\lim_{n\to\infty} P_n(A) = P(A)$ *para todo* A *P-continuo.*

DEMOSTRACIÓN:

$[a \Rightarrow b]$ Sea $\varepsilon_n = \rho(P_n, P) + 1/n$ para $n \geq 1$, y $f \in C_0(S)$ con $f \geq 0$

$$\int f dP_n = \int_0^{\|f\|} P_n(f \geq t)dt \leq \int_0^{\|f\|} P(\{f \geq t\}^{\varepsilon_n})dt + \varepsilon_n\|f\|$$

para toda n, sacando el límite superior

$$\begin{aligned}\limsup_{n\to\infty} \int f dP_n &\leq \lim_{n\to\infty} \int_0^{\|f\|} P(\{f \geq t\}^{\varepsilon_n})dt \\ &= \int_0^{\|f\|} P(\{f \geq t\})dt = \int f dP\end{aligned}$$

Ahora para $f \in C_0(S)$ arbitraria, $\|f\| + f \geq 0$ y $\|f\| - f \geq 0$

$$\limsup_{n\to\infty} \int (\|f\| + f) dP_n \leq \int (\|f\| + f) dP \qquad \text{y}$$

$$\limsup_{n\to\infty} \int (\|f\| - f) dP_n \leq \int (\|f\| - f) dP$$

Por tanto $P_n \Rightarrow P$.

$[b \Rightarrow c]$ Por definición.

$[c \Rightarrow d]$ Sea $F \subset S$ cerrado, y para cada $\varepsilon > 0$ definamos

$$f_\varepsilon(x) = \left(1 - \frac{d(x,F)}{\varepsilon}\right) \vee 0 \qquad x \in S$$

de manera que $f_\varepsilon \in C_0(S)$ es uniformemente continua y $f_\varepsilon(x) = 1$ para toda $x \in F$, entonces

$$\limsup_{n\to\infty} P_n(F) \leq \lim_{n\to\infty} \int f_\varepsilon dP_n = \int f_\varepsilon dP$$

tomando el límite cuando ε se aproxima a cero, llegamos a

$$\limsup_{n\to\infty} P_n(F) \leq \lim_{\varepsilon\to 0} \int f_\varepsilon dP = P(F).$$

$[d \Rightarrow e]$ Sea $G \subset S$ abierto, entonces

$$\liminf_{n\to\infty} P_n(G) = 1 - \limsup_{n\to\infty} P_n(G^c) \geq 1 - P(G^c) = P(G).$$

$[e \Rightarrow f]$ Sea A un conjunto P-continuo, entonces

$$\begin{aligned}
\limsup_{n\to\infty} P_n(A) &\leq \limsup_{n\to\infty} P_n(\overline{A}) \\
&= 1 - \liminf_{n\to\infty} P_n(\overline{A}^c) \\
&\leq 1 - P(\overline{A}^c) \\
&= P(A)
\end{aligned}$$

Por otro lado

$$\liminf_{n\to\infty} P_n(A) \geq \liminf_{n\to\infty} P_n(A^\circ) \geq P(A^\circ) = P(A)$$

por tanto $\lim_{n\to\infty} P_n(A) = P(A)$.

[$f \Rightarrow b$] Sea $f \in C_0(S)$ de manera que $f \geq 0$, entonces $\partial\{f \geq t\} \subset \{f = t\}$, de aquí que $\{f \geq t\}$ es P- continuo, entonces

$$\begin{aligned} \lim_{n\to\infty} \int f dP_n &= \lim_{n\to\infty} \int_0^{\|f\|} P_n\{f \geq f\} dt \\ &= \int_0^{\|f\|} P\{f \geq t\} dt = \int f dP \end{aligned}$$

para $f \in C_0(S)$ considerando nuevamente $\|f\| + f$ y $\|f\| - f$ para concluir $P_n \Rightarrow P$.

[$e \Rightarrow a$ **con** S **separable**] Sea $\varepsilon > 0$ y $E_i \in \mathscr{B}(S)$ con diámetro menor que $\varepsilon/2$ $i = 1, 2, \ldots$ una partición de S. Sea n el menor entero tal que $P(\cup_{i=1}^n E_i) > 1 - \varepsilon/2$, y $\mathscr{G}$ la colección de los conjuntos abiertos de la forma $(\cup_{i\in I} E_i)^{\varepsilon/2}$ donde $I \subset \{1, \ldots, n\}$. Como $\mathscr{G}$ es finito existe n_0 de manera que $P(G) \leq P_n(G) + \varepsilon/2$, para todo $G \in \mathscr{G}$ y $n \geq n_0$. Sea $F \in \mathscr{F}$ y

$$F_0 = \bigcup \{E_i : 1 \leq i \leq n, E_i \cap F \neq \emptyset\}$$

entonces $F^{\varepsilon/2} \in \mathscr{G}$ y

$$\begin{aligned} P(F) &\leq P(F_0^{\varepsilon/2}) + \varepsilon/2 \\ &\leq P_n(F_0^{\varepsilon/2}) + \varepsilon \leq P_n(F^\varepsilon) + \varepsilon \end{aligned}$$

Esto es $\rho(P_n, P) \leq \varepsilon$ para $n \geq n_0$. □

Corolario 1.4 *Sean $P_n, P \in \mathscr{P}(S)$ con $n = 1, 2, \ldots,$ y $S' \in \mathscr{B}(S)$. Supongamos que $P_n(S') = P(S') = 1$ para toda n, y sean P_n' y P las restricciones de P_n y P a $\mathscr{B}(S')$. Entonces $P_n \Rightarrow P$ en S si y sólo si $P_n' \Rightarrow P$ en S'.*

DEMOSTRACIÓN: Sea G' abierto de S', entonces $G' = G \cap S$ para algún G abierto de S; como $P_n \Rightarrow P$ tenemos

$$\limsup_{n\to\infty} P_n'(G') = \limsup_{n\to\infty} P_n(G) \geq P(G) = P'(G')$$

por tanto $P_n' \Rightarrow P'$ en S'. El regreso es de manera similar; si G es un abierto de S, entonces $G' = G \cap S$ es abierto de S' y entonces

$$\limsup_{n\to\infty} P_n(G) = \limsup_{n\to\infty} P_n'(G') \geq P'(G') = P(G)$$

es decir $P_n \Rightarrow P$ en S. □

Corolario 1.5 *Sea (S,d) un espacio métrico, (X_n, Y_n) y X variables aleatorias $(S \times S)$-valuadas y S-valuadas respectivamente con $n = 1, 2, \ldots$. Si $X_n \Rightarrow X$ y $d(X_n, Y_n) \to 0$ en probabilidad, entonces $Y_n \Rightarrow X$.*

DEMOSTRACIÓN: Si $f \in C_0(S)$ es uniformemente continua, tenemos

$$\lim_{n\to\infty} E[f(X_n) - f(Y_n)] = 0$$

ya que $d(X_n, Y_n) \to 0$, con lo que resulta

$$\lim_{n\to\infty} E[f(Y_n)] = \lim_{n\to\infty} E[f(X_n)] = E[f(X)]$$

Por lo cual $Y_n \Rightarrow X$. □

1.4 Compacidad de $\mathscr{P}(S)$

Con ayuda del concepto de convergencia débil se probará la compacidad de $\mathscr{P}(S)$ siempre y cuando S sea un espacio métrico compacto para ello sera necesario el siguiente teorema.

Teorema 1.9 *El espacio métrico S es homeomorfo al subconjunto $\Delta = \{\delta_x : x \in S\}$ de $\mathscr{P}(S)$.*

DEMOSTRACIÓN: Para x en S y $\{x_n\}$ una sucesión en S tal que $x_n \to x$, se tiene $\delta_{x_n} \Rightarrow \delta_x$.

Supongamos que $\delta_{x_n} \Rightarrow \delta_x$ y $\{x_n\}$ no converge a x, entonces existe A un vecindad de x y un subsucesión $\{x_{n'}\}$ tal que $x_{n'}$ no esta en A para toda n'. Por el Lema de Urysohn existe g una función continua con $0 \le g(x) \le 1$, $g(x) = 0$ y $g(y) = 1$ para $y \in A^c$. Entonces $\int g d\delta_{x_{n'}} = 1$, sin embargo $\int g d\delta_x = 0$ lo que es una contradicción. □

Terminamos este primer capítulo con un resultado interesante ya esperado.

Teorema 1.10 *El espacio $\mathscr{P}(S)$ es compacto si y sólo si S es un espacio métrico compacto.*

DEMOSTRACIÓN: Supongamos que S es un espacio métrico compacto, entonces $C_0(S)$ es separable. Sea $\{g_n\}$ un subconjunto denso en la bola unitaria con centro en cero de $C(S)$, tal que $g_1 \equiv 1$, $\|g_n\| \le 1$. Sea $T : \mathscr{P}(S) \to I^\infty$, dada por $T(\mu) = (\int g_1 d\mu, \int g_2 d\mu, \ldots)$ donde I^∞ es el producto numerable del intervalo $[-1,1]$ y además I^∞ es un espacio métrico compacto. Veamos que T es un homeomorfismo.

Supongamos que $T(\mu) = T(\nu)$, entonces $\int g_n d\mu = \int g_n d\nu$ para toda n, como $\{g_n\}$ es denso en en la bola unitaria $\int g d\mu = \int g d\nu$ para toda g en en la bola unitaria de $C_0(S)$ esto es $\mu = \nu$ y T es inyectiva.

Comprobemos la continuidad de T. Sea $\{\mu_\alpha\}$ una sucesión de medidas tal que $\mu_\alpha \Rightarrow \mu$ con $\mu_\alpha, \mu \in \mathscr{P}(S)$, entonces $\int g_n d\mu_\alpha \to \int g_n d\mu$ para toda n y $T(\mu_\alpha) \to T(\mu)$, es decir T es continua. Que pasa con T^{-1}; sea $\{\mu_\alpha\}$ una sucesión de elementos de $\mathscr{P}(S)$, tal que $T(\mu_\alpha) \to T(\mu)$ es decir $\int g_n d\mu_\alpha \to \int g_n d\mu$ para toda n. Para $g \in C_0(S)$, con $\|g\| \leq 1$ se tiene

$$\begin{aligned}
\left|\int g d\mu_\alpha - \int g d\mu\right| &= \left|\int (g - g_n) d\mu_\alpha + \int (g_n - g) d\mu + \right. \\
&\qquad \left. \int g_n d\mu_\alpha - \int g_n d\mu\right| \\
&\leq 2\|g - g_n\| + \left|\int g_n d\mu_\alpha - \int g_n d\mu\right|
\end{aligned}$$

por tanto

$$\limsup_\alpha \left|\int g d\mu_\alpha - \int g d\mu\right| \leq 2\|g - g_n\|$$

dado que existe una subsucesión $\{g_{n'}\}$ tal que $\|g - g_{n'}\| \to 0$ y para g en $C(S)$ con $\|g\| \leq 1$ se obtiene

$$\limsup_\alpha \left|\int g d\mu_\alpha - \int g d\mu\right| = 0$$

es decir $\mu_\alpha \Rightarrow \mu$ y T es un homeomorfismo.

Ahora probaremos que $T(\mathscr{P}(S))$ es un subconjunto cerrado de I^∞. Supongamos que $\{\mu_n\}$ es una sucesión en $\mathscr{P}(S)$ tal que $T(\mu_n)$ converge a $(\alpha_1, \alpha_2, \dots)$ en I^∞. Sea g en $C_0(S)$ tal que $\|g\| \leq 1$, por tanto existe una sucesión $\{g_{n'}\}$ tal que $\|g_{n'} - g\| \to 0$, por las mismas maniobras que se emplearon en la prueba de la continuidad de T^{-1} se tiene

$$\left|\int g d\mu_n - \int g d\mu_m\right| \leq 2\|g - g_{n'}\| + \left|\int g_{n'} d\mu_n - \int g_{n'} d\mu_m\right|$$

por tanto

$$\limsup_{n,m\to\infty} \left|\int g d\mu_n - \int g d\mu_m\right| \leq 2\|g - g_{n'}\|$$

tomando $n' \to \infty$, se llega a

$$\limsup_{n,m\to\infty}\left|\int g d\mu_n - \int g d\mu_m\right| = 0$$

Para toda g en la bola unitaria S_0 $\lim \int g d\mu$ existe y lo llamaremos Λg. Si $f \in C_0(S)$, existe una cosntante $c \neq 0$ tal que $cf \in C_0(S)$, de esta manera definamos $\Lambda(f) = c\Lambda(f/c)$. Λ es una funcional no negativa en $C_0(S)$ tal que $\Lambda(1) = 1$. Por el Teorema de Representación de Riesz existe una medida μ tal que $\Lambda(g) = \int g d\mu$. En particular $\Lambda(g_k) = \alpha_k$, por lo cual $T(\mu) = (\alpha_1, \alpha_2, \dots)$, es decir $T(\mathscr{P}(S))$ es cerrado, en consecuencia $\mathscr{P}(S)$ es compacto.

Supongamos que $\mathscr{P}(S)$ es un conjunto compacto, por el Teorema 1.9 S es homeomorfo a un subconjunto cerrado de $\mathscr{P}(S)$, por tanto S es compacto. □

2 Espacios C y D

Trataremos el caso particular de dos espacios métricos importantes para el estudio de procesos estocásticos, el primero de ellos es el espacio de las funciones reales continuas con valores en $[0,1]$ con la métrica del supremo, el segundo es el espacio de las funciones reales con valores en $[0,1]$ continuas por la derecha con límite por la izquierda con la métrica de Skorohod. El primero de estos espacios se le conoce como el espacio C y al otro como el espacio D.

Para ambos espacios se exponen criterios de tensión de medidas, que es fundamental para obtener convergencia débil, así como también se caracterizan a los conjuntos compactos. Para el caso del espacio C se prueba la existencia de la Medida de Wiener, que da la descripción del Movimento Browniano, proceso estocástico que R. Wiener formaliza matemáticamente al cues -tionamiento que R. Brown hizo a finales del siglo XIX, sobre la simulación del movimiento que sigue un grano de polen disperso en agua. A partir de la Medida de Wiener se define el Puente Browniano, que jugará un papel importante en los Teoremas de Donsker que en el capítulo 4 se tratarán. Además se presentan las dos versiones del Principio de Invarianza.

Por último se hace una comparación entre convergencia débil de medidas en C con la topología uniforme y en D con la topología de Skorohod.

Los resultados de este capítulo corresponden al libro Billingsley [1968].

2.1 El Espacio C

Denotemos por $C = C[0,1]$ al espacio de las funciones reales continuas definidas en el intervalo $[0,1]$, dotado de la *métrica del supremo:*

$$d(x,y) = \sup_{0\leq t\leq 1} |x(t) - y(t)| \quad x, y \in C$$

Con esta métrica el espacio (C,d) resulta ser un espacio métrico completo y separable.

Para los puntos $t_1, \ldots, t_k$ del intervalo $[0,1]$, designemos por $\pi_{t_1,\ldots,t_k}(x)$ la proyección natural, es decir, la función de C en $\mathbb{R}^k$ dada por $\pi_{t_1,\ldots,t_k}(x) = (x(t_1), \ldots, x(t_k))$; llamemos *conjuntos de dimensión finita* a los subconjuntos de C de la forma $\pi^{-1}_{t_1,\ldots,t_k} H$ con $H \in \mathscr{B}(\mathbb{R}^k)$; si la función $\pi_k : \mathbb{R}^\infty \to \mathbb{R}^k$, es la proyección de las primeras k componentes y P es una medida de probabilidad en $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$, la *distribución de dimensión finita de* P es la medida $P\pi_k^{-1}$ en $(\mathbb{R}^k, \mathscr{B}(\mathbb{R})^k)$, para el caso de $(C, \mathscr{B}(C))$ las distribuciones de dimensión finita estan dadas por $P\pi^{-1}_{t_1,\ldots,t_k}$.

Por ejemplo, consideremos los puntos $0, 1/2, 1$, y la función continua $x(t) = (1-t^2)/2$, entonces $\pi_{0,1/2,1}(x) = (1/2, 3/8, 0)$, y tomando el conjunto $H = \{(u,v,w) \in \mathbb{R}^3 : u^2 + v^2 + w^2 \leq 1\}$, el conjuto de finita dimensión $\pi^{-1}_{0,1/2,1} H$ es el conjunto de funciones continuas tal que al ser evaluadas en los puntos $0, 1/2, 1$, el vector $(x(0), x(1/2), x(1))$ esta en la bola unitaria con centro en el origen, en este caso $x \in \pi^{-1}_{0,1/2,1} H$.

Teorema 2.1 *Sean P_n, P medidas de probabilidad en $(C, \mathscr{B}(C))$. Si las distribuciones de dimensión finita de P_n convergen débilmente a las de P, y si $\{P_n\}$ es un conjunto tenso, entonces $P_n \Rightarrow P$.*

DEMOSTRACIÓN: Dado que $\{P_n\}$ es tensa, toda subsucesión $\{P_{n'}\}$ contiene una subsucesión $\{P_{n''}\}$ que converge débilmente a alguna medida Q. Entonces las distribuciones de dimensión finita de Q son el límite de las correspondientes distribuciones de dimensión finita de $P_{n''}$, y por tanto coinciden con las de P. Como toda medida en C esta totalmente determinada por las distribuciones de dimensión finita tenemos que $Q = P$.

Esto es, toda subsucesión contiene una subsucesión que converge a P, por tanto $P_n \Rightarrow P$. □

Este teorema es el camino a seguir para demostrar la convergencia débil de medidas, por lo que es importante tenerlo en mente. Aquí resalta lo crucial que es la caracterización de los conjuntos relativamente compactos, o los conjuntos de medidas tensas. Que esperamos para empezar a conocer como son estos conjuntos, para el caso particular en que las medidas toman valores en $\mathscr{B}(C)$.

El *módulo de continuidad* de un elemnto x de C se define como:

$$w_x(\delta) = w(x,\delta) = \sup_{|s-t|<\delta} |x(s) - x(t)|, \quad 0 < \delta \leq 1$$

Como toda función continua sobre $[0,1]$ es uniformemente continua, (es decir, dado $\varepsilon > 0$ existe $\delta > 0$ tal que para todo s, t en $[0,1]$ con $|s-t| < \delta$, implica que $|x(s) - x(t)| < \varepsilon$) podemos establecer que para cada x en C

$$\lim_{\delta \to 0} w_x(\delta) = 0$$

Teorema 2.2 *Sea $\{P_n\}$ una sucesión de medidas de probabilidad definidas en $(C, \mathscr{B}(C))$, la sucesión es tensa si y sólo si satisface*

(i) Para cada $\eta > 0$, existe a de manera que

$$P_n\{x : |x(0)| > a\} \leq \eta, \quad n \geq 1$$

(ii) Para cada $\varepsilon, \eta > 0$ existe δ, con $0 < \delta < 1$, y n_0 entero positivo tal que

$$P_n\{x : w_x(\delta) > \varepsilon\} \leq \eta, \quad n \geq n_0.$$

La condición (i) establece que la sucesión de medidas $\{P_n \pi_0^{-1}\}$ en $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ es tensa.

DEMOSTRACIÓN: Supongamos que la sucesión $\{P_n\}$ es tensa; dado η existe K compacto de manera que $P_n(K) > 1 - \eta$ para toda n. Por el Teorema de Arzelà-Ascoli, $K \subset \{x : |x(0)| \leq a\}$ para alguna a, y $K \subset \{x : w_x(\delta) \leq \varepsilon\}$ para alguna δ, de donde concluimos (i) y (ii) con $n_0 = 1$.

En la condición (ii) podemos poner $n_0 = 1$, ya que si P es una medida en $(C, \mathscr{B}(C))$, en tensa, por tanto existe δ tal que $P\{x : w_x(\delta) \geq \varepsilon\} \leq \eta$. Entonces supongamos que $\{P_n\}$ satisface (ii) para δ' y $n \geq n_0$, tomemos n tal que $n < n_0$, entonces existe δ_n tal que $P_n\{x : w_x(\delta_n) \geq \varepsilon\} \leq \eta$ llamemos $\delta = \min\{\delta', \delta_1, \ldots, \delta_{n_0-1}\}$, de aqui se sigue la condición (ii) para $n_0 = 1$.

Supongamos que $\{P_n\}$ satisface (i) y (ii), con $n_0 = 1$; dado η podemos elegir a tal que si $A = \{x : |x(0)| \leq a\}$ entonces $P_n(A^c) \leq \frac{1}{2}\eta$ para toda n, y δ_k de manera que si el conjunto

$$A_k = \left\{x : w_x(\delta_k) < \frac{1}{k}\right\}$$

entonces $P_n(A_k^c) \leq \eta/2^{k+1}$ para toda n. Sea K la cerradura de $A \cap (\cap_{k=1}^{\infty} A_k)$, entonces $P_n(K) \geq 1 - \eta$ para toda n, además por el Teorema de Arzelà-Ascoli, el conjunto K es compacto, de donde $\{P_n\}$ es tensa. □

Ahora veamos otro teorema que nos ayudara con el concepto de tensión de medidas en el espacio C.

Teorema 2.3 *Sea $\{P_n\}$ una sucesión de medidas de probabilidad definidas en $(C, \mathscr{B}(C))$, que satisface* *(i) Para cada $\eta > 0$, existe a de manera que*

$$P_n\{x : |x(0)| > a\} \leq \eta, \quad n \geq 1$$

(ii) Para cada $\varepsilon, \eta > 0$ existe δ, con $0 < \delta < 1$, y n_0 entero positivo tal que

$$\frac{1}{\delta} P_n \left\{ x : \sup_{t \leq s \leq t+\delta} |x(s) - x(t)| \geq \varepsilon \right\} \leq \eta, \quad n \geq n_0.$$

para toda t.

entonces $\{P_n\}$ es tensa.

En el caso en que $t + \delta > 1$, nos limitaremos a $t \leq s \leq 1$.
DEMOSTRACIÓN: Basta demostrar que la condición (ii) implica la condición (ii) del teorema anterior. Para δ fija definamos el conjunto

$$A_t = \left\{ x : \sup_{t \leq s \leq t+\delta} |x(s) - x(t)| \geq \varepsilon \right\}$$

Tomemos una partición uniforme del intervalo $[0, 1]$ en subintervalos de longitud δ, si $|s - t| < \delta$ entonces ambos puntos están en un mismo intervalo o bien en intervalos vecinos, de donde

$$P_n\{x : w_x(\delta) \geq 3\varepsilon\} \leq P_n \left\{ \bigcup_{i < \delta^{-1}} A_{i\delta} \right\} \leq \sum_{i < \delta^{-1}} P_n(A_{i\delta})$$

entonces $P_n\{x : w_x(\delta) \geq 3\varepsilon\} \leq (1 + [1/\delta])\delta\eta < 2\eta\delta < 2\eta$ □

2.1.1 Funciones Aleatorias

Sea X una función de $(\Omega, \mathcal{B}, P)$ a C, esto es, $X(\omega)$ es una función continua en $[0, 1]$ cuyo valor en t se denota por $X(t, \omega)$. Para t fija $X(t)$ denotará una función real en Ω, con valor $X(t, \omega)$ en ω, es decir $X(t)$ es la composición $\pi_t X$; de manera similar $(X(t_1), \ldots, X(t_k))$ es una función de Ω a $\mathbb{R}^k$ con valor $(X(t_1, \omega), \ldots, X(t_k, \omega))$ en ω.

Para t fijo consideremos el conjunto cerrado $A = \{x \in C : x(t) \leq \alpha\}$, entonces $\{\omega : X(t, \omega) \leq \alpha\} = X^{-1}A$, por consiguiente si X es un elemento aleatorio en C, $X(t)$ es una variable aleatoria. Ahora supongamos que para cada t, $X(t)$ es una variable aleatoria; si B es una bola cerrada en C, con centro en x de radio δ, entonces $X^{-1}B = \{\omega : X(\omega) \in B\} = \cap_r\{\omega : -\delta \leq$

$X(r,\omega) - x(r) \leq \delta\}$ donde la intersección corre sobre todos los racionales del intervalo $[0,1]$, entonces $X^{-1}B \in \mathcal{B}$; como C es separable, tenemos que X es un elemento aleatorio en C.

Sea $\{X_n\}$ una sucesión de funciones aleatorias, se dice que la sucesión es tensa si la sucesión de las distribuciones correspondientes $\{PX_n^{-1}\}$ es tensa.

El Teorema 2.2 asegura que la sucesión $\{X_n\}$ es tensa si y sólo si $\{X_n(0)\}$ es tensa y si para cada ε, η positivos existe δ con $0 < \delta < 1$ y n_0 entero positivo tal que

$$P\{w(X_n,\delta) \geq \varepsilon\} \leq \eta \qquad n \geq n_0 \tag{2.1}$$

donde $w(x,\delta)$ representa $w_x(\delta)$. Esta condición restringe el comportamiento de las funciones X_n, es decir, establece que las funciones X_n no tengan oscilaciones muy violentas.

De la misma manera el Teorema 2.3 establece que la sucesión $\{X_n\}$ es tensa si $\{X_n(0)\}$ es tensa y si para cada ε, η positivos existe δ, con $0 < \delta < 1$ y n_0 entero positivo tal que

$$\frac{1}{\delta}P\left\{\sup_{t\leq s\leq t+\delta} |X_n(s) - X_n(t)| \geq \varepsilon\right\} \leq \eta \tag{2.2}$$

para $n \geq n_0$ y $0 \leq t \leq 1$.

Sean $\xi_1, \xi_2, \ldots, \xi_n$ variables aleatorias definidas en $(\Omega, \mathcal{B}, P)$, se define $S_n = \xi_1 + \cdots + \xi_n$ y $S_0 = 0$. Pasamos a definir la variable aleatoria $X_n(t,\omega)$ que es de interés para el desarrollo del espacio C. Para cada $i/n \in [0,1]$ se define $X_n(i/n,\omega)$ como:

$$X_n\left(\frac{i}{n},\omega\right) = \frac{1}{\sigma\sqrt{n}}S_i(\omega)$$

y de manera lineal para el resto de los puntos del dominio, esto es, si $t \in [(i-1)/n, i/n]$, entonces

$$\begin{aligned} X_n(t,\omega) &= \frac{(i/n) - t}{1/n}X_n\left(\frac{i-1}{n}\right) + \frac{t-(i-1)/n}{1/n}X_n\left(\frac{i}{n}\right) \\ &= \frac{1}{\sigma\sqrt{n}}S_{i-1}(\omega) + n\left(t - \frac{i-1}{n}\right)\frac{1}{\sigma\sqrt{n}}\xi_i(\omega) \end{aligned}$$

Si $t \in [(i-1)/n, i/n]$, $i-1 = [nt]$, por tanto la expresión anterior se puede simplificar a

$$X_n(t,\omega) = \frac{1}{\sigma\sqrt{n}}S_{[nt]}(\omega) + (nt - [nt])\frac{1}{\sigma\sqrt{n}}\xi_{[nt]+1}(\omega) \tag{2.3}$$

Como S_i es una variable aleatoria, $X_n(t)$ es una variable aleatoria para cada t, por lo cual X_n es una función aleatoria.

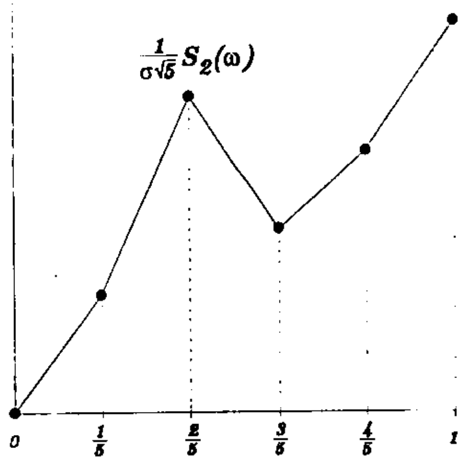

Figura 2.1: Gráfica de X_n, para $n = 5$.

Ahora nos enfocamos a determinar condiciones bajo las cuales las funciones aleatorias definidas anteriormente formarán un conjunto tenso. Pero antes notemos que la sucesión $\{X_n(0)\}$ es tensa, ya que $X_n(0) = 0$ para toda n. Si $t = k/n$ y $t + \delta = j/n$, con k, j enteros la ecuación (2.2) se puede reescribir según la definición de $X_n(t)$ como:

$$\frac{1}{\delta} P \left\{ \max_{i \leq n\delta} \frac{1}{\sigma\sqrt{n}} |S_{k-1} - S_k| \geq \varepsilon \right\} \leq \eta \tag{2.4}$$

claro que ambas difieren por le supuesto de tomar a t y $t + \delta$ como múltiplos de $1/n$; pero para nuestros objetivos esto no representará ninguna desventaja.

Teorema 2.4 *Sea $\{X_n\}$ una sucesión de funciones aleatorias definidas como en (2.3), y tal que para cada ε positivo existe $\lambda > 1$ y n_0 entero positivo, de manera que si $n \geq n_0$, entonces*

$$P \left\{ \max_{i \leq n} |S_{k+i} - S_k| \geq \lambda\sigma\sqrt{n} \right\} \leq \frac{\varepsilon}{\lambda^2} \tag{2.5}$$

para toda k. Entonces $\{X_n\}$ es tensa.

DEMOSTRACIÓN: Si remplazamos ε por $\eta\varepsilon^2$, entonces existe λ positivo y n_1 tal que

$$P\left\{\max_{i\leq n}|S_{k+1}-S_k|\geq \lambda\sigma\sqrt{n}\right\}\leq \frac{\eta\varepsilon^2}{\lambda^2} \tag{2.6}$$

para $n \geq n_1$ y $k \geq 1$. Llamemos $\delta = \varepsilon^2/\lambda^2$, como $\lambda > 1 > \varepsilon$, δ es menor que 1. Sea n_0 un entero mayor que n_1/δ. Si $n \geq n_0$, entonces $[n\delta] \geq n_1$, y por (2.6)

$$P\left\{\max_{i\leq[n\delta]}|S_{k+1}-S_k|\geq \lambda\sigma\sqrt{[n\delta]}\right\}\leq \frac{\eta\varepsilon^2}{\lambda^2}$$

Como $\lambda\sqrt{[n\delta]} \leq \varepsilon\sqrt{n}$ y $\eta\delta = \eta\varepsilon^2/\lambda^2$, llegamos a la ecuación (2.4) para toda k y $n \geq n_0$. Por tanto $\{X_n\}$ es tensa. □

2.1.2 Medida de Wiener

La proyección π_t, definida como $\pi_t(x) = x(t)$ con x elemento de C, es una variable aleatoria en $(C, \mathscr{B}(C))$ que denotaremos por x_t. Si P es una medida de probabilidad en $(C, \mathscr{B}(C))$, $\{x_t : 0 \leq t \leq 1\}$ es un proceso estocástico, donde t es el parámetro de tiempo, x_t son las variables coordenadas y su distribución depende de P.

La *Medida de Wiener* que denataremos por W, es una medida de probabilidad en $(C, \mathscr{B}(C))$ que satisface las siguientes dos condiciones:

- Para cada t, la variable aleatoria x_t se distribuye normal con media 0 y varianza t bajo W:

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

  y para $t = 0$, $W\{x_0 = 0\} = 1$.

- El proceso estocástico $\{x_t : 0 \leq t \leq 1\}$ tiene incrementos independientes bajo W: Si $0 \leq t_0 \leq t_1 \leq \cdots \leq t_k \leq 1$ entonces las variables aleatorias

  $$x_{t_1} - x_{t_0}, x_{t_2} - x_{t_1}, \ldots, x_{t_k} - x_{t_{k-1}},$$

  son independientes bajo W.

Por las caracteristicas antes mencionadas, resulta que el proceso $\{x_t\}$ tiene incrementos estacionarios, dado que la distribución de $x_t - x_s$ depende unicamente de $t - s$.

Probaremos la existencia de la medida de Wiener en $(C, \mathscr{B}(C))$.

Teorema 2.5 *Existe en* $(C, \mathcal{B}(C))$ *una medida de probabilidad* W, *con los requerimientos antes mencionados.*

DEMOSTRACIÓN: Sean $\xi_1, \xi_2, \ldots$ variables aleatorias independientes, con distribución normal, media 0 y varianza 1, definidas en un mismo espacio de probabilidad $(\Omega, \mathcal{B}, P)$. Sea X_n una función aleatoria definida como (2.3) con $\sigma = 1$, es decir:

$$X_n(t,\omega) = \frac{1}{\sqrt{n}} S_{[nt]}(\omega) + (nt - [nt]) \frac{1}{\sqrt{n}} \xi_{[nt]+1}(\omega) \tag{2.7}$$

Sea P_n la distribución de la variable aleatoria X_n en C. Para $t_1, \ldots, t_k$ en $[0,1]$ consideramos $P_n \pi^{-1}_{t_1,\ldots,t_k}$ la distribución de dimensión finita, que corresponde a la distribución del vector aleatorio $(X_n(t_1), \ldots, X_n(t_k))$. Para t fijo, la variable aleatoria X_n se distribuye normal con media cero, y varianza

$$\frac{[nt]}{n} + \frac{(nt - [nt])^2}{n}$$

que converge a t, cuando $n \to \infty$, por tanto $X_n(t) \Rightarrow N(0,t)$ Sin ningún inconveniente podemos considerar más de un punto, para concluir que las distribuciones de dimensión finita convergen débilmente a las correspondientes distribuciones de dimensión finita de W. Resta demostrar que $\{P_n\}$ es tensa, esto será con el apoyo del Teorema 2.4. Consideremos los conjuntos ajenos

$$E_i = \left\{ \max_{j<i} |S_j| < 2\lambda\sqrt{n} \le |S_i| \right\} \tag{2.8}$$

con $\lambda > 0$, por ser ξ_i independientes y con incrementos estacionarios llegamos a

$$\begin{aligned} P\left\{ \max_{i \le n} |S_i| \ge 2\lambda\sqrt{n} \right\} &\le P\{|S_n| \ge \lambda\sqrt{n}\} + \\ &\quad \sum_{i=1}^{n-1} P(E_i \cap \{|S_n - S_i| \ge \lambda\sqrt{n}\}) \\ &= P\{|S_n| \ge \lambda\sqrt{n}\} + \\ &\quad \sum_{i=1}^{n-1} P(E_i) P\{|S_{n-i}| \ge \lambda\sqrt{n}\} \\ &\le P\{|S_n| \ge \lambda\sqrt{n}\} + \\ &\quad \sum_{i=1}^{n-1} P(E_i) P\{|S_{n-i}| \ge \lambda\sqrt{n-i}\} \end{aligned} \tag{2.9}$$

por la desigualdad de Chebyshev, con a el tercer momento de una distribución normal con media 0, y varianza 1, se tiene

$$P\left\{\max_{i\leq n}|S_i| \geq 2\lambda\sqrt{n}\right\} \leq \frac{a}{\lambda^3} + \sum_{i=1}^{n-1} P(E_i)\frac{a}{\lambda^3} \leq \frac{2a}{\lambda^3}$$

para toda n. Para cualquier ε positivo, elegimos $\lambda > 1$ tal que $2a/\lambda < \varepsilon$, para obtener así

$$P\left\{\max_{i<n}|S_i| \geq 2\lambda\sqrt{n}\right\} \leq \frac{\varepsilon}{\lambda^2}$$

por tanto $\{P_n\}$ es tensa según el Teorema 2.4. □

En efecto la medida de Wiener existe. Indistintamente usaremos W para denotar un elemnto aleatorio con valores en C cuya distribución es la medida de Wiener, asi como también lo hemos hecho la medida medida de Wiener en C. Llamaremos *proceso de Wiener* a $\{W_t, 0 \leq t \leq 1\}$ donde $W_t(\omega)$ es el valor en t de la función aleatoria $W(\omega)$.

2.1.3 El Puente Browniano

Un elemento aleatorio X en C es *Gaussiano*, si todas sus distribuciones de dimensión finita son normales. La distribución de un elemento Gaussiano en C está totalmente determinada por su media $E\{X(t)\}$, y por el producto de momentos $E\{X(s)X(t)\}$, para $0 \leq s,t \leq 1$, ya que éstos determinan las distribuciones de dimensión finita. Para W estos son

$$E\{W_t\} = 0$$

y como $W_t - W_s$ se distribuye normal con media cero y varianza $t-s$, $E\{(W_t - W_s)^2\} = t - s$ de donde

$$E\{W_sW_t\} = s \qquad \text{si } s \leq t$$

Para el estudio las distribuciones empíricas sera necesario el elemento aleatorio W°, llamado *Puente Browniano* o *Browniano Atado* que se define como:

$$W_t^\circ = W_t - tW_1 \quad \text{con } t \in [0,1]$$

que es un elemento Gaussiano en C, y satisface $E\{W_t^\circ\} = 0$, $E\{W_s^\circ W_t^\circ\} = s(1-t)$ con $s \leq t$, además $W_0^\circ = W_1^\circ = 0$ de donde biene el nombre de Puente Browniano. Si la función $h : C \to C$ es tal a cada x le asigna el valor $x(t) - tx(1)$ en t, entonces las medidas W° y W estan relacionadas por $W^\circ = Wh^{-1}$.

2.1.4 El Principio de Invarianza

En esta sección se expone el Principo de Invarianza, que habla sobre convergencia en distribución de las variables aletorias X_n, antes de presentarlo enunciamos un lema que sera necesario para la prueba del Principio de Invarianza.

Sean $\xi_1, \xi_2, \ldots$ variables aleatorias definidas en $(\Omega, \mathscr{B}, P)$ y $S_n = \xi_1 + \cdots + \xi_n$, de nueva cuenta definamos el elemento aleatorio X_n en C como

$$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}(\omega) \tag{2.10}$$

Lema 2.1 *Sean $\xi_1, \ldots, \xi_n$ variables aleatorias independintes con media cero y varianza finita σ_i^2, sea $S_i = \xi_1 + \cdots + \xi_i$, y $s_i = \sigma_1^2 + \cdots + \sigma_i^2$, entonces*

$$P\left\{\max_{i \le m} |S_i| \ge \lambda s_m\right\} \le 2P\left\{|S_m| \ge (\lambda - \sqrt{2}) s_m\right\}$$

DEMOSTRACIÓN: Supongamos que $\lambda > \sqrt{2}$ de otro caso no hay nada que demostrar, consideremos los conjuntos

$$E_i = \left\{\max_{j<i} |S_j| < \lambda s_m \le |S_i|\right\}.$$

al igual que en la prueba de la existencia de la medida de Wiener calculamos

$$\begin{aligned} P\left\{\max_{i \le m} |S_i| \ge \lambda s_m\right\} \quad \le \quad & P\{|S_m| \ge (\lambda - \sqrt{2}) s_m\} + \\ & \sum_{i=1}^{m-1} P(E_i \cap \{|S_m| < (\lambda - \sqrt{2}) s_m\}) \end{aligned} \tag{2.11}$$

Si $|S_i| \ge \lambda s_m$ y $|S_m| < (\lambda - \sqrt{2}) s_m$, se disponde de la desigualdad $|S_m - S_i| \ge \sqrt{2} s_m$, por la desigualdad de Chebyshev podemos acotar superiormente el segundo sumando de (2.11) de la forma siguiente

$$\begin{aligned} \sum_{i=1}^{m-1} P(E_i) P\{|S_m - S_i| \ge \sqrt{2} s_m\} \quad & \le \quad \sum_{i=1}^{m-1} P(E_i) \frac{1}{2 s_m^2} \sum_{k=i+1}^{m} \sigma_k^2 \\ & \le \quad \frac{1}{2} \sum_{i=1}^{m-1} P(E_i) \\ & \le \quad \frac{1}{2} P\left\{\max_{i \le m} |S_i| \ge \lambda s_m\right\} \end{aligned}$$

finiquitando esta demostración. □

El próximo teorema es conocido como el Teorema de Donsker y como el Principio de Invarianza, al hacer referencia a él lo haremos por el Principio de Invarianza ya que existe otro teorema que también lleva el nombre de Donsker, el cual es de mayor interés para nuestro objetivo, de él se hablara ampliamente en el capítulo 4.

Teorema 2.6 (Principio de Invarianza) *Sean $\xi_1, \xi_2, \ldots$ variables aleatorias independientes e idénticamente distribuídas con media cero y varianza finita σ^2 positiva. Entonces*

$$X_n \Rightarrow W \tag{2.12}$$

donde las funciones aleatorias X_n están definidas como en (2.10).

DEMOSTRACIÓN: Primero demostremos que las distribuciones de dimensión finita convergen a las deseadas. Consideremos un sólo punto s en $[0,1]$, por la definición de X_n se obtiene la desigualdad

$$\left| X_n(s) - \frac{1}{\sigma\sqrt{n}} S_{[ns]} \right| \leq \frac{1}{\sigma\sqrt{n}} \xi_{[ns]+1} \tag{2.13}$$

donde la cota superior converge a cero en probabilidad por la útil desigualdad de Chebyshev. Por el Teorema del Límite Central y el hecho que $[ns]/n \to s$

$$\frac{1}{\sigma\sqrt{n}} S_{[ns]} \Rightarrow W_s$$

junto con (2.13) se tiene $X_n(s) \Rightarrow W_s$.

Ahora consideremos dos puntos s, t en $[0,1]$ con $s < t$, deseamos probar que $(X_n(s), X_n(t)) \Rightarrow (W_s, W_t)$. Como las variables aleatorias $S_{[ns]}$ y $S_{[nt]} - S_{[ns]}$ son independientes, nuevamente por el Teorema del Límite Central y la desigualdad (2.13) obtenemos la convergencia

$$\left(\frac{1}{\sigma\sqrt{n}} S_{[ns]}, \frac{1}{\sigma\sqrt{n}} (S_{[nt]} - S_{[ns]}) \right) \Rightarrow (W_s, W_t - W_s).$$

De igual manera se puede tomar un conjunto con tres o más puntos, logrando la convergencia de las distribuciones de dimensión finita. Para comprobar que $\{X_n\}$ es tensa nos basaremos en el Lema 2.1. Para $\lambda > 2\sqrt{2}$ gozamos de la desigualdad $P\{\max_{i\leq n} |S_i| \geq \lambda\sigma\sqrt{n}\} \leq 2P\{|S_n| \geq \frac{1}{2}\lambda\sigma\sqrt{n}\}$, por el Teorema del Límite Central se obtiene la convergencia

$$P\left\{ |S_n| \geq \frac{1}{2}\lambda\sigma\sqrt{n} \right\} \to P\left\{ |N| \geq \frac{1}{2}\lambda \right\} \leq \frac{8}{\lambda^3} E|N|^3$$

entonces para λ suficientemente grande se obtiene

$$\limsup_{n\to\infty} P\left\{\max_{i\leq n}|S_i| \geq \lambda\sigma\sqrt{n}\right\} < \frac{\varepsilon}{\lambda^2}$$

junto con el Toerema 2.4, se tiene la propiedad de tensión para $\{X_n\}$. □

A este teorema se le llama el Principio de Invarianza, por que no interesa la distribución de las variables aleatorias ξ_i, salvo que tengan media cero y varianza finita. Por ello si se desea calcular una distribución límite, se toman variables aleatorias sencillas de manipular y después se aplica el Principio de Invarianza.

2.1.5 Producto de Momentos

Supongamos que existen enteros no negativos $u_1, \ldots, u_m$ que satisfacen

$$E\{|S_j - S_i|^\gamma \cdot |S_k - S_j|^\gamma\} \leq \left(\sum_{i\leq l\leq j} u_l\right)^\alpha \left(\sum_{j\leq l\leq k} u_l\right)^\alpha \tag{2.14}$$

con $0 \leq i \leq j \leq k \leq m$, γ y α positivos. Por ejemplo para $\gamma = 2$, $\alpha = 1$ y ξ_i indepéndientes con media cero, tomamos u_i como la varianza de ξ_i satisfacen (2.14). Dado que $xy \leq (x+y)^2$ para x, y no negativos, de la desigualdad anterior llegamos a

$$E\{|S_j - S_i|^\gamma \cdot |S_k - S_j|^\gamma\} \leq \left(\sum_{i\leq l\leq k} u_l\right)^{2\alpha} \tag{2.15}$$

Si $|S_j - S_i| \geq \lambda$ y $|S_k - S_j| \geq \lambda$ para λ positiva, entonces $|S_j - S_i|^\gamma \cdot |S_j - S_i|^\gamma \geq \lambda^{2\gamma}$, por la desiguladad de Chebyshev y (2.15) se tiene

$$P\{|S_j - S_i| \geq \lambda, |S_k - S_j| \geq \lambda\} \leq \frac{1}{\lambda^{2\gamma}} \left(\sum_{i\leq l\leq k} u_l\right)^{2\alpha} \tag{2.16}$$

con $0 \leq i \leq j \leq k \leq m$.

El siguiente teorema representa el primer paso para el Teorema de Donsker que el capítulo 5 se estudiará.

Teorema 2.7 *Si $\gamma \geq 0$, $\alpha > \frac{1}{2}$ y se satisface (2.16) para λ positiva y*

$$M'_m = \max_{0 \leq k \leq m} \min\{|S_k|, |S_m - S_k|\}$$

entonces

$$P\{M'_m \geq \lambda\} \leq \frac{K_{\gamma,\alpha}}{\lambda^{2\gamma}}(u_1 + \cdots + u_m)^{2\alpha} \tag{2.17}$$

para toda λ positiva y $K_{\gamma,\alpha}$ una constante que depende solamente de γ y α.

DEMOSTRACIÓN: Sea $\delta = 1/(2\gamma + 1)$, $0 \leq \delta \leq 1$, como $1 < 2^{(2\alpha-1)\delta}$ se elige K lo suficientemente grande tal que ($K > 1$)

$$2^{\delta}\left[\frac{1}{2^{2\alpha\delta}} + \frac{1}{K^{\delta}}\right] \leq 1$$

Una vez establecido K, probaremos la desigualdad (2.17) por inducción sobre m. Para $m = 2$, $M'_2 = \min\{|S_1|, |S_2 - S_1|\}$, por (2.16) contamos con

$$\begin{aligned} P\{M'_2 \geq \lambda\} &\leq P\{|S_1| \geq \lambda, |S_2 - S_1| \geq \lambda\} \\ &\leq \frac{1}{\lambda^{2\gamma}}(u_1 + u_2)^{2\alpha} \\ &\leq \frac{K}{\lambda^{2\gamma}}(u_1 + u_2)^{2\alpha} \end{aligned}$$

Supongamos que la desigualdad (2.17) se satisface para todo entero positivo menor que m. Llamemos $u = u_1 + \cdots + u_m$ y supongamos $u > 0$, sea h un entero $1 \leq h \leq m$ tal que

$$\frac{u_1 + \cdots + u_{h-1}}{u} \leq \frac{1}{2} \leq \frac{u_1 + \cdots + u_h}{u}$$

en el caso $h = 1$, el lado izquierdo se considera como 0; ahora bien consideremos las siguientes variables aleatorias:

$$\begin{aligned} U_1 &= \max_{0 \leq i \leq h-1} \min\{|S_i|, |S_{h-1} - S_i|\} \\ U_2 &= \max_{h \leq j \leq m} \min\{|S_j - S_h|, |S_m - S_j|\} \\ D_1 &= \min\{|S_{h-1}|, |S_m - S_{h-1}|\} \\ D_2 &= \min\{|S_h|, |S_m - S_h|\} \end{aligned}$$

Dado que $h-1<m$, contamos con la hipótesis de inducción, esto es, para las variables aleatorias $\xi_1,\dots,\xi_{h-1}$ y $u_1,\dots,u_{h-1}$ se tiene la desigualdad

$$P\{U_1\geq\lambda\}\leq\frac{K}{\lambda^{2\gamma}}(u_1+\cdots+u_{h-1})^{2\alpha}\leq\frac{u^{2\alpha}}{\lambda^{2\gamma}}\frac{K}{2^{2\alpha}} \tag{2.18}$$

como $m-h<m$, de nueva cuenta se hace uso de la hipótesis de inducción para las variables aleatorias $\xi_{h+1},\dots,\xi_m$ para obtener

$$P\{U_2\geq\lambda\}\leq\frac{K}{\lambda^{2\gamma}}(u_{h+1}+\cdots+u_m)^{2\alpha}\leq\frac{u^{2\alpha}}{\lambda^{2\gamma}}\frac{K}{2^{2\alpha}}$$

Por la desigualdad (2.16) contamos con

$$P\{D_1\geq\lambda\}\leq\frac{1}{\lambda^{2\lambda}}(u_1+\cdots+u_m)^{2\alpha}\leq\frac{u^{2\alpha}}{\lambda^{2\gamma}} \tag{2.19}$$

y con $P\{D_2\geq\lambda\}\leq u^{2\alpha}/\lambda^{2\gamma}$. Para $0\leq i\leq h-1$ probaremos que $\mu_i=\min\{|S_i|,|S_m-S_i|\}\leq U_1+D_1$.

- Para el caso en que $|S_i|\leq U_1$ se tiene $\mu_i\leq|S_i|\leq U_1+D_1$.
- Supongamos $|S_{h-1}-S_i|\leq U_1$, si $|S_{h-1}|=D_1$ entonces

$$\mu_i\leq|S_i|\leq|S_{h-1}-S_i|+|S_{h-1}|\leq U_1+D_1$$

- Supongamos que $|S_{h-1}-S-i|\leq U_i$ y $|S_m-S_{h-1}|=D_1$, entonces

$$\mu_i\leq|S_m-S_i|\leq|S_{h-1}-S_i|+|S_m-S_h|\leq U_1+D_1$$

De esta manera contamos con $\min\{|S_i|,|S_m-S_i|\}\leq U_1+D_1$ para $0\leq i\leq h-1$. De manera ánaloga se tiene también que $\min\{|S_j|,|S_m-S_j|\}\leq U_2+D_2$ con $h\leq j\leq m$; de estas dos últimas desigualdades resulta

$$M'_m\leq\max\{U_1+U_1,U_2+D_2\}$$

por consiguiente $P\{M'_m\geq\lambda\}\leq P\{U_1+D_1\geq\lambda\}+P\{U_2+D_2\geq\lambda\}$. Sean λ_1,λ_2 núemros positivos tales que $\lambda_1+\lambda_2=\lambda$, por las desigualdades (2.18) y (2.19) se llega a

$$P\{U_1+D_1\geq\lambda\}\leq P\{U_1\geq\lambda_0\}+P\{D_1\geq\lambda_1\}\leq\frac{u^{2\alpha}}{\lambda_0^{2\gamma}}\frac{K}{2^{2\alpha}}+\frac{u^{2\alpha}}{\lambda_1^{2\gamma}}$$

Sean C_0, C_1 números positivos, entonces

$$\min_{\lambda_0,\lambda_1>0}\left[\frac{C_0}{\lambda_0^{2\gamma}}+\frac{C_1}{\lambda_1^{2\gamma}}\right]=\frac{1}{\lambda^{2\gamma}}[C_0^\delta+C_1^\delta]^{1/\delta}\qquad\text{con } \lambda_0+\lambda_1=\lambda$$

por lo cual, tomando $C_0 = K/2^{2\alpha}$ y $C_1 = 1$

$$P\{U_1+D_1\geq\lambda\}\leq\frac{u^{2\alpha}}{\lambda^{2\gamma}}\left[\left(\frac{K}{2^{2\alpha}}\right)^\delta+1\right]^{1/\delta}$$

al igual se puede establecer para la variable aleatoria $U_2 + D_2$

$$P\{M_m'\geq\lambda\}\leq\frac{u^{2\alpha}}{\lambda^{2\gamma}}2\left[\left(\frac{K}{2^{2\alpha}}\right)^\delta+1\right]^{1/\delta}$$

por la manera en que se eligio K, el término de la derecha se puede acotar por el merecido $Ku^{2\alpha}/\lambda^{2\gamma}$, dando fin a la prueba del teorema. □

Por último mostramos un teorema que sera utilizado en la el la prueba del Teorema de Donsker.

Teorema 2.8 *Sea $\gamma \geq 0$, $\alpha > 1/2$ y*

$$P\{|S_j-S_i|\geq\lambda,|S_k-S_j|\geq\lambda\}\leq\frac{1}{\lambda^{2\gamma}}\left(\sum_{i\leq l\leq k}u_l\right)^{2\alpha}$$

se satisface para toda λ positiva con $0 \leq i \leq j \leq k \leq m$, entonces para toda λ positiva se tiene

$$P\{M_m''\geq\lambda\}\leq\frac{K_{\gamma,\alpha}''}{\lambda^{2\gamma}}(u_1+\cdots+c_m)^{2\alpha}$$

donde $K''_{\gamma,\alpha}$ es una constante que depende solamente de γ y α.

2.2 El Espacio D

En esta sección también trabajaremos es un espacio de funciones, pero con menos restricciones que el espacio C, a saber, el conjunto de las funciones reales definidas en $[0, 1]$ continuas por la derecha con límite por la izquierda, es decir, con funciones definidas en $[0, 1]$ que satisfacen:

(*i*) $x(t+) = \lim_{s \to t^+} x(s)$, existe para $0 \leq t < 1$, y $x(t+) = x(t)$

(*ii*) $x(t-) - \lim_{s \to t^-} x(s)$, existe para $0 < t \leq 1$

A este conjunto lo denotaremos por $D = D[0,1]$, dichas funciones también reciben el nombre de funciones con discontinuidades de primer clase.

Para $x \in D$ y $T_0 \subset [0,1]$, se define

$$w_x(T_0) = \sup\{|x(s) - x(t)| : s, t \in T_0\}$$

entonces el módulo de continuidad de un elemento x de C se puede expresar como:

$$w_x(\delta) = \sup_{0 \leq t \leq 1-\delta} w_x[t, t+\delta]$$

Si x es elemento de C, es uniformemente continua, el siguiente lema es un resultado similar pero para el espacio D.

Lema 2.2 *Sea x un elemento del espacio D y $\varepsilon > 0$, entonces existen puntos $t_0, \ldots, t_r$ tal que $0 = t_0 < t_1 < \cdots < t_r = 1$ y*

$$w_x[t_{i-1}, t_i) < \varepsilon \tag{2.20}$$

DEMOSTRACIÓN: Sea

$$\tau = \sup\{t \in [0,1] : [0,t) = \cup_{i=1}^{r}[t_{i-1}, t_i), \text{con}$$
$$0 = t_0 < \cdots < t_r = t \text{ y } w_x[t_{i-1}, t_i) < \varepsilon\}$$

Dado que $x(0) = x(0+)$, τ está bien definido y es positivo. Como $x(\tau-)$ existe, τ es el máximo del conjunto, y como $x(\tau) = x(\tau+)$, $\tau = 1$. □

De este lema nos podemos dar cuenta que los elementos de D tienen a lo más un número finito de saltos mayor que ε positivo, como consecuencia tienen a lo más un número numerable de discontinuidades, además son acotados y por último se pueden aproximar por funciones simples, con lo que aseguramos que las funciones de D son funciones medibles.

Como un intento de obtener resultados análogos a los que se obtuvieron en el estudio del espacio C, empezaremos con la definición de *módulo de continuidad* en D. Para $0 < \delta < 1$ y x en D sea

$$w'_x(\delta) = \inf_{\{t_i\}} \max_{0 < i \leq r} w_x[t_{i-1}, t_i)$$

donde el ínfimo es sobre el conjunto finito $\{t_i\}$ tal que

$$\begin{aligned} &0 = t_0 < t_1 < \cdots < t_r = 1 \qquad \text{y} \\ &t_i - t_{i-1} > \delta \qquad i = 1, 2, \ldots, r \end{aligned}$$

Con esta definición el Lema 2.2 establece que todo miembor x de D satisface

$$\lim_{\delta \to 0} w'_x(\delta) = 0$$

Para $\delta < 1/2$, es posible dividir el intervalo $[0, 1)$ en subintervalos $[t_{i-1}, t_i)$ tal que $\delta < t_i - t_{i-1} \leq 2\delta$, para obtener la relación

$$w'_x(\delta) \leq w_x(2\delta)$$

Sea $x \in C$, y $\varepsilon > 0$, tomemos una partición $\{t_i\}$ tal que $0 = t_0 < \cdots < t_r = 1$, $t_i - t_{i-1} > \delta$ y

$$\max_{0 < i \leq r} w_x[t_{i-1}, t_i) < w'_x(\delta) + \varepsilon \tag{2.21}$$

si s, t son puntos tales que $|s - t| < \delta$, entonces ambos están en el intervalo $[t_{i-1}, t_i)$ ó en intervalos vecinos, por ser x continua y por la desigualdad (2.21), $w_x(\delta) \leq 2w'_x(\delta) + 2\varepsilon$ de donde

$$w_x(\delta) \leq 2w'_x(\delta)$$

esto es, los módulos w_x y w'_x son casi iguales para las funciones continuas.

2.2.1 La Topología de Skorohod

Sea Λ el conjunto de las funciones continuas estrictamente crecientes del intervalo $[0, 1]$ sobre si mismo, por tanto si λ pertenece a Λ, $\lambda 0 = 0$ y $\lambda 1 = 1$. Para x, y en D definamos $d(x, y)$ como el ínfimo sobre $\varepsilon > 0$ tal que existe $\lambda \in \Lambda$ que satisface

$$\sup_t |\lambda t - t| \leq \varepsilon \qquad \text{y} \qquad \sup_t |x(t) - y(\lambda t)| \leq \varepsilon$$

Como todo elemento de D es acotado, tomando $\lambda \equiv t$ se tiene que d es una función finita.

Teorema 2.9 *(D, d) es un espacio métrico.*

DEMOSTRACIÓN: Por definición d es no negativa. Supongamos que $d(x, y) = 0$, tomando $\lambda \equiv t$ se tiene

$$\inf\{\varepsilon > 0 : \sup_t |x(t) - y(t)| \leq \varepsilon\} = 0$$

dado $n > 0$, existe ε_n tal que $0 \leq \varepsilon_n < 1/n$ y sobre todo $\sup |x(t) - y(t)| \leq \varepsilon_n$ para cada n, por tanto $x(t) = y(t)$, es decir, $x = y$.

Sea $\lambda \in \Lambda$, entonces λ es una función biyectiva por lo que podemos hablar de λ^{-1} sin inconveniente, más aún $\lambda^{-1} \in \Lambda$, como λ y λ^{-1} son funciones simétricas respecto a la función identidad, se tiene

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

Dado t en $[0, 1]$, sea $t_0 = \lambda t$, entonces $x(t) - y(\lambda t) = x(\lambda^{-1} t_0) - y(t_0)$, como λ es una función biyectiva

$$\sup_t |x(t) - y(\lambda t)| = \sup_t |x(\lambda^{-1} t) - y(t)|$$

asegurando la simetría de d.

Sean $\lambda_1, \lambda_2 \in \Lambda$, entonces $\lambda_2\lambda_1$ también pertenece a Λ, y de las siguientes desigualdades

$$\sup_t |\lambda_2\lambda_1 t - t| \leq \sup_t |\lambda_1 t - t| + \sup_t |\lambda_2 t - t|$$

y

$$\sup_t |x(t) - y(\lambda_2\lambda_1 t)| \leq \sup_t |x(t) - y(\lambda_1 t)| + \sup_t |x(t) - y(\lambda_2 t)|$$

se logra la desigualdad del triángulo para d. □

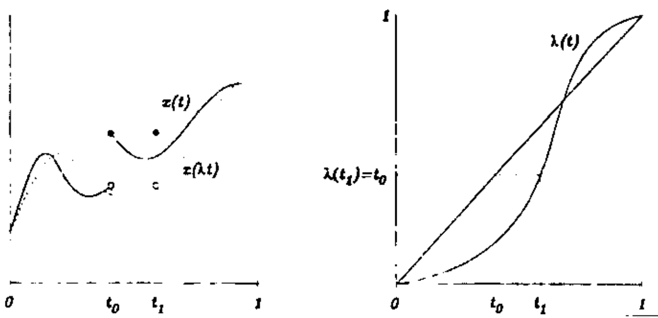

Figura 2.2

La topología que genera la métrica d se le llama *topología de Skorohod.* Diremos que una sucesión $\{x_n\}$ en D converge a x en la topología de Skorohod si y sólo si existen funciones $\lambda_n \in \Lambda$ tal que

$$\lim_n x_n(\lambda_n t) = x(t) \quad \text{y} \quad \lim_n \lambda_n t = t$$

uniformemente en t ambos límites. Si $\{x_n\}$ converge uniformemente a x, entonces converge en la topología de Skorohod, basta tomar $\lambda_n t \equiv t$ para toda n. Pero el recíproco no es cierto, ya que en la topología de Skorohod se tiene la convergencia

$$1\{[0, 1/2 + 1/n)\} \to 1\{[0, 1/2)\}$$

pero dicha sucesión no converge uniformemente.

Sin embargo dado que

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

convergencia en la topología de Skorohod implica $x_n(t) \to x(t)$ en todo punto de continuidad t de x; si x es uniformemente continua entonces convergencia en la topología de Skorohod implica convergecia uniforme. Por tanto, la topología de Skorohod restringida al espacio C coincide con la topología generada por la métrica uniforme en C, es decir, con la topología uniforme.

Consideremos el conjunto de las funciones x con valor racional en $t = 1$, y valor racional en cada uno de los intervalos $[(i-1)/k, i/k)$ para algún entero k con $0 < i \leq k$, este conjunto es un conjunto numerable, y como todo elemento y de D es medible, entonces y se puede aproximar por funciones simples, es decir, por funciones que pertenecen al conjunto antes descrito, resultando (D, d) un espacio métrico separable.

Pero no todo es de agrado, pués D no es completo bajo la métrica d, consideremos la sucesión $x_n = 1\{[1/2, 1/2 + 1/n)\}$, entonces $d(x_n, x_m) \leq |n^{-1} - m^{-1}|$, es decir $\{x_n\}$ es fundamental pero no es convergente bajo la métrica d. Para solucionar este problema introduciremos otra métrica d_0 equivalente a d, bajo la cual D resultará un espacio completo.

Sea λ una función no decreciente en $[0, 1]$ con $\lambda 0 = 0$ y $\lambda 1 = 1$, definamos

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

Si $\|\lambda\|$ es finito tenemos que las pendientes de las cuerdas de λ están acotadas por cero e infinito, por consiguiente λ es continua y creciente. Se puede dar el caso que $\|\lambda\|$ sea infinito y λ pertenesca a Λ, esto justificará la definición de d_0 que mencionamos a continuación.

Para x, y en D sea $d_0(x,y)$ el ínfimo de los $\varepsilon > 0$ tal que existe λ en Λ que satisface

$$\|\lambda\| \leq \varepsilon \quad \text{y} \quad \sup_t |x(t) - y(\lambda t)| \leq \varepsilon \tag{2.22}$$

Nuevamente toamando $\lambda t \equiv t$, resulta que la función d_0 es finita y por las siguientes relaciones tenemos que d_0 es una métrica en D.

$$\|\lambda^{-1}\| = \|\lambda\| \quad \text{y} \quad \|\lambda_2\lambda_1\| \leq \|\lambda_1\| + \|\lambda_2\|$$

Con d_0 como métrica en D, que pasa ahora con la sucesión $x_n = 1\{[1/2, 1/2 + 1/n)\}$. Resulta que para $m, n > 3$

$$d_0(x_n, x_m) = \min\left\{1, \left|\log \frac{m}{n}\right|\right\}$$

es decir, la sucesión no es fundamental bajo d_0, obteniendo un poco de tranquilidad por el momento.

Si $d_0(x,y) < \varepsilon < \frac{1}{4}$ y λ en Λ satisface (2.22) se tiene

$$\log(1-2\varepsilon) < -\varepsilon \leq \log \frac{\lambda t}{t} \leq \varepsilon < \log(1+2\varepsilon)$$

como $0 \leq t \leq 1$, $|\lambda t - t| \leq 2\varepsilon$, es decir

$$d(x,y) \leq 2d_0(x,y) \qquad \text{si } d_0(x,y) < \frac{1}{4} \tag{2.23}$$

Lema 2.3 *Sean x, y en D tal que $d(x,y) < \delta^2$, con $0 < \delta < \frac{1}{4}$, entonces $d_0(x,y) \leq 4\delta + w'_x(\delta)$.*

Demostración: Sea $\delta < 1/2$ y $\{t_i\}$ una partición del intervalo $[0,1]$ tal que $t_i - t_{i-1} > \delta$, con $1 \leq i \leq r$, y

$$w_x[t_{i-1}, t_i) < w'_x(\delta) + \delta \tag{2.24}$$

y μ en Λ tal que

$$\sup_t |x(t) - y(\mu t)| < \delta^2 \quad \text{y} \quad \sup_t |\mu t - t| < \delta^2 \tag{2.25}$$

Definamos una función λ como $\lambda t_i = \mu t_i$ y lineal un los puntos intermedios, resultando λ una función continua y creciente, dado que $\mu^{-1}\lambda$ fija a los puntos

t_i y es creciente, los puntos t y $\mu^{-1}\lambda t$ estan en un mismo intervalo $[t_{i-1}, t_i)$. Por consiguiente

$$\begin{aligned} |x(t) - y(\lambda t)| &\leq |x(t) - x(\mu^{-1}\lambda t)| + |x(\mu^{-1}\lambda t) - y(\lambda t)| \\ &< w'_x(\delta) + \delta + \delta^2 \\ &< 4\delta + w'_x(\delta) \end{aligned}$$

Dado que $\lambda t_i = \mu t_i$ por (2.25)

$$|(\lambda t_i - \lambda t_{i-1}) - (t_i - t_{i-1})| < 2\delta^2 < 2\delta(t_i - t_{i-1})$$

como λ es creciente, para todos los puntos s, t se da la desigualdad $|(\lambda t - \lambda s) - (t - s)| < 2\delta|t - s|$. Con esto llegamos a

$$\log(1 - 2\delta) \leq \log \frac{\lambda t - \lambda s}{t - s} \leq \log(1 + 2\delta)$$

de donde $\|\lambda\| \leq 4\delta$ y $d_0(x, y) \leq 4\delta + w'_x(\delta)$. □

Ahora si, seamos testigos de la prueba de que las métricas d y d_0 generan la misma topología en D.

Teorema 2.10 *Las métricas d y d_0 sobre el espacio D son equivalentes.*

DEMOSTRACIÓN: Denotemos por $S_d(x, \varepsilon)$ y $S_{d_0}(x, \varepsilon)$ la bola abierta con centro en x de radio ε con respecto a la métrica d y d_0 respectivamente. Por (2.23), es posible encontrar δ tal que $S_{d_0}(x, \delta)$ este contenida en $S_d(x, \varepsilon)$.

El Lema 2.3 establece que si

$$\delta < 1/4 \quad \text{y} \quad 4\delta + w'_x(\delta) < \varepsilon \tag{2.26}$$

entonces $S_d(x, \delta^2) \subset S_{d_0}(x, \varepsilon)$. Dados x y ε, existe δ que satisface (2.26), ya que $w'_x(\delta) \to 0$ cuando δ tiene a cero. Por tanto d y d_0 son métricas equivalentes en D. □

2.2.2 La Completitud de D

Teorema 2.11 *El espacio métrico (D, d_0) es completo.*

DEMOSTRACIÓN: Sea $\{x_n\}$ una sucesión fundamental en D, que satisface $d(x_n, x_{n+1}) < 2^{-n}$, y μ_n funciones en Λ tal que

$$\sup_t |x_n(t) - x_{n+1}(\mu_n t)| < \frac{1}{2^n} \quad \text{y} \quad \|\mu_n\| < \frac{1}{2^n} \tag{2.27}$$

entonces para $m \geq 1$

$$\begin{aligned}\sup_t |\mu_{n+m+1}\cdots\mu_{n+1}\mu_n t - \mu_{n+m}\cdots\mu_{n+1}\mu_n t| &= \sup_s |\mu_{n+m+1}s - s| \\ &\leq \frac{1}{2^{n+m}}\end{aligned}$$

para n fija la sucesión $\{\mu_{n+m}\cdots\mu_{n+1}\mu_n\}$ es fundamental para toda t, entonces se tiene la convergencia uniforme

$$\lambda_n t = \lim_{m\to\infty} \mu_{n+m}\cdots\mu_{n+1}\mu_n t.$$

De esta manera las funciones λ_n son continuas, no decrecientes, $\lambda_n 0 = 0$ y $\lambda_n 1 = 1$, falta probar que $\|\lambda_n\|$ es finito, para ello veamos que

$$\begin{aligned}\left|\log \frac{\mu_{n+m}\cdots\mu_n t - \mu_{n+m}\cdots\mu_n s}{t-s}\right| &\leq \|\mu_{n+m}\cdots\mu_{n+1}\mu_n\| \\ &\leq \|\mu_n\| + \|\mu_{n+1}\| + \cdots + \|\mu_{n+m}\| \\ &\leq \frac{1}{2^{n-1}}\end{aligned}$$

si $m \to \infty$ resulta $\|\lambda_n\| \leq 2^{1-n}$, es decir $\lambda_n \in \Lambda$. Por definición de λ_n se tiene $\lambda_n = \lambda_{n+1}\mu_n$, sustituyendo en (2.27) se obtiene

$$\sup_t |x_n(\lambda_n^{-1}t) - x_{n+1}(\lambda_{n+1}^{-1}t)| = \sup_s |x_n(s) - x_{n+1}(\mu_n s)| < \frac{1}{2^n}$$

Entonces las funciones $x_n(\lambda_n^{-1}t)$ forman una sucesión fundamental para toda t y convergen a una función $x(t)$, que pertenece a D. Como $\sup_t |x_n(\lambda_n^{-1}t) - x(t)| \to 0$ y $\|\lambda_n\| \to 0$ resulta $d_0(x_n, x) \to 0$, lo que prueba que toda sucesión fundamental en (D, d_0) es convergente. □

2.2.3 Los conjuntos compactos en D

Ahora demos paso a caracterizar los conjuntos compactos de D, en el caso de el espacio C el Teorema de Arzeà-Ascoli fue un factor importante, por ello enunciamos el sigueinte teorema análogo para el caso del espacio D.

Teorema 2.12 *Un subconjunto A de D es relativamente compacto en la topología de Skorohod si y sólo si*

$$\sup_{x\in A}\sup_t |x(t)| < \infty \quad y \quad \lim_{\delta\to 0}\sup_{x\in A} w'_x(\delta) = 0 \tag{2.28}$$

DEMOSTRACIÓN: Para demostrar que A es relativamente compacto, es suficiente con demostrar que es totalmente acotado con respecto a d_0, ya que (D, d_0) es completo, pero se probará primero respecto a d.

Dado $\varepsilon > 0$ sea k un entero posititivo tal que $1/k < \varepsilon$ y $w'_x(1/k) < \varepsilon$ para toda $x \in A$. Sea H una ε-red finita del intervalo $[-\alpha, \alpha]$ donde

$$\alpha = \sup_{x \in A} \sup_t |x(t)|$$

y B un subconjunto finito de funciones $y \in D$ tal que en los intervalos $[(u-1)/k, u/k)$ toman un valor constante de H, y $y(1) \in H$.

Probaremos que el conjunto B es una 2ε-red respecto a d; sea x en A, dado que $w'_x(1/k) < \varepsilon$ existe una partición $\{t_0, \dots, t_r\}$ del intervalo $[0,1]$ tal que

$$t_i - t_{i-1} > \frac{1}{k} \quad \text{y} \quad w'_x[t_{i-1}, t_i) < \varepsilon \tag{2.29}$$

sean u_i enteros distintos tal que $u_i/k \leq t_i < (u_i + 1)/k$ para $i = 1, \dots, r$, por tanto existe $\lambda \in \Lambda$ tal que $\lambda(u_i/k) = t_i$ y es lineal en los intervalos $[u_i/k, u_i + 1/k)$. Sea y en B que satisface

$$\left| y\left(\frac{u}{k}\right) - x\left(\lambda \frac{u}{k}\right) \right| < \varepsilon \qquad \text{con } 0 \leq u \leq k$$

como $[\lambda(u/k), \lambda((u+1)/k)) \subset [\lambda(u_i/k), \lambda(u_{i+1}/k)) = [t_i, t_{i+1})$ y por (2.29) la función $x(\lambda t)$ no puede variar más de ε en el intervalo $[u/k, (u+1)/k)$, y como y es constante en estos intervalos, tenemos $|y(t) - x(\lambda t)| < 2\varepsilon$ para toda t. Además $|t_i - \frac{u_i}{k}| < 1/k < \varepsilon$ por tanto $\sup_t |\lambda t - t| < \varepsilon$ y $d(x, y) < 2\varepsilon$, es decir, B un una 2ε-red respecto a d.

Dado η positiva, se elige δ tal que $0 < \delta < 1/4$, entonces $4\delta + w'_x(\delta) < \eta$ para toda x en A, y tomemos ε de manera que $0 < 2\varepsilon < \delta^2$. Por el Lema 2.3, B es una η-red respecto a d_0, con lo que aseguramos que A es totalmente acotado respecto a d_0, y su cerradora es compacta en (D, d_0).

Si A es relativamente compacto, es acotado por lo que resulta la primera parte de (2.28). La prueba de la otra ecuación es más elaborada, probaremos que $w'_x(\delta)$ es semicontinua por arriba en x para cada δ.

Dados x, δ y ε hay que exbhir η tal que $d(x, y) < \eta$ implique $w'_x(\delta) < w'_y(\delta) + \varepsilon$. De nuevo sea $\{t_0, \dots, t_r\}$ una partición del intervalo $[0,1]$ tal que $t_i - t_{i-1} > \delta$ y

$$w_x[t_{i-1}, t_i) < w'_x[\delta) + \frac{\varepsilon}{2} \qquad \text{con } i = 1, 2, \dots, r \tag{2.30}$$

tomemos η de manerea que $t_i - t_{i-1} > \delta + 2\eta$ para $0 \leq i \leq r$ y $\eta < \varepsilon/4$. Si $d(x, y) < \eta$ entonces para alguna λ en Λ se tiene

$$\sup_t |\lambda t - t| < \eta \quad \text{y} \quad \sup_t |x(t) - y(\lambda t)| < \eta$$

Llamemos $s_i = \lambda^{-1} t_i$, por consiguiente

$$s_i - s_{i-1} > t_i - t_{i-1} - 2\eta > \delta$$

Si s y t estan en $[s_{i-1}, s_i)$ entonces λs y λt estan en $[t_{i-1}, t_i)$ por tanto llegamos a $|y(s) - y(t)| < w'_x(\delta) + \varepsilon$, más aún tenemos la desigualdad

$$w_y[s_{i-1}, s_i) < w'_x(\delta) + \varepsilon$$

entonces $w'_y(\delta) < w'_x(\delta) + \varepsilon$. □

Continuemos con la caracterización de los conjuntos compctos en (D, d_0), pero antes definamos una variante de w'_x, a saber

$$w''_x(\delta) = \sup \min\{|x(t) - x(t_1)|, |x(t_2) - x(t)|\}$$

donde el supremo se toma sobre todos los puntos t, t_1 y t_2 tal que

$$t_1 \leq t \leq t_2 \quad \text{y} \quad t_2 - t_1 \leq \delta \tag{2.31}$$

Para ε y δ dados dividamos el intervalo $[0, 1)$ en subintervalos $[s_{i-1}, s_i)$ tal que $s_i - s_{i-1} > \delta$ y $w_x[s_{i-1}, s_i) < w'_x(\delta) + \varepsilon$. Si se tienen puntos como en (2.31), entonces en el caso en que los puntos t_1, t_2 esten en un mismo intervalo $[s_{i-1}, s_i)$ se tiene $|x(t) - x(t_1)| < w'_x(\delta) + \varepsilon$ y $|x(t_2) - x(t)| < w'_x(\delta) + \varepsilon$, si estan en intervalos vecinos $[s_{i-1}, s_i)$ y $[s_i, s_{i+1})$ se satisface $|x(t) - x(t_1)| < w'_x(\delta) + \varepsilon$ para $t_1 \leq t < s_i$ y $|x(t_2) - x(t)| < w'_x(\delta) + \varepsilon$ para $s_i \leq t \leq t_2$, obteniendo la relación

$$w''_x(\delta) \leq w'_x(\delta) \tag{2.32}$$

Teorema 2.13 *Un conjunto A es relativamente compacto en la topología de Skorohod si y sólo si*

$$\sup_{x \in A} \sup_t |x(t)| < \infty$$

y además satisface

$$\begin{aligned}
&\lim_{\delta \to 0} \sup_{x \in A} w''_x(\delta) = 0 \\
&\lim_{\delta \to 0} \sup_{x \in A} w_x[0, \delta) = 0 \\
&\lim_{\delta \to 0} \sup_{x \in A} w_x[1 - \delta, 1) = 0
\end{aligned} \tag{2.33}$$

DEMOSTRACIÓN: Por el Teorema 2.12 probaremos que las equaciones (2.33) son equivalentes a

$$\lim_{\delta\to 0} w'_x(\delta) = 0 \tag{2.34}$$

La implicación de (2.34) a (2.33) se sigue de la relación $w''_x(\delta) \le w'_x(\delta)$.

Dado ε positivo, se elige δ tal que para toda x en A

$$w''_x(\delta) < \varepsilon, \quad w_x[0,\delta) < \varepsilon \quad \text{y} \quad w_x[1-\delta, 1) < \varepsilon$$

Sean t_1, t_2 tale que $t_2 - t_1 \le \delta$, si $t_1 \le s \le t \le t_2$ y $|x(s) - x(t_1)| > \varepsilon$ entonces $|x(t) - x(s)| < \varepsilon$ y $|x(t_2) - x(s)| < \varepsilon$, por consiguiente $|x(t_2) - x(t)| < 2\varepsilon$ y

$$\min\{|x(s) - x(t_1)|, |x(t_2) - x(t)|\} < 2\varepsilon \tag{2.35}$$

Supongamos que x tiene un salto mayor que 2ε, en los puntos τ_1, τ_2 y además $0 < \tau_2 - \tau_1 < \delta$, para los puntos $t_1 \le s \le t \le t_2$ con $t_2 - t_1 \le \delta$; $\tau_1 = s$, $\tau_2 = t_2$ y t_1 cercano a τ_1 y t cercano a τ_2 se da la desigualdad (2.35) que es un contradicción. Resumiendo, la función x en el intervalo $[0, 1]$ no puede tener saltos mayores que 2ε en dos puntos cuya distancia sea menor que δ, lo mismo se puede decir para los intervalos $[0, \delta)$ y $[1 - \delta, 1)$.

Si x tiene un salto mayor que 2ε es un punto s_i, entonces estos puntos son tales que $0 = s_0 < s_1 < \cdots < s_m = 1$ y $s_i - s_{i-1} \ge \delta$. Si $s_i - s_{i-1} > \delta$ agregamos el punto medio del intervalo $[s_{i-1}, s_i]$ a la partición de esta manera se tiene una nueva partición $s_0, \ldots, s_r$ $(r \ge m)$ que satisface

$$\frac{1}{2}\delta < s_i - s_{i-1} \le \delta \qquad \text{para } i = 1, 2, \ldots r$$

Sean t_1, t_2 tal que $s_{i-1} \le t_1 < t_2 < s_i$; sea σ_1 el supremos de $\sigma \in [t_1, t_2]$ tal que $\sup_{t_1 \le u \le \sigma} |x(u) - x(t_1)| \le 2\varepsilon$, y σ_2 el ínfimo de $\sigma \in [t_1, t_2]$ tal que $\sup_{\sigma \le u \le t_2} |x(t_2) - x(u)| \le 2\varepsilon$. Si $\sigma_1 < \sigma_2$ existe s mayor que σ_1 tal que $|x(s) - x(t_1)| > 2\varepsilon$ y t menor que σ_2 tal que $|x(t_2) - x(t)| > 2\varepsilon$, como $t_2 - t_1 < \delta$ resulta una contradicción a (2.35); por lo tanto $\sigma_2 \le \sigma_1$ y $|x(\sigma_1-) - x(t_1)| \le 2\varepsilon$ y $|x(t_2) - x(\sigma_1)| \le 2\varepsilon$.

Como $t_1 < \sigma_1 \le t_2$, $\sigma_1 \in (s_{i-1}, s_i)$ y el salto en σ_1 es a lo más de 2ε, es decir $|x(\sigma_1) - x(\sigma_1-)| < 2\varepsilon$ por lo que

$$w_x[s_{i-1}, s_i) \le 6\varepsilon$$

de donde $w'_x(\delta/2) \le 6\varepsilon$, esto es $\lim_{\delta\to 0} w'_x(\delta) = 0$. □

2.2.4 Conjuntos de dimensión finita

Para los puntos $t_1, \ldots, t_k$ en $[0,1]$ se define la proyección natural $\pi_{t_1,\ldots,t_k}$ del espacio D a $\mathbb{R}^k$ como

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

Las funciones π_0 y π_1 son continuas, pero para t en $(0,1)$ que pasa con π_t. Consideremos los puntos $0 < t < 1$, si $\{x_n\}$ converge a x en la topología de Skorohod y además x es continua en t_0, se tiene $x_n(t_0) \to x(t_0)$; ahora supongamos que x no es continua en t_0, si $\lambda_n \in \Lambda$ es lineal en los intervalos $[0,t_0]$ y $[t_0,1]$, $\lambda_n t_0 = t_0 - 1/n$ y la sucesión $x_n(s) \equiv x(\lambda_n s)$, entonces $\{x_n\}$ converge a x en la topología de Skorohod pero $x_n(t_0)$ no converge a $x(t_0)$.

Lema 2.4 *Para $0 < t < 1$, π_t es continua en x si y sólo si x es una función continua en t.*

Falta ver que $\pi_{t_1,\ldots,t_k}$ es medible respecto a $\mathcal{B}(D)$ la σ-álgebra generada por los conjuntos abiertos de la topología de Skorohod. Sólo consideremos un punto t, si $\{x_n\}$ converge a x en la topología de Skorohod, entonces $x_n(s) \to x(s)$ para todo puntos de continuidad de x. Como las funciones x_n son uniformemente acotadas

$$\frac{1}{\varepsilon}\int_t^{t+\varepsilon} x_n(s)ds \to \frac{1}{\varepsilon}\int_t^{t+\varepsilon} x(s)ds$$

para cada ε positivo. Entonces $h_\varepsilon(x) = \varepsilon^{-1}\int_t^{t+\varepsilon} x(s)ds$ es continua respecto a la topología de Skorohod, y $h_\varepsilon(x) \to \pi_t(x)$ para cada x cuando $\varepsilon \to 0$ ya que la función x es continua por la derecha. Entonces π_t es medible y en consecuencia $\pi_{t_1,\ldots,t_k}$.

Los *conjuntos de dimensión finita* de D son los conjuntos de la forma $\pi^{-1}_{t_1,\ldots,t_k}H$ con $k \geq 1$ y H en $\mathcal{B}(\mathbb{R}^k)$, y además son conjuntos medibles.

Sea T_0 un subconjunto de $[0,1]$, y $\mathcal{F}_{T_0}$ la clase de los conjuntos $\pi^{-1}_{t_1,\ldots,t_k}H$, con t_i elementos de T_0 y $H \in \mathcal{B}(\mathbb{R}^k)$. De esta manera $\mathcal{F}_{T_0}$ es un álgebra aditiva finita, y $\mathcal{F}_{[0,1]}$ es la clase de todos los conjuntos de dimensión finita.

Teorema 2.14 *Sea T_0 un subconjunto de $[0,1]$ denso tal que $1 \in T_0$, entonces $\mathcal{F}_{T_0}$ genera a $\mathcal{B}(D)$.*

DEMOSTRACIÓN: Sea $\{t_n\}$ un subconjunto de T_0 que es denso en $[0,1]$ con $t_1 = 1$, y $S_{d_0}(x,r)$ una bola abierta respcto a d_0, probaremos que $S_{d_0}(x,\varepsilon)$ está en la σ-álgebra generada por $\mathcal{F}_{T_0}$. Para $0 < \varepsilon < r$ y $k \geq 1$, llamemos

$A_k(\varepsilon)$ al conjunto de las funciones y en D para la cual existe $\lambda \in \Lambda$ que cumple con

$$\|\lambda\| < r - \varepsilon \quad \text{y} \quad \max_{1\leq i\leq k} |y(t_i) - x(\lambda t_i)| < r - \varepsilon$$

Veamos que $A_k(\varepsilon)$ pertenece a $\mathscr{F}_{T_0}$. Para ε y k fijos, sea H_1 el conjunto de los puntos $(x(\lambda t_1), \ldots, x(\lambda t_k))$ donde λ recorre todas las funciones en Λ tal que $\|\lambda\| < r-\varepsilon$, H_2 el conjunto de puntos $(\alpha_1, \ldots, \alpha_k)$ tal que $|\alpha_i - \beta_i| < r-\varepsilon$ para algún $(\beta_1, \ldots, \beta_k)$ de H_1. H_2 es un conjunto abierto en $\mathbb{R}^k$ y $A_k(\varepsilon) = \pi^{-1}_{t_1,\ldots,t_k} H_2$, por tanto $A_k(\varepsilon) \in \mathscr{F}_{T_0}$.

Ahora porbaremos que

$$S_{d_0}(x,r) = \cup_\varepsilon \cap_{k=1}^{\infty} A_k(\varepsilon)$$

donde la unión es sobre todos los racionales ε del intervalo $(0,r)$, justificando que $S_{d_0}(x,r)$ es un elemento de $\mathscr{F}_{T_0}$. Veamos primero que

$$\cap_{k=1}^{\infty} A_k(\varepsilon) \subset S_{d_0}(x,r)$$

Sea y en $\cap_{k=1}^{\infty} A_k(\varepsilon)$, para cada k elegimos λ_k en Λ que satisface

$$\|\lambda_k\| < r - \varepsilon \quad \text{y} \quad \max_{1\leq i\leq k} |y(t_i) - x(\lambda_k t_i)| < r - \varepsilon \tag{2.36}$$

Por el Teorema de Helly, existe una subsecesión $\{\lambda_{k'}\}$ y una función λ no decreciente, tal que

$$\lim_{k'} \lambda_{k'} t = \lambda t$$

para todo punto t de continuidad de λ.

Sean s, t puntos de continuidad de λ, entonces

$$\left|\log \frac{\lambda t - \lambda s}{t-s}\right| = \lim_{k'} \left|\log \frac{\lambda_{k'} t - \lambda_{k'} s}{t-s}\right| \leq r - \varepsilon$$

por lo cual λ es continua y creciente, además $\|\lambda\| \leq r - \varepsilon$. Por (2.36), $|y(t_1) - y(t_1)| < r - \varepsilon$ ya que $t_1 = 1$, y si $i > 1$, $|y(t_i) - x(\lambda_{k'} t_i)| < r - \varepsilon$ para $k' \geq i$. Como $\lambda_{k'} t \to \lambda t$ se tiene $|y(t_i) - x(\lambda t_i)| < r - \varepsilon$ o $|y(t_i) - x((\lambda t_i)-)| < r - \varepsilon$, como $\{t_i\}$ es denso $\sup_t |y(t) - x(\lambda t)| < r - \varepsilon$, y $y \in S_{d_0}(x,r)$. La otra contención no presenta problema.

Con esto terminamos la demostración ya D es un espacio separable. □

2.2.5 Distribuciones de dimensión finita

Si P es una medida de probabilidad en $(D, \mathscr{B}(D))$, las *distribuciones de dimensión finta* de P son las medidas $P\pi^{-1}_{t_1,\ldots,t_k}$ dadas por $P\pi^{-1}_{t_1,\ldots,t_k}H = P(\pi^{-1}_{t_1,\ldots,t_k}H)$ en $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$.

Como (D, d_0) es completo, una familia de medidas de probabilidad es relativamente compacto si y sólo si es tensa, ya se estudió la caracterización de los conjuntos compactos en el espacio D que es fundamental pero hay un problema, las proyecciones no son necesariamente continuas. Esta dificultad se tratará de solucionar en esta sección.

Sea P una medida de probabilidad en $(D, \mathscr{B}(D))$, denotemos por T_P al conjunto de los puntos t en $[0,1]$ tal que π_t es continua excepto en un conjunto de P-medida cero $(0, 1 \in T_P)$. Sea $0 < t < 1$, t pertenece a T_P si y sólo si $P(J_t) = 0$ donde

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

el conjunto de las funciones x que son discontinuas en t.

Recordemos que los elementos de D tienen a lo más un número numerable de saltos, por lo que se probará que $P(J_t) > 0$ a lo más para un número numereble de puntos t. Sea $J_t(\varepsilon)$ el conjunto de las funciones x que tienen un salto mayor que ε en t. Para ε y δ fijos $P(J_t(\varepsilon)) \geq \delta$, para un número finito de puntos, de lo contrario si es cierta para los puntos $t_1, t_2, \ldots$ todos ellos distintos, entonces el conjunto $\limsup_n(J_{t_n}(\varepsilon))$ tendría probabilidad δ por lo menos, es decir, sería no vacio lo cual lleva a una contradicción, ya que los elementos de D tienen sólo un número finito de puntos con salto mayor que ε positivo. Para ε fija, $P(J_t(\varepsilon)) > 0$ para un número numerable de puntos t, entonces $P(J_t(\varepsilon)) \to P(J_t)$ cuando $\varepsilon \to 0$.

Resumiendo, el conjunto T_P contiene a los puntos 0, 1, y su complemento respecto a $[0,1]$ es a lo más numerable. Por lo tanto si $t_1, t_2, \ldots, t_k$ pertenecen a T_P, entonces $\pi_{t_1,t_2,\ldots,t_k}$ es continua, excepto en un conjunto de P-medida cero.

Supongamos que $P_n \Rightarrow P$, donde P_n, P son medidas de probabilidad en $(D, \mathscr{B}(D))$ entonces

$$P_n\pi^{-1}_{t_1,t_2,\ldots,t_k} \Rightarrow P\pi^{-1}_{t_1,t_2,\ldots,t_k}$$

si todos los puntos t_i pertenecen a T_P. Es necesario que todos los puntos sean miembros de T_P por ejemplo: Si P es la delta de Dirac en el punto $1\{[0,1/2)\}$, y P_n también es delta de Dirac pero en el punto $1\{[0, 2^{-1} + n^{-1})\}$, entonces $P_n \Rightarrow P$ pero $P_n\pi_{2^{-1}}$ no converge a $P\pi_{2^{-1}}$.

Teorema 2.15 *Sea $\{P_n\}$ una sucesión de medidas de probabilidad definidas en $(D, \mathscr{B}(D))$ tensa y $P_n \pi^{-1}_{t_1,t_2,\dots,t_k} \Rightarrow P\pi^{-1}_{t_1,t_2,\dots,t_k}$ para todos $t_1, t_2, \dots, t_k$ en T_P, entonces $P_n \Rightarrow P$.*

DEMOSTRACIÓN: Por la tensión de $\{P_n\}$ cada subsucesión $\{P_{n'}\}$ contiene un subsucesión $\{P_{n''}\}$ que converge débilmente a una medida Q. Si $t_1, t_2, \dots, t_k \in T_P \cap T_Q$ se tiene $P_{n''}\pi^{-1}_{t_1,t_2,\dots,t_k} \Rightarrow P\pi^{-1}_{t_1,t_2,\dots,t_k}$ y también $P_{n''}\pi^{-1}_{t_1,t_2,\dots,t_k} \Rightarrow Q\pi^{-1}_{t_1,t_2,\dots,t_k}$, por tanto

$$P\pi^{-1}_{t_1,t_2,\dots,t_k} = Q\pi^{-1}_{t_1,t_2,\dots,t_k} \quad t_1, t_2, \dots, t_k \in T_P \cap T_Q$$

Dado que T_P y T_Q contienen todos los puntos del intervalo $[0,1]$ excepto por un número a lo más numerable, lo mismo se puede establecar para $T_P \cap T_Q$, además contiene al 1 y es denso. Por el Teorema 2.14, $\mathscr{B}(D)$ es generada por los conjuntos de dimensión finita respecto a los elementos de $T_P \cap T_Q$. Por lo tanto $P = Q$ y $P_n \Rightarrow P$. □

2.2.6 Tensión

Para determinar tensión de una familia de medidas de probabilidad en C, fue util el Teorema de Arzelà-Ascoli, que caracteriza a los conjuntos compactos en C, para el caso del espacio D ya se tiene un resultado análogo al Teorema de Arzelà-Ascoli, por tanto demos inicio a determinar a las familias tensas en D.

Del Teorema 2.13 que caracteriza a los conjuntos compactos en D, se obtiene este resultado, que caracteriza a los conjuntos tensos en D.

Teorema 2.16 *Una sucesión $\{P_n\}$ de medidas de probabilidad definidas en $(D, \mathscr{B}(D))$ es tensa si y sólo si satisface:*

(i) Para cada η positiva, existe a tal que

$$P_n\{x : \sup_t |x(t)| > a\} \le \eta \qquad n \ge 1$$

(ii) Para cada ε y η positivos existe δ, con $0 < \delta < 1$, y n_0 entero positivo, tal que

$$P_n\{x : w'_x(\delta) > \varepsilon\} \le \eta \qquad n \ge n_0$$

Utilizando el Teorema 2.13 se tiene otra opción para el criterio de tensión.

Teorema 2.17 *Una una sucesión $\{P_n\}$ de medidas de probabilidad definidas en $(D, \mathscr{B}(D))$ es tensa si y sólo si satisface:*

(i) Para cada η positiva, existe a tal que

$$P_n\{x : \sup_t |x(t)| > a\} \leq \eta \qquad n \geq 1$$

(ii) Para cada ε y η positivos existe δ, con $0 < \delta < 1$, y n_0 entero positivo, tal que

$$P_n\{x : w''_x(\delta) > \varepsilon\} \leq \eta \tag{2.37}$$

$$P_n\{x : w_x[0, \delta) > \varepsilon\} \leq \eta \tag{2.38}$$

$$P_n\{x : w_x[1-\delta, 1) > \varepsilon\} \leq \eta \tag{2.39}$$

para $n \geq n_0$

Teorema 2.18 *Sea $\{P_n\}$ una sucesión de medidas de probabilidad definidas en $(D, \mathscr{B}(D))$ tal que*

$$P_n\pi^{-1}_{t_1,t_2,\ldots,t_k} \Rightarrow P\pi^{-1}_{t_1,t_2,\ldots,t_k} \tag{2.40}$$

para todos los puntos $t_1, t_2, \ldots, t_k$ en T_P. Supongamos además que $P(J_1) = 0$, y que para cada ε y η positivos existe δ con $0 < \delta < 1$ y un entero positivo n_0 tal que

$$P_n\{x : w''(\delta) \geq \varepsilon\} \leq \eta \qquad n \geq n_0 \tag{2.41}$$

Entonces $P_n \Rightarrow P$.

DEMOSTRACIÓN: Es suficiente con probar que $\{P_n\}$ es tensa y para ello se empleará el Teorema 2.17. Sea η positivo y $\varepsilon = 1$, entonces existe δ y n_0 que satisfacen (2.41). Dado que T_P es denso en $[0, 1]$, contiene puntos $t_1, \ldots, t_k$ tal que $0 = t_1 < \cdots < t_k = 1$ y $t_i - t_{i-1} < \delta$. Por hipótesis se tiene que $\{P_n\pi^{-1}_{t_1,t_2,\ldots,t_k}\}$ es tensa en $\mathbb{R}^k$ por lo cual

$$P_n\left\{x : \max_{1\leq i\leq k} |x(t_i)| > a_0\right\} \leq \eta \quad n \geq 1$$

para algún entero a_0. Si $|x(t_i)| \leq a_0$ para $i = 1, \ldots, k$ y $w''_x(\delta) < 1$, entonces $|x(t)| < a_0 + 1$ para toda t. Tomando $a = a_0 + 1$ y junto con (2.41) llegamos a $P_n\{\sup_t |x(t)| > a\} \leq 2\eta$ para $n \geq n_0$; y como se hizo en el Teorema 2.2 se puede tomar $n_0 = 1$ incrementando a.

Demos paso a probar (2.38). Como los elemetos de D son continuos por la derecha, para ε y η dados existe δ tal que

$$P\{x : |x(\delta) - x(0)| \geq \varepsilon\} \leq \eta \tag{2.42}$$

ahora tomemos a δ como la menor que satisface la desigualdad anterior, la desigualdad (2.41) para algún n_0 y además pertenezca a T_P. Por lo anterior $P_n\pi_{0,\delta}^{-1} \Rightarrow P\pi_{0,\delta}^{-1}$ y como consecuencia

$$P_n\{x : |x(\delta) - x(0)| \geq \varepsilon\} \leq \eta$$

para n suficientemete grande. Si $w''_x(\delta) < \varepsilon$ y $|x(\delta) - x(0)| < \varepsilon$ entonces $|x(s) - x(0)| < 2\varepsilon$ para toda s en $[0, \delta]$, de donde $w_x[0, \delta] < 4\varepsilon$, que junto con (2.41) se obtiene $P_n\{x : w_x[0, \delta) \geq 4\varepsilon\} \leq 2\eta$, probando la desigualdad (2.38).

En forma análoga se demuestra la desigualdad (2.39); en lugar de (2.42) se considera

$$P\{x : |x(1) - x(1 - \delta)| \geq \varepsilon\} \leq \eta \tag{2.43}$$

para δ pequeña, que no es cierta en general ya que los elementos de D no son necesariamente continuos por la izquierda, pero contamos con $P(J_1) = 0$, entonces x es contiua por la izquierda en $t = 1$, excepto por un conjunto de medida cero.

Por tanto $\{P_n\}$ es tensa y $P_n \Rightarrow P$. □

El siguiente resultado trata de nuevo el módulo de continuidad $w_x(\delta)$, en lugar de los módulos $w'_x(\delta)$ y $w''_x(\delta)$ que se han empleado en el espacio D.

Teorema 2.19 *Supongamos que para cada η positiva, existe a tal que*

$$P_n\{x : |x(0)| > a\} \leq \eta \quad n \geq 1 \tag{2.44}$$

y que para cada ε y η positivos, existe δ, con $0 < \delta < 1$ y n_0 entero positivo tal que

$$P_n\{x : w_x(\delta) > \varepsilon\} \leq \eta \quad n \geq n_0 \tag{2.45}$$

Entonces $\{P_n\}$ es tensa, y si además P es el límite débil de una subsucesión, entonces $P(C) = 1$.

DEMOSTRACIÓN: Probaremos tensión con ayuda del Teorema 2.16. Como $w'_x(\delta) \leq w_x(2\delta)$ para $\delta < 2^{-1}$, se satisface la condición (ii) del Teorema 2.16. Para algún entero positivo k se tiene

$$|x(t)| \leq |x(0)| + \sum_{i=1}^{k} |x(it/k) - x((i-1)t/k)|$$

y junto con (2.45) se tiene la condición (i), por tanto $\{P_n\}$ es tensa.

Si $w_y(\delta/2) \geq 2\varepsilon$, entonces y esta en el interior del conjunto $\{x : w_x(\delta) \geq \varepsilon\}$, como $P_{n'} \Rightarrow P$ se tiene

$$P\{y : w_y(\delta/2) \geq 2\varepsilon\} \leq \liminf_{n'} P_{n'}\{x : w_x(\delta) \geq \varepsilon\}$$

Dados ε y η, existe δ y n_0 que satisfacen (2.45), de donde $P\{y : w_y(\delta/2) \geq 2\varepsilon\} \leq \eta$. Para cada k existe δ_k tal que si $A_k = \{y : w_y(\delta_k) \geq 1/k\}$ entonces $P(A_k) < 1/k$. Tomando $A = \liminf_k A_k$, $P(A) = 0$, si x no es elemento de A, $\lim_{\delta \to 0} w_x(\delta) = 0$, por lo que $P(C) = 1$. □

2.2.7 Un criterio de convergencia

Sea X es una función de $(\Omega, \mathcal{B}, P)$ a D, es decir, $X(\omega)$ pertenece a D, y el valor que toma en t se denota por $X(t, \omega)$. Para cada t, $X(t)$ denotará la función $\pi_t X$ de Ω a $\mathbb{R}$, al igual que el espacio C, X es un *elemento aleatorio* en D si y sólo si $X(t)$ es una variable aleatoria para cada t. Una sucesión $\{X_n\}$ de elementos aleatorios en D es tensa si la sucesión de las distribuciones correspondientes es tensa.

Obtendremos resultados de convergencia de distribuciones que serán utiles para el Teorema de Donsker que se estudiara más adelante.

Sean X_n, X elementos aleatorios en D, si P es la distribución de X escribiremos T_X en lugar de T_P; recordemos que el conjunto T_X contiene los puntos 0 y 1.

Teorema 2.20 *Supongamos que*

$$(X_n(t_1), \ldots, X_n(t_k)) \Rightarrow (X(t_1), \ldots, X(t_k))$$

para cualesquiera $t_1, \ldots, t_k$ *en* T_X, $P\{X(1) \neq X(1-)\} = 0$ *y que*

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

para $t_1 \leq t \leq t_2$ *y* $n \geq 1$, *donde* $\gamma \geq 0$, $\alpha > 1/2$ *y* F *es una función continua en* $[0, 1]$, *no decreciente. Entonces* $X_n \to X$.

DEMOSTRACIÓN: Sean τ, δ, n y m fijos, con n y m enteros positivos, consideremos las variables aleatorias

$$\xi_i = X_n\left(\tau + \frac{i}{m}\delta\right) - X_n\left(\tau + \frac{i-1}{m}\delta\right), \quad i = 1, \ldots, m$$

Sea

$$M''_m = \max\min\left\{\left|X_n\left(\tau + \frac{j}{m}\delta\right) - X_n\left(\tau + \frac{i}{m}\delta\right)\right|,\right.$$
$$\left.\left|X_n\left(\tau + \frac{k}{m}\delta\right) - X_n\left(\tau + \frac{j}{m}\delta\right)\right|\right\}$$

donde el máximo se toma sobre $0 \leq i \leq j \leq k \leq m$, por el Teorema 2.8 junto con la hipótesis se llega a

$$P\{M''_m \geq \varepsilon\} \leq \frac{K}{\varepsilon^{2\gamma}}[F(\tau+\delta) - F(\tau)]^{2\alpha} \tag{2.46}$$

con $K = K''_{\gamma,\alpha}$ que depende unicamente de γ y α.

Sea

$$w''(X_n, [\tau, \tau+\delta]) = \sup\min\{|X_n(t) - X_n(t_1)|, |X_n(t_2) - X_n(t)|\}$$

donde el supremo es sobre los puntos t, t_1, t_2 tal que $\tau \leq t_1 \leq t \leq t_2 \leq \tau + \delta$. Como X_n es continua por la derecha, si m tiende a infinito de la ecuación (2.46) se tiene

$$P\{w''(X_n, [\tau, \tau+\delta]) > \varepsilon\} \leq \frac{K}{\varepsilon^{2\gamma}}[F(\tau+\delta) - F(\tau)]^{2\alpha} \tag{2.47}$$

Si $\delta = (2u)^{-1}$ y suponemos además que

$$w''(X_n, [2i\delta, (2i+1)\delta]) \leq \varepsilon \quad 0 \leq i \leq u-1 \tag{2.48}$$

y

$$w''(X_n, [(2i+1)\delta, (2i+2)\delta]) \leq \varepsilon \quad 0 \leq i \leq u-2 \tag{2.49}$$

Si $t_1 \leq t \leq t_2$ y $t_2 - t_1 \leq \delta$, entonces los puntos t_1 y t_2 pertenecen a un mismo intervalo de los que están involucrados en (2.48) o en (2.49), por tanto

$$\min\{|X_n(t) - X_n(t_1)|, |X_n(t_2) - X_n(t)|\} \leq \varepsilon$$

de donde se tiene que $w''(X_n, \delta) \leq \varepsilon$, y por la desigualdad (2.47) se tiene

$$P\{w''(X_n, \delta) \geq \varepsilon\} \leq \frac{K}{\varepsilon^{2\gamma}}(\Sigma' + \Sigma'')$$

donde Σ' y Σ'' son de la forma

$$\sum_{k=1}^{r}[F(t_k) - F(t_{k-1})]^{2\alpha}$$

con $0 \leq t_1 \leq \cdots \leq t_r \leq 1$ y $t_k - t_{k-1} \leq 2\delta$, entonces

$$P\{w''(X_n, \delta) \geq \varepsilon\} \leq \frac{2K}{\varepsilon^{2\gamma}}[F(1) - F(0)][w_F(2\delta)]^{2\alpha - 1}$$

Como $2\alpha > 1$ y F es continua, el lado derecho de esta desigualdad converge a cero. □

2.2.8 El Principio de Invarianza

El espacio D presenta ventajas para tratar el Principo de Invarianza, ya que las variables X_n no serán tan complejas como lo fueron en C. Dadas las variables aleatorias $\xi_1, \xi_2, \ldots$ en $(\Omega, \mathscr{B}, P)$, $S_n = \xi_1 + \cdots + \xi_n$, se define $X_n(\omega)$ una función en D como

$$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) \tag{2.50}$$

Dado que $X_n(t)$ es una variable aleatoria, X_n es una función aleatoria.

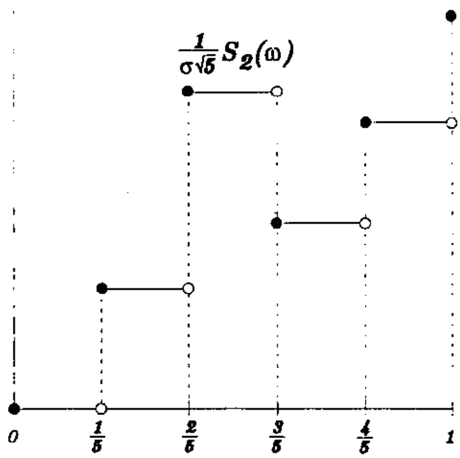

Figura 2.3: Gráfica de X_n, para $n = 5$

Deseamos probar que X_n converge débilmente a W, la medida de Wiener. Pero hay algo extraño, W es una medida en $(C, \mathscr{B}(C))$, y necesitamos que este definida en $(D, \mathscr{B}(D))$. La medida W se extenderá de la siguiente manera: como $C \in \mathscr{B}(D)$, y la topología de Skorohod restringida a C es la topología uniforme, para A en $\mathscr{B}(D)$, $A \cap C \in \mathscr{B}(C)$, así $W(A) = W(A \cap C)$. Cuando se hable de W, se entenderá como una media en $(D, \mathscr{B}(D))$ o como un elemento aleatorio en D cuya distribución es la medida de Wiener en D.

Teorema 2.21 *[Principio de Invarianza] Supongamos que las variables aleato rias ξ_i son independientes e idénticamente distribuídas con media cero y varianza finita σ^2, esto es*

$$E\{\xi_n\} = 0 \quad y \quad E\{\xi_n^2\} = \sigma^2 \tag{2.51}$$

Entonces las funciones aleatorias X_n definidas como (2.50) satisfacen

$$X_n \Rightarrow W$$

DEMOSTRACIÓN: La prueba de la convergencia de las distribuciones de dimensión finita es la misma que la elaborada para el espacio C. Sólo nos dedicaremos a probar que $\{X_n\}$ es tensa basándonos en el Teorema 2.20.

Como las variables ξ_i son independientes, entonces para $t_1 \leq t \leq t_2$

$$\begin{aligned} E\{|X_n(t) - X_n(t_1)|^2 \cdot |X_n(t_2) - X_n(t)|^2\} &= \\ &\frac{1}{\sigma^4 n^2} E\{|S_{[nt]} - S_{[nt_1]}|^2\} E\{|S_{[nt_2]} - S_{[nt]}|^2\} \\ &= \frac{1}{n^2}([nt] - [nt_1])([nt_2] - [nt]) \\ &\leq \left\{\frac{[nt_2] - [nt_1]}{n}\right\}^2 \end{aligned} \tag{2.52}$$

si $t_2 - t_1 \geq 1/n$, (2.52) se puede acotar por $4(t_2 - t_1)^2$, ahora si $t_2 - t_1 < 1/n$, t_1 y t están en un intervalo de la forma $[(i-1)/n, i/n)$ o bien t y t_2 lo están, esto es $X_n(t) - X(t_1) = 0$ ó $X_n(t_2) - X(t) = 0$, por tanto $X_n \Rightarrow W$. □

2.2.9 Topología Uniforme

Por un momento trabajaremos con el espacio D, pero ahora con la *topología uniforme*, es decir la que es generada por la métrica

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

El espacio (D, ρ) es completo, pero no es separable, por ejemplo tomemos los puntos x_θ con $0 < \theta < 1$

$$x_\theta(t) = \begin{cases} 0 & \text{si } 0 \leq t < \theta \\ 1 & \text{si } \theta \leq t \leq 1 \end{cases} \tag{2.53}$$

todo ellos distan uno y son un conjunto no numerable. El prefijo U se refiere a la topología uniforme, y el prefijo S a la topología de Skorohod del espacio

D, d representará cualquiera de las métricas dadas al espacio D y ρ la métrica uniforme, que satisfacen $d(x,y) \leq \rho(x,y)$, esto es, la topología uniforme es más fina que la topología de Skorohod , de donde U-convergencia implica S-convergencia, por otro lado si $\{x_n\}$ es S-convergente a x en C, entonces $\{x_n\}$ es U-convergente.

Sea $\mathscr{B}(U)$ la clase de los conjuntos U-Borel y $\mathscr{B}(D)$ la clase de los conjuntos S-Borel. Como un conjunto S-abierto es U-abierto se tiene

$$\mathscr{B}(D) \subset \mathscr{B}(U)$$

Sean P_n, P medidas de probabilidad definidas en $\mathscr{B}(U)$, denotaremos por $P_n \Rightarrow P[U]$ convergencia débil en el sentido de la topología uniforme y $P_n \Rightarrow P[S]$ en el sentido de la topología de Skorohod . Entonces si $P_n \Rightarrow P[U]$ implica que

$$\int f dP_n \to \int f dP \tag{2.54}$$

para toda función f U-continua y acotada; y $P_n \Rightarrow P[S]$ implica (2.54) para toda función f D-continua y acotada, por tanto $P_n \Rightarrow P[U]$ implica $P_n \Rightarrow P[S]$.

Supongamos que P_n, P estan definidas en $\mathscr{B}(U)$, $P_n \Rightarrow P[S]$ y $P(C) = 1$. Veremos que $P_n \Rightarrow P[U]$, para A en $\mathscr{B}(U)$ denotemos por $Cl_U(A)$ la U-cerradura de A, y por $Cl_S(A)$ la S-cerradura de A, $Cl_U(A) \subset Cl_S(A)$, como $P_n \Rightarrow P[S]$ y $P(C) = 1$, $\limsup_n P_n(A) \leq \limsup_n P_n(Cl_S(A)) \leq P(Cl_S(A)) = P(Cl_S(A) \cap C)$. Como una sucesión S-convergente a un punto en C, es U-convergente, $Cl_S(A) \cap C \subset Cl_U(A)$, de donde $\limsup_n P_n(A) \leq P(Cl_U(A))$, por tanto $P_n \Rightarrow P[U]$. Hemos probado el siguiente resultado.

Teorema 2.22 *Supongamos que P_n y P son medias de probabilidad definidas en $\mathscr{B}(U)$*

(i) Si $P_n \Rightarrow P[U]$ entonces $P_n \Rightarrow P[S]$.

(ii) Si $P_n \Rightarrow P[S]$ y $P(C) = 1$, entonces $P_n \Rightarrow P[U]$

Sean P_n las distribuciones de las variables aleatorias X_n involucradas en el Principio de Invarianza para el espacio D. Por el Teorema 2.21, $P_n \Rightarrow W[S]$, como $W(C) = 1$ se tiene $P_n \Rightarrow W[U]$.

3 Teoremas de Glivenko-Cantelli

Se tiene una moneda que tiene probabilidad p de que caiga "*sol*", si la moneda se lanza repetidas veces se espera que las frecuencias $n^{-1}\sum X_n$ se aproximen a p, donde X_i es igual a 1 si en el lanzamiento i se obtiene sol y 0 en otro caso.

La Teoría de Probabilidad ha tratado de dar una justificación mátematica a esto obteniendo entre otros resultados la *Ley fuerte de los grandes números*, que dice: Sean $X_1, X_2, \ldots$ variables aleatorias independientes e idénticamente distribuídas tal que $E|X_1| < \infty$, entonces

$$\frac{X_1 + \cdots + X_n}{n} \to EX_1 \text{ c.s.}$$

Uno de los factores que intervienen en esta ley es $n^{-1}\sum X_i$, un primer paso que se da para generalizar este resultado es el de distribuciones empíricas término que H. Cramér introduce en 1928. Consideremos las variables aleatorias $X_1, \ldots, X_n$, entonces la función

$$F_n(t,\omega) = \frac{1}{n}\sum_{i=1}^{n} 1_{[0,t]}(X_i(\omega))$$

es llamada la *distribución empírica* de las variables aleatorias X_i

La primera generalización de esta ley la fue hecha por V. Glivenko y F. P. Cantelli en trabajos independientes en el año de 1933 obteniendo así una nueva versión de la Ley de los Grandes Números para el caso de distribuciones empíricas de variables aleatorias independientes, este resultado es conocido como el Teorema de Glivenko-Cantelli.

Presentaremos el teorema clasico de Glivenko-Cantelli, y posteriormente dos generalizaciones, antes de esto se estudiarán las desigualdades máximales como la técnica de simetrización, que serán utilizadas en la prueba de las generalizaciones de los Teoremas de Glivenko-Cantelli, y de los Teorema de Donsker en el siguiente capítulo.

El contenido de este capítulo esta basado en el libro Van der Vart y Wellner [1996], la demostración. del Teorema de Glivenko-Cantelli es del libro Billingsley [1994].

3.1 El Teorema de Glivenko-Cantelli

Sean $\xi_1, \xi_2, \ldots$ variables aleatorias, definidas en mismo un espacio de probabilidad $(\Omega, \mathcal{B}, P)$ tal que $0 \leq \xi_i(\omega) \leq 1$. La función de *distribución empírica* $F_n(t, \omega)$ correspondiente a $\xi_1(\omega), \ldots, \xi_n(\omega)$, se define como:

$$F_n(t, \omega) = \frac{1}{n} \sum_{i=1}^{n} 1_{[0,t]}(\xi_i(\omega)) \tag{3.1}$$

Sea $\mathcal{F}$ una clase de funciones $f : \mathbb{R} \to \mathbb{R}$ medibles, y $X_1, X_2, \ldots$ variables aleatorias independientes e idénticamente distribuídas con $E|X_1| < \infty$, el camino a seguir para obtener una generalización de la Ley de los Grandes Números es determinar la convergencia

$$\frac{f(X_1) + \cdots + f(X_n)}{n} \to Ef(X_1) \quad \text{c.s.}$$

uniformemente sobre f. El Teorema de Glivenko-Cantelli prueba esta convergencia para la caso de las funciones indicadoras del intervalo $[0, 1]$.

Teorema 3.1 (Glivenko-Cantelli) *Supongamos que $\xi_1, \xi_2, \ldots$ son las variables aleatorias independientes con distribución común F, sea $D_n(\omega) = \sup_t |F_n(t, \omega) - F(t)|$. Entonces $D_n \to 0$ con probabilidad 1.*

DEMOSTRACIÓN: Para cada t, $F_n(t, \omega)$ es una variable aleatoria, por tanto D_n es una variable aleatoria.

Cada uno de los sumandos de (3.1) es independiente, y para t fija son identicamente distribuidas, por la Ley de los Grandes Números, existe un conjunto A_t con probabilidad 0, tal que

$$\lim_n F_n(t, \omega) = F(t) \qquad \omega \notin A_t$$

Ahora consideremos las funciones $1_{[0,t)}$, aplicando de nueva cuenta la Ley de los Grandes Números llegamos a $\lim_n F_n(t-, \omega) = F(t-)$ exepto en un conjunto B_t de probabilidad cero, donde $F(t-) = P(X < t)$. Sea $\varphi(u) = \inf\{t : u \leq F(t)\}$ con $0 < u < 1$ y $t_{m,k} = \varphi(k/m)$ con $m \geq 1$ y $1 \leq$

$k \leq m$. Entonces $F(\varphi(u)-) \leq u \leq F(\varphi(u))$ y con esto se tiene $F(t_{m,k}-) - F(t_{m,k-1}) \leq m^{-1}$, $F(t_{m,1}-) \leq m^{-1}$ y $F(t_{m,m}) \geq 1-m^{-1}$. Llamemos $D_{m,n}(\omega)$ el máximo de $|F_n(t_{m,k},\omega) - F(t_{m,k})|$ y $|F_n(t_{m,k}-,\omega) - F(t_{m,k}-)|$ para $k = 1,\dots,m$. Si $t_{m,k-1} \leq t < t_{m,k}$, entonces

$$\begin{aligned} F_n(t,\omega) &\leq F_n(t_{m,k}-,\omega) \\ &\leq F(t_{m,k}-) + D_{m,n}(\omega) \leq F(t) + m^{-1} + D_{m,n}(\omega) \end{aligned}$$

y

$$\begin{aligned} F_n(t,\omega) &\geq F_n(t_{m,k-1}-,\omega) \\ &\geq F(t_{m,k-1}) - D_{m,n}(\omega) \geq F(t) - m^{-1} - D_{m,n}(\omega) \end{aligned}$$

de manera similar se tratan los casos $t < t_{m,1}$ y $t \geq t_{m,m}$, para llegar a $D_n(\omega) \leq D_{m,n}(\omega) + m^{-1}$. Si ω no pertenece a $A = \cup(A_{t_{mk}} \cup B_{t_{mk}})$ entonces $\lim_n D_{m,n}(\omega) = 0$, por tanto $\lim_n D_n(\omega) = 0$, donde A tiene probabilidad 0. $\square$

Para dar paso generalizaciones del Teorema de Glivenko-Cantelli serán necesario nuevos conceptos que damos a continuación.

3.2 Distribuciones Empíricas

La *medida empírica* $\mathbb{P}_n$ de una colección de elementos aleatorios $X_1,\dots,X_n$ en $(\Omega, \mathscr{B})$ es la medida discreta dada por

$$\mathbb{P}_n(C) = \frac{1}{n}\#(1 \leq i \leq n : X_i \in C)$$

($\#A$ denota el número de elementos del conjunto A) con C en $\mathscr{B}$, que también puede escribirse como

$$\mathbb{P}_n = \sum_{i=1}^{n} \delta_{X_i}$$

Si P es una medida y f una función medible, al escribir Pf nos referimos a $\int fP$; sea $\mathscr{F}$ una colección de funciones medibles $f : \Omega \to \mathbb{R}$, entonces bajo la medida empírica es posible establecer una función de $\mathscr{F}$ a $\mathbb{R}$ dada por

$$f \to \mathbb{P}_n f$$

Dada una función f medible, la Ley de los Grandes Números nos garantiza la convergencia

$$\mathbb{P}_n f \to Pf \ \text{ c.s.}$$

si Pf existe y las variables aleatorias X_i son independientes e idénticamente distribuídas; en este capítulo se darán condiciones necesarias para garantizar esta convergencia uniformemente sobre una familia de funciones $\mathcal{F}$, que basicamente tratarán de ver que tan "*grande*" es la familia $\mathcal{F}$.

Denotemos por $\|Q\|_{\mathcal{F}} = \sup\{|Qf| : f \in \mathcal{F}\}$, entonces la versión uniforme de la Ley de los Grandes Números se puede escirbir como

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0$$

en probabilidad exterior o casi seguramente, una clase de funciones $\mathcal{F}$ que satisface la convergencia anterior se le llama *clase Glivenko-Cantelli.* Por ejemplo, la clase de las funciones indicadoras del intervalo $[0, 1]$ es una clase Glivenko-Cantelli.

Para poder determinar si una clase $\mathcal{F}$ dc funciones es Glivenko-Cantelli dependera del tamaño de $\mathcal{F}$, para lograr esto sera necesario introducir nuevos conceptos, que basicamente ilustrarán el número de bolas necesarias para cubrir a $\mathcal{F}$. De esto se desprende que $(\mathcal{F}, \|\cdot\|)$ sea un subconjunto de algún espacio normado de las funciones $f : \Omega \to \mathbb{R}$. Las normas que serán de mayor interés son las normas $L_p(P)$ donde P una medida de probabilidad y la norma esta dada por

$$\|f\|_{P,p} = \left(\int |f|^p dP\right)^{1/p}$$

Definición. *El número cubriente $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ es el mínimo número de bolas $\{g : \|g - f\| < \varepsilon\}$ necesarias para cubrir el conjunto $\mathcal{F}$. El centro de dichas bolas no tienen que ser necesariamente elementos de $\mathcal{F}$, sin embargo deben tener norma finita.*

Dadas dos funciones g y h, el *bracket* $[g, h]$ es el conjunto de todas las funciones f tal que $g \leq f \leq h$. Un ε-bracket es un bracket $[g, h]$ tal que $\|h - g\| < \varepsilon$.

Definición. *El número bracketing $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ es el mínimo número de ε-brackets necesarios para cubrir $\mathcal{F}$.*

Llamaremos números de entropía al logaritmo del núero cubriente o al logaritmo del número bracketing. Trataremos sólo con normas que satisfacen la propiedad de Riesz: es decir, para cualquier par de funciones f, g si $|f| \leq |g|$, entonces $\|f\| \leq \|g\|$. Para dichas normas contamos con la relación

$$N(\varepsilon, \mathcal{F}, \|\cdot\|) \leq N_{[]}(2\varepsilon, \mathcal{F}, \|\cdot\|)$$

ya que si f pertenece a un 2ε-bracket $[g, h]$, entonces esta en la bola de radio ε con centro en $(g-h)/2$.

Una *función envolvente* F de una clase $\mathcal{F}$ de funciones es tal que $|f(x)| \leq F(x)$ para toda x y f en $\mathcal{F}$, la mínima función envolvente es la función dada por $x \rightarrow \sup_f |f(x)|$. Siempre que tratemos de una función F envolvente supondremos que es finita.

3.3 Desigualdades Maximales

En esta sección trabajaremos sobre varias desigualdades que resultarán importantes para la prueba de los Teoremas de Glivenko-Cantelli de este capítulo, como los Teoremas de Donsker del capítulo siguiente.

Teorema 3.2 *Sea ψ una función convexa no decreciente tal que $\psi(0) = 0$ y X una variable aleatoria, entonces*

$$\|X\|_\psi = \inf\left\{C > 0 : E\psi\left(\frac{|X|}{C}\right) \leq 1\right\}$$

es una norma.

DEMOSTRACIÓN: Por definición $\|X\|_\psi \geq 0$; Sea $c \in \mathbb{R}$, entonces

$$\begin{aligned}
\|cX\|_\psi &= \inf\left\{C > 0 : E\psi\left(\frac{|c||X|}{C}\right) \leq 1\right\} \\
&= \inf\left\{|c|C' > 0 : E\psi\left(\frac{|X|}{C'}\right) \leq 1\right\} \\
&= |c|\|X\|_\psi
\end{aligned}$$

Sea $\varepsilon > 0$ y $r_1, r_2 > 0$ tal que

$$\|X\|_\psi \leq r_1 < \|X\|_\psi + \varepsilon \qquad \|Y\|_\psi \leq r_2 < \|Y\|_\psi + \varepsilon$$

si $C_0 = r_1 + r_2$, entonces

$$\begin{aligned} E\psi\left(\frac{|X+Y|}{C_0}\right) &\leq E\psi\left(\frac{|X|r_1}{C_0 r_1} + \frac{|Y|r_2}{C_0 r_2}\right) \\ &\leq E\frac{r_1}{C_0}\psi\left(\frac{|X|}{r_1}\right) + E\frac{r_2}{C_0}\psi\left(\frac{|Y|}{r_2}\right) \\ &\leq \frac{r_1}{C_0} + \frac{r_2}{C_0} = 1 \end{aligned}$$

de donde $\|X+Y\|_\psi \leq C_0 \leq \|X\|_\psi + \|Y\|_\psi + 2\epsilon$, por tanto $\|X+Y\|_\psi \leq \|X\|_\psi + \|Y\|_\psi$. □

A la norma del teorema anterior se le llama *norma de Orlicz.* Un caso particular y de interés en el presente trabajo es cuando $\psi(x) = x^p$ para $p \geq 1$, que corresponde a la norma L_p, en este caso se escribirá como

$$\|X\|_\psi = \|X\|_p = (E|X|^p)^{1/p}$$

Otro caso de interés es cuando la función esta dada por $\psi_p(x) = e^{x^p} - 1$ con $p \geq 1$. Como $x^p \leq \psi_p(x)$ para x no negativa se tiene $\|X\|_p \leq \|X\|_{\psi_p}$ para cada p, del hecho $x^p \leq (e^x - 1)p!$ resulta $\|X\|_p \leq p!\|X\|_{\psi_1}$ con $p \geq 1$. Para $p \leq q$ se da la desigualdad

$$\|X\|_{\psi_p} \leq \|X\|_{\psi_q}(\log 2)^{p/q}$$

La norma de Orlicz es útil para obtener una cota de las colas de la distribución, ya que por la desigualdad de Chebyshev

$$P(|X| > x) \leq P\left(\psi(|X|/\|X\|_\psi) \geq \psi(x/\|X\|_\psi)\right) \leq \frac{1}{\psi(x/\|X\|_\psi)}.$$

Lema 3.1 *Sea X un variable aleatoria tal que $P(|X| > x) \leq Ke^{-Cx^p}$ para toda x, con K, C constantes y $p \geq 1$. Entonces*

$$\|X\|_{\psi_p} \leq \left(\frac{1+K}{C}\right)^{1/p}$$

DEMOSTRACIÓN: Por el Teorema de Fubini

$$\begin{aligned} E(e^{D|X|^p} - 1) &= E\int_0^{|X|^p} De^{Ds}ds \\ &= \int_0^\infty P(|X| > s^{1/p})De^{Ds}ds \\ &\leq \int_0^\infty Ke^{-Cs}De^{Ds}ds = \frac{KD}{C-D} \end{aligned}$$

y es menor igual que 1 si y sólo si $D^{-1/p} \geq ((1+K)/C)^{1/p}$, es decir $\|X\|_{\psi_p} \leq ((1+K)/C)^{1/p}$ □

Analicemos el comportamiento la norma de Orlicz para un número finito o numerable de variables aleatorias, dado el hecho que $\max|X_i|^p \leq \sum |X_i|^p$ para el caso de la norma L_p se tiene

$$\left\| \max_{1\leq i\leq m} X_i \right\|_p = \left(E \max_{1\leq i\leq m} |X_i|^p \right)^{1/p} \leq m^{1/p} \max_{1\leq i\leq m} \|X_i\|_p$$

Una desigualdad similar se puede obtener para el caso de las normas exponenciales de Orlicz, veamos el caso general.

Lema 3.2 *Sea ψ una función convexa, no decreciente y no nula, tal que $\psi(0) = 0$ y $\limsup_{x,y\to\infty} \psi(x)\psi(y)/\psi(cxy) < \infty$ para alguna constante c. Entonces las variables aleatorias $X_1, \ldots, X_m$ satisfacen*

$$\left\| \max_{1\leq i\leq m} X_i \right\|_\psi \leq K\psi^{-1}(m) \max_{1\leq i\leq m} \|X_i\|_\psi$$

para una constante K que depende sólo de ψ.

Demostración: Para $x, y \geq 1$ supongamos que $\psi(x)\psi(y) \leq \psi(cxy)$, por tanto $\psi(x/y) \leq \psi(cx)/\psi(y)$ para $x \geq y \geq 1$; para $y \geq 1$ y cualquier constante C se tiene

$$\begin{aligned} \max_{1\leq i\leq m} \psi\left(\frac{|X_i|}{Cy}\right) &\leq \max_{1\leq i\leq m} \left[\frac{\psi(c|X_i|/C)}{\psi(y)} + \psi\left(\frac{|X_i|}{Cy}\right) 1\left\{\frac{|X_i|}{Cy} < 1\right\} \right] \\ &\leq \sum_{i=1}^{n} \frac{\psi(c|X_i|/C)}{\psi(y)} + \psi(1) \end{aligned}$$

ahora bien, tomando a $C = c\max\|X_i\|_\psi$ y aplicando esperanza a ambos lados de la desigualdad se tiene

$$E\psi\left(\frac{\max|X_i|}{Cy}\right) \leq \frac{m}{\psi(y)} + \psi(1)$$

supongamos que $\psi(1) \leq 1/2$, entonces el lado derecho de la desigualdad anterior es menor igual que 1 para $y = \psi^{-1}(2m)$, y como

$$\left\| \max_{1\leq i\leq m} |X_i| \right\|_\psi \leq E\psi\left(\frac{\max|X_i|}{Cy}\right)$$

se tiene que

$$\left\| \max_{1 \le i \le m} |X_i| \right\|_\psi \le Cy \le \psi^{-1}(2m)c \max_{1 \le i \le m} \|X_i\|_\psi$$

Como ψ es una función convexa y $\psi(0) = 0$, $\psi^{-1}(2m) \le 2\psi^{-1}(m)$. Todo esto bajo el supuesto de $\psi(1) \le 1/2$.

Para el caso general de ψ, existen constantes $\sigma \le 1$ y $\tau > 0$ tal que la función $\phi(x) = \sigma\psi(\tau x)$ es convexa no decreciente y sobre todo $\phi(1) \le 1/2$, aplicando la desigualdad a ϕ junto con $\sigma\tau\|X\|_\psi \le \|X\|_\phi$ y $\|X\|_\phi \le \|X\|_\psi \tau$ se obtiene la desigualdad para ψ. □

El valor de la constante K se puede tomar como 1 para el caso de las normas L_p, pero lo importante es que en la cota aparece ψ^{-1}, la cual no da una cota pequeña cuando ψ crece demasiado. Este lema no es útil si se tiene un conjunto de variables aleatorias con cardinalidad numerable o mayor, para sobrepasar este problema se presenta otro teorema.

Definición. *Sea (T, d) un espacio métrico, el número cubriente $N(\varepsilon, d)$ es el mínimo número de bolas de radio ε necesarias para cubrir T.*

Una colección de puntos están ε-separados si la distancia entre cualquier par de puntos es mayor que ε.

Definición. *Sea (T, d) un espacio métrico, el número packing $D(\varepsilon, d)$ es el máximo número de puntos ε-separados en T.*

Los números de entropía son el logaritmo del número cubriente y del numero packing respectivamente, la relación que guardan los números cubriente y los números packing es

$$N(\varepsilon, d) \le D(\varepsilon, d) \le N(\varepsilon/2, d)$$

Además T es totalmente acotado si y sólo si para todo $\varepsilon > 0$ los números cubrientes y los números packing son finitos.

Se dice que un proceso $\{X_t : t \in T\}$ estocástico es separable, si el número $\sup_{d(s,t)<\delta} |X_s - X_t|$ con s, t en T, no cambia al momento de intercambiar T por un conjunto numerable.

Teorema 3.3 *Sea ψ una función convexa, no decreciente, no nula, tal que $\psi(0) = 0$ y $\limsup_{x,y\to\infty} \psi(x)\psi(y)/\psi(cxy) < \infty$ para alguna constante c. Sea $\{X_t : t \in T\}$ un proceso estocástico separable tal que*

$$\|X_s - X_t\|_\psi \le Cd(s, t) \qquad \textit{para todo } s, t$$

donde d es una semimétrica en T y C un constante. Entonces para $\eta, \delta > 0$

$$\left\| \sup_{d(s,t)\leq\delta} |X_s - X_t| \right\|_\psi \leq K \left[\int_0^\eta \psi^{-1}(D(\varepsilon, d))d\varepsilon + \delta\psi^{-1}(D^2(\eta, d)) \right]$$

para una constante K que depende sólo de ψ y C.

DEMOSTRACIÓN: Supongamos que el número packing es finito. Sean $T_0 \subset T_1 \subset T_2 \cdots \subset T$ conjuntos tal que cada T_i es un conjunto con el máximo de puntos que satisfacen $d(s,t) > \eta 2^{-i}$ para todo $s, t \in T_i$; el número de elementos de T_i es menor igual que $D(\eta 2^{-i}, d)$ por definición de número packing.

Para cada punto t_{j+1} en T_{j+1} le asociamos un único punto t_j en T_j de manera que $d(t_j, t_{j+1}) \leq \eta 2^{-j}$. De esta manera a cada punto t_{k+1} se le asocia un punto t_0 en T_0 formando una cadena de puntos $t_{k+1}, t_k, \ldots, t_0$. Para los puntos s_{k+1}, t_{k+1} en T_{k+1} se tiene

$$\begin{aligned} |(X_{s_{k+1}} - X_{s_0}) - (X_{t_{k+1}} - X_{t_0})| &= \left| \sum_{j=0}^{k} (X_{s_{j+1}} - X_{s_j}) - \sum_{j=0}^{k} (X_{t_{j+1}} - X_{s_j}) \right| \\ &\leq \sum_{j=0}^{k} |X_{s_{j+1}} - X_{s_j}| + \sum_{j=0}^{k} |X_{t_{j+1}} - X_{t_j}| \\ &\leq 2\sum_{j=0}^{k} \max |X_u - X_u| \end{aligned}$$

donde el máximo es sobre todas las parejas (u, v) en $T_{j+1} \times T_j$ con j fijo, el número de parejas es a lo más el numero de elementos de T_{j+1} y $\|X_u - X_v\|_\psi \leq Cd(u,v) \leq C\eta 2^{-j}$. Por el Lema 3.2 se tiene

$$\begin{aligned} \left\| \max_{s,t\in T_{k+1}} |(X_s - X_{s_0}) - (X_t - X_{t_0})| \right\|_\psi &\leq 2\sum_{j=0}^{k} K_j \psi^{-1}(m_j) \max \|X_u - X_v\|_\psi \\ &\leq K' \sum_{j=0}^{k} \psi^{-1}(D(\eta 2^{-j-1}, d))\eta 2^{-j} \\ &\leq 4K \int_0^\eta \psi^{-1}(D(\varepsilon, d))d\varepsilon \qquad (3.2) \end{aligned}$$

donde s_0 y t_0 son los puntos finales de las cadenas que empiezan en los puntos s y t respectivamente y la consntante K depende de ψ y C.

Para cada par de puntos s_0 y t_0 en T_0 con $d(s_0, t_0) < \delta$ se elige un par de puntos s_{k+1} y t_{k+1} en T_{k+1} cuyos puntos finales de la cadena sean s_0 y t_0 y además $d(s_{k+1}, t_{k+1}) < \delta$, dando un total de parejas menor o igual que $D^2(\eta, d)$. Ahora de la desigualdad

$$|X_{s_0} - X_{t_0}| \leq |(X_{s_0} - X_{s_{k+1}}) - (X_{t_0} - X_{t_{k+1}})| + |X_{s_{k+1}} - X_{t_{k+1}}|$$

tomamos el máximo sobre todas las parejas en T_{k+1} que disten menos que δ y junto con (3.2) se llega a

$$\left\| \max_{d(s,t)<\delta} |X_s - X_t| \right\|_\psi \leq 8K \int_0^\eta \psi^{-1}(D(\epsilon, d))d\epsilon + \| \max |X_{s_{k+1}} - X_{t_{k+1}}| \|_\psi$$

con $s, t \in T_{k+1}$ cuyos puntos finales son s_0 y t_0. El último sumando esta acotado por $K\psi^{-1}(D^2(\eta, d)\delta$ según el Lema 3.2.

Como $\{X_t\}$ es un proceso separable y haciendo tender k a infinito, resulta la desigualdad buscada. □

3.3.1 Desigualdades Sub-Gaussianas

Seguiremos obteniendo resultados para las colas de variables aleatorias, el siguiente lema determinará cuando un proceso estocástico es sub-Gaussiano, como veremos más adelante.

Lema 3.3 (Desigualdad de Hoeffding) *Sean $a_1, \ldots, a_n$ constantes arbitrarias y $\xi_1, \ldots, \xi_n$ variables aleatorias tal que $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$. Entonces*

$$P\left(\left| \sum_{i=1}^n \xi_i a_i \right| > x \right) \leq 2\exp\left\{ -\frac{1}{2} \frac{x^2}{\|a\|^2} \right\}$$

donde $\|a\|$ representa la norma Euclideana del vector $(a_1, \ldots, a_n)$. En particular $\| \sum \xi_i a_i \|_{\psi_2} \leq \sqrt{6}\|a\|$.

Demostración: Para una constante $\lambda > 0$ se tiene $Ee^{\lambda\xi} = (e^\lambda + e^{-\lambda})/2 \leq e^{\lambda^2/2}$, por la desigualdad de Chebyshev se llega a

$$P\left(\sum_{i=1}^n a_i \xi_i > x \right) \leq e^{-\lambda x} E \exp\left\{ \lambda \sum_{i=1}^n a_i \xi_i \right\} \leq \exp\left\{ \left(\frac{\lambda^2}{2} \right) \|a\|^2 - \lambda x \right\}$$

la expresión de la derecha alcanza su mínimo para $\lambda = x/\|a\|^2$, resultado la cota de $\exp\{-\frac{1}{2}x^2/\|a\|^2\}$, de manera análoga se puede acotar $P(\sum a_i \xi_i < -x)$, para obtener la desigualdad buscada.

Por el Lema 3.1 tomando $K = 2$ y $C = (2\|a\|^2)^{-1}$ se obtiene $\| \sum \xi_i a_i \|_{\psi_2} \leq \sqrt{6}\|a\|$. □

Un proceso estocástico es *sub-Gaussiano* respecto la métrica d si:

$$P(|X_s - X_t| > x) \leq 2 \exp\left\{-\frac{x^2}{2d^2(s,t)}\right\}$$

para todo s, t en T y $x > 0$.

Todo proceso Gaussiano es un proceso sub-Gaussiano, considerando la semimétrica $d(s,t) = \sigma(X_s - X_t)$, un ejemplo de un proceso sub-Gaussiano es

$$X_a = \sum_{i=1}^{n} a_i \xi_i \qquad a \in \mathbb{R}^n$$

donde $P\{\xi_i = 1\} = P\{\xi_i = -1\} = 1/2$ con la métrica $d(a,b) = \|a - b\|$, según el lema anterior.

Corolario 3.1 *Sea $\{X_t : t \in T\}$ un proceso sub-Gaussiano separable. Entonces para cada $\delta > 0$,*

$$E \sup_{d(s,t) \leq \delta} |X_s - X_t| \leq K \int_0^{\delta} \sqrt{\log D(\varepsilon, d)}\, d\varepsilon$$

para una constante universal K. En particular para cualquier t_0

$$E \sup_t |X_t| \leq E|X_{t_0}| + K \int_0^{\infty} \sqrt{\log D(\varepsilon, d)}\, d\varepsilon$$

DEMOSTRACIÓN: Aplicamos el Teorema 3.3 con ψ_2 y $\eta = \delta$ entonces

$$\left\| \sup_{d(s,t) \leq \delta} |X_s - X_t| \right\|_{\psi_2} \leq K \left[\int_0^{\delta} \psi_2^{-1}(D(\varepsilon, d))\, d\varepsilon + \delta \psi_2^{-1}(D^2(\delta, d)) \right]$$

como $\psi_2^{-1}(m) = \sqrt{\log(1+m)}$, $\psi_2^{-1}(D^2(\delta,d)) \leq \sqrt{2}\psi_2^{-1}(D(\delta,d))$, por lo que resulta

$$\begin{aligned} K \left[\int_0^{\delta} \psi_2^{-1}(D(\varepsilon, d))\, d\varepsilon + \delta \psi_2^{-1}(D^2(\delta, d)) \right] &\leq K \left[\int_0^{\delta} \sqrt{\log(1 + D(\varepsilon, d))}\, d\varepsilon + \right. \\ &\qquad \left. \delta \sqrt{2 \log(1 + D(\delta, d))} \right] \\ &\leq K' \int_0^{\delta} \sqrt{\log(1 + D(\varepsilon, d))}\, d\varepsilon \end{aligned}$$

Como $\log(1+m) \leq 2\log m$ para $m \geq 2$ y $D(\varepsilon, d) \geq 2$ para todo ε menor que el diámetro de T resulta la desigualdad buscada. El caso particular es inmediato. □

Lema 3.4 *Sean $X_1, \ldots, X_n$ variables aleatorias que satisfacen*

$$P(|X_i| > x) \leq 2\exp\left\{-\frac{1}{2}\frac{x^2}{b+ax}\right\} \tag{3.3}$$

para toda x positiva y constantes positivas a, b. Entonces

$$\left\|\max_{1\leq i\leq m} X_i\right\|_{\psi_1} \leq K\left(a\log(1+m) + \sqrt{b}\sqrt{\log(1+m)}\right)$$

para una cosntante universal K.

Demostración: De (3.3) se obtiene

$$P(|X_i| > x) \leq \begin{cases} 2\exp\left\{-x^2/(4b)\right\} & \text{para } x \leq b/a \\ 2\exp\left\{-x/(4a)\right\} & \text{para } x > b/a \end{cases}$$

entonces para toda x se tiene $P(|X_i|1\{|X_i| \leq b/a\}) \leq 2\exp\{-x^2/4b\}$ y $P(|X_i|1\{|X_i| > b/a\}) \leq 2\exp\{-x/4a\}$. Por el Lema 3.1 se tiene

$$\||X_i|1\{|X_i| \leq b/a\}\|_{\psi_2} \leq \sqrt{12b} \quad \text{y} \quad \||X_i|1\{|X_i| > b/a\}\|_{\psi_1} \leq 12a$$

Por tanto desarrollando llegamos a

$$\begin{aligned}\left\|\max_i |X_i|\right\|_{\psi_1} &\leq \left\|\max_i |X_i|1\{|X_i| \leq b/a\}\right\|_{\psi_2} + \\ &\quad \left\|\max_i |X_i|1\{|X_i| > b/a\}\right\|_{\psi_1} \\ &\leq K\left(\sqrt{\log(1+m)}\sqrt{b} + \log(1+m)a\right)\end{aligned}$$

según el Lema 3.2. □

3.4 Simetrización

Sean $\varepsilon_i, \ldots, \varepsilon_n$ variables aleatorias independientes e idénticamente distribuídas tales que $P\{\varepsilon_i = 1\} = P\{\varepsilon_i = -1\} = 1/2$, en esta sección se considerará el proceso simétrico

$$f \to \mathbb{P}_n^\circ f = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i f(X_i)$$

con $\varepsilon_i, \ldots, \varepsilon_n$ independientes de las variables aleatorias $X_1, \ldots, X_n$.

Resulta que el teorema del límite central o la ley de los grandes númeroes es cierta para el proceso empírico si es cierta para el proceso simétrico, quien sera de ayuda para seguir este procedimiento es la técnica de simetrización que presentamos en esta sección.

En la demostración de este lema como el de los siguientes, haremos abuso de la notación, ya que al escribir

$$\left\|\sum f(X_i)\right\|_{\mathscr{F}}$$

es en realidad

$$\sup_{f\in\mathscr{F}}\left|\sum f(X_i)\right|$$

Lema 3.5 (Simetrización) *Sea $\Phi : \mathbb{R} \to \mathbb{R}$ una función convexa no decreciente y $\mathscr{F}$ una clase de funciones medibles, entonces*

$$E^*\Phi\left(\|\mathbb{P}_n - P\|_{\mathscr{F}}\right) \leq E^*\Phi\left(2\|\mathbb{P}_n^{\circ}\|_{\mathscr{F}}\right) \tag{3.4}$$

DEMOSTRACIÓN: Sean $Y_1, \ldots, Y_n$ copias independientes de $X_1, \ldots, X_n$ definidas como las últimas n proyecciones del espacio $(\Omega^n, \mathscr{B}^n, P^n) \times (\Omega^n, \mathscr{B}^n, P^n)$. Para $X_1, \ldots, X_n$ fijas se tiene

$$\begin{aligned}
\|\mathbb{P}_n - P\|_{\mathscr{F}} &= \sup_{f\in\mathscr{F}}\frac{1}{n}\left|\sum_{i=1}^{n}(f(X_i) - Ef(Y_i))\right| \\
&\leq E_Y^*\sup_{f\in\mathscr{F}}\frac{1}{n}\left|\sum_{i=1}^{n}(f(X_i) - f(Y_i))\right| \\
&= E_Y^*\left\|\frac{1}{n}\sum_{i=1}^{n}(f(X_i) - f(Y_i))\right\|_{\mathscr{F}} \\
&= E_Y\left\|\frac{1}{n}\sum_{i=1}^{n}(f(X_i) - f(Y_i))\right\|_{\mathscr{F}}^{*}
\end{aligned}$$

donde E_Y^* es la esperanza exterior respecto a $Y_1, \ldots, Y_n$, por la desigualdad de Jensen

$$\Phi(\|\mathbb{P}_n - P\|_{\mathscr{F}}) \leq E_Y\Phi\left(\left\|\frac{1}{n}\sum_{i=1}^{n}(f(X_i) - f(Y_i))\right\|_{\mathscr{F}}^{*}\right)$$

dado que Φ es no decreciente y continua, el $(*)$ se puede colocar en E_Y^*. Si se toma la esperanza respecto a $X_1, \ldots, X_n$ resulta

$$E^*\Phi(\|\mathbb{P}_n - P\|_{\mathscr{F}}) \leq E_X^* E_Y^* \Phi\left(\left\|\frac{1}{n}\sum_{i=1}^{n}(f(X_i) - f(Y_i))\right\|_{\mathscr{F}}\right)$$

por el Lema A.7, el lado derecho de la desigualdad se puede acotar por E_{XY}^*. Como Y_i es una copia de X_i, $f(X_i) - f(Y_i)$ tiene la misma distribución que $f(Y_i) - f(X_i)$, por lo que

$$E^*\Phi\left(\left\|\frac{1}{n}\sum_{i=1}^{n} e_i(f(X_i) - f(Y_i))\right\|_{\mathscr{F}}\right)$$

es igual para cualquier $(e_1, \ldots, e_n) \in \{-1, 1\}^n$, por la desigualdad del triángulo y como Φ es una función convexa se tiene

$$\begin{aligned}
E^*\Phi\left(\|\mathbb{P}_n - P\|\right) &\leq E_\varepsilon E_{XY}^* \Phi\left(\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(f(X_i) - f(Y_i))\right\|_{\mathscr{F}}\right) \\
&\leq \frac{1}{2}E_\varepsilon E_{XY}^*\Phi\left(2\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathscr{F}}\right) + \\
&\quad \frac{1}{2}E_\varepsilon E_{XY}^*\Phi\left(2\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f(Y_i)\right\|_{\mathscr{F}}\right) \\
&\leq \frac{1}{2}E_\varepsilon E_X^*\Phi\left(2\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathscr{F}}\right) + \\
&\quad \frac{1}{2}E_\varepsilon E_Y^*\Phi\left(2\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f(Y_i)\right\|_{\mathscr{F}}\right) \\
&\leq E_\varepsilon E_X^*\Phi\left(2\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i f(X_i)\right\|_{\mathscr{F}}\right)
\end{aligned}$$

Por tanto $E^*\Phi(\|\mathbb{P}_n - P\|_{\mathscr{F}}) \leq E^*\Phi(2\|\mathbb{P}_n^o\|_{\mathscr{F}})$ □

El lema de simetrización se aplicará a cualquier clase $\mathscr{F}$, y se empleará en las demostraciones de los Teorema de Glivenko-Cantelli y en en los Teoremas de Donsker, con Φ la función identidad.

Como el Teorema de Fubini falla para la esperanza exterior, pediremos que el integrando del lado derecho de (3.4) sea una función medible, esto es

$$(X_1, \dots, X_n) \mapsto \left\| \sum_{i=1}^{n} e_i f(X_i) \right\|_{\mathscr{F}} \tag{3.5}$$

es medible para cualquier $(e_1, \dots, e_n) \in \{-1, 1\}$ y así utilizar el Lema de Simetrización y el Teorema de Fubini sin ningún inconveniente.

Definición. *Una clase $\mathscr{F}$ de funciones $f : \Omega \to \mathbb{R}$ medibles en el espacio $(\Omega, \mathscr{B}, P)$ es una clase P-medible, si la función (3.5) es mebible en $(\Omega^n, \mathscr{B}^n, P^n)$ para cada n y para todo vector $(e_1, \dots, e_n) \in \mathbb{R}^n$.*

3.5 Dos Teoremas de Glivenko-Cantelli

Demos paso a probar condiciones necesarias para que una familia de funciones sea una clase Glivenko-Cantelli.

Teorema 3.4 *Sea $\mathscr{F}$ una clase de funciones medibles tal que para cada $\varepsilon > 0$, $N_{[]}(\varepsilon, \mathscr{F}, L_1(P)) < \infty$. Entonces $\mathscr{F}$ es Glivenko-Cantelli.*

DEMOSTRACIÓN: Sea $\varepsilon > 0$, y $[g_i, h_i]$ un número finito de ε-brackets de manera que la unión de ellos contenga a $\mathscr{F}$ y $P(h_i - g_i) < \varepsilon$ para toda i. Dado $f \in \mathscr{F}$, existe un ε-bracket tal que $g_i \leq f \leq h_i$ y

$$\begin{aligned} (\mathbb{P}_n - P)f &\leq (\mathbb{P}_n - P)h_i + P(h_i - f) \\ &\leq (\mathbb{P}_n - P)h_i + \varepsilon \end{aligned}$$

de donde

$$\sup_{f \in \mathscr{F}} (\mathbb{P}_n - P)f \leq \max_i (\mathbb{P}_n - P)h_i + \varepsilon$$

por la Ley de los Grandes Números el lado derecho converge a ε. De manera análoga se obtiene

$$\inf_{f \in \mathscr{F}} (\mathbb{P}_n - P)f \geq \min_i (\mathbb{P}_n - P)g_i + \varepsilon$$

Por lo cual $\limsup \|\mathbb{P}_n - P\|^*_{\mathscr{F}} \leq \varepsilon$ para todo $\varepsilon > 0$. Por tanto $\mathscr{F}$ es Glivenko-Cantelli. □

El último teorema de Glivenko-Cantelli es un poco más complicado.

Teorema 3.5 *Sea $\mathscr{F}$ una clase P-medible de de funciones medibles con envolvente F tal que $P^*F < \infty$. Sea $\mathscr{F}_M$ la clase de funciones $f1\{F \leq M\}$ con f en $\mathscr{F}$. Si $\log N(\varepsilon, \mathscr{F}_M, L_1(\mathbb{P}_n)) = o_P^*(n)$ para cada ε y M positivos, entonces $\|\mathbb{P}_n - P\|_{\mathscr{F}}^* \to 0$ c.s. y en media. En particular $\mathscr{F}$ es Glivenko-Cantelli.*

DEMOSTRACIÓN: Por el Lema de Simetrización tomando $\Phi(x) = x$ y el Teorema de Fubini

$$
\begin{aligned}
E^*\left(\|\mathbb{P}_n - P\|_{\mathscr{F}}\right) &\leq E^*(2\|\mathbb{P}_n^\circ\|) \\
&= 2E\left(\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathscr{F}}\right) \\
&= 2E_X E_\varepsilon\left(\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathscr{F}}\right) \\
&= 2E_X E_\varepsilon\left(\sup_{f\in\mathscr{F}} \frac{1}{n}\left|\sum_{i=1}^n (f1\{F \leq M\} + f1\{F > m\})\right|\right) \\
&\leq 2E_X E_\varepsilon\left(\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathscr{F}_M} + \left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathscr{F}}\right) \\
&\leq 2E_X E_\varepsilon\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathscr{F}_M} + 2P^*F1\{F > M\}
\end{aligned}
$$

para toda $M > 0$. Para M suficientemente grande, $P^*F1\{F > M\}$ es pequeño, ya que $P^*F < \infty$, pondremos atención en acotar el primer sumando considerando M fija. Para $X_1, \ldots X_n$ fijas, sea $\mathscr{G}$ una ε-red de $\mathscr{F}_M$ bajo la norma $L_1(\mathbb{P}_n)$, el tamaño del conjunto $\mathscr{G}$ puede ser elegido de manera que sea igual a $N(\varepsilon, \mathscr{F}_M, L_1(\mathbb{P}_n))$. Dado $f \in \mathscr{F}_M$ existe g en $\mathscr{G}$ tal que $\|f - g\|_1 < \varepsilon$, es decir,

$$\mathbb{P}_n|f - g| = \frac{1}{n}\sum_{i=1}^n |f - g|(X_i) < \varepsilon$$

de donde

$$E_\varepsilon\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathscr{F}_M} \leq E_\varepsilon\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathscr{G}} + \varepsilon$$

Ahora bien, desarrollemos el primer sumando de la desigualdad anterior con ayuda del Lema 3.2

$$\begin{aligned}
E_\varepsilon \left\| \frac{1}{n} \sum \varepsilon_i f(X_i) \right\|_{\mathscr{G}} &= \left\| \max_{f\in\mathscr{G}} \frac{1}{n} \left| \sum_{i=1}^n \varepsilon_i f(X_i) \right| \right\|_{1|X} \\
&\leq \left\| \max_{f\in\mathscr{G}} \frac{i}{n} \left| \sum_{i=1}^n \varepsilon_i f(X_i) \right| \right\|_{\psi_2|X} \\
&\leq K\sqrt{1+\log N(\varepsilon, \mathscr{F}_M, L_1(\mathbb{P}_n))} \cdot \\
&\qquad \max_{f\in\mathscr{G}} \left\| \frac{1}{n} \sum \varepsilon_i f(X_i) \right\|_{\psi^2|X}
\end{aligned}$$

donde la norma $\|\cdot\|_{\psi^2|X}$ es sobre $\varepsilon_1, \ldots, \varepsilon_n$ con $X_1, \ldots, X_n$ fijas. Por la desigualdad de Hoeffding la última norma se puede acotar de la manera siguiente

$$\begin{aligned}
\left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\psi^2|X} &\leq \frac{\sqrt{6}}{n} [\varepsilon_1^2 f(X_1)^2 + \cdots + \varepsilon_n^2 f(X_n)^2]^{1/2} \\
&= \sqrt{\frac{6}{n}} (\mathbb{P}_n f^2)^{1/2} \leq \sqrt{\frac{6}{n}} M
\end{aligned}$$

Por tanto

$$E_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathscr{F}_M} \to \varepsilon$$

cuando $n \to \infty$, de donde $E_\varepsilon \|n^{-1} \sum \varepsilon_i f(X_i)\|_{\mathscr{F}_M}$ converge a cero en probabilidad.

Como $E_\varepsilon \|n^{-1} \sum \varepsilon_i f(X_i)\|_{\mathscr{F}_M} < M$, por el Teorema de la Convergencia Dominada se tiene que

$$E_X E_\varepsilon \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{\mathscr{F}_M} \to 0$$

es decir $\|\mathbb{P}_n - P\|_{\mathscr{F}}^*$ converge a cero en media, la convergencia casi segura se obtiene del siguiente lema, ya que $\|\mathbb{P}_n - P\|_{\mathscr{F}}^*$ es una martingala decreciente. □

Lema 3.6 *Sea $\mathscr{F}$ una clase de funciones medibles con función envolvente F, tal que $PF^* < \infty$. Definamos las σ-álgebras Σ_n como la filtración generada* *por las funciones medibles $h : \Omega^\infty \longrightarrow \mathbb{R}$ que son simétricas en sus primeros n argumentos. Entonces*

$$E\left(\|\mathbb{P}_n - P\|^*_{\mathscr{F}} \mid \Sigma_{n+1}\right) \geq \|\mathbb{P}_{n+1} - P\|^*_{\mathscr{F}} \quad c.s.$$

Más aún, existen versiones medibles de $\|\mathbb{P}_n - P\|^*_{\mathscr{F}}$ adaptadas a la filtración, tales versiones forman una submartingala decreciente que converge casi seguramente a una constante.

DEMOSTRACIÓN: Supongamos que $Pf = 0$ para toda f. Consideremos la función $h : \Omega^n \to \mathbb{R}$ dada por $h(x_1, \ldots, x_n) = \|n^{-1}\sum_{i=1}^n f(x_i)\|_{\mathscr{F}}$ que es simétrica en sus argumentos, además existe h^* que también es simétrica en sus argumentos, por tanto $h^*(X_1, \ldots, X_n)$ es una versión Σ_n-medible de $\|\mathbb{P}_n\|^*_{\mathscr{F}}$

Sea $\mathbb{P}^i_n$ la medida empírica de las variables $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$, entonces $\mathbb{P}_{n+1}f = (n+1)^{-1}\sum_{i=1}^{n+1}\mathbb{P}^i_n f$ para cualquier f, de donde

$$\|\mathbb{P}_{n+1}\|^*_{\mathscr{F}} \leq \frac{1}{n+1}\sum_{i=1}^{n+1}\|\mathbb{P}^i_n\|^*_{\mathscr{F}} \quad \text{c.s.}$$

para cualquier versión medible, de esta manera se toma la versión de la izquierda que sea Σ_{n+1}-medible, entonces si sacamos la esperanza condicional respecto a Σ_{n+1} llegamos a

$$\|\mathbb{P}_{n+1}\|^*_{\mathscr{F}} \leq \frac{1}{n+1}\sum_{i=1}^{n+1} E\left(\|\mathbb{P}^i_n\|^*_{\mathscr{F}} \mid \Sigma_{n+1}\right) \quad \text{c.s.}$$

Notemos que si cambiamos $\|\mathbb{P}^i_n\|^*_{\mathscr{F}}$ por otra versión no se altera la desigualdad, ya que sólo difieren en un conjunto nulo. Entonces tomemos una versión de $h^*(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots X_{n+1})$ para $\|\mathbb{P}^i_n\|^*_{\mathscr{F}}$. Por ser h^* una función simétrica la $E\left(\|\mathbb{P}^i_n\|^*_{\mathscr{F}} \mid \Sigma_{n+1}\right)$ no depende de i, por lo cual

$$E\left(\|\mathbb{P}^i_n\|^*_{\mathscr{F}} \mid \Sigma_{n+1}\right) = E\left(\|\mathbb{P}^{n+1}_n\|^*_{\mathscr{F}} \mid \Sigma_{n+1}\right)$$

□

Otras extenciones del Teorema de Glivenko-Cantelli fueron propuestas en 1971 por Vapnik-Červonenkis, basadas en ideas de tipo combinatorio, aquellas calses de conjuntos en los que se puede extender el Teorema de Glivenko-Cantelli se les llama *clases V-C*. La idea se basa en ver cual es la cardinalidad máxima del coonjunto que es despedazado por una clase dde conjuntos, en el sentido de que todos los subconjuntos se puedan expresar como la intersección de algún elemento de la clase con el conjunto dado. R. Dudley [1984] da un recuento de algunos resultados posteriores que permiten facilmente verificar cuando una clase esn $V - C$, como se le conoce en la literatura. D. Pollard [1984] también de algunas extensiones y contraejemplos interesantes.

4 Teoremas de Donsker

El Teorema del Límite Central dice: Sean $X_1, X_2, \dots$ variables aleatorias independientes e idénticamente distribuídas con media cero y varianza 1 –lo referente a media y varianza es para simlificar notación, lo indispensable es que las variables aleatorias tengan segundo momento finito– entonces

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \Rightarrow N(0,1)$$

Notemos que al igual que la Ley de los Grandes Números, también aparece el factor $n^{-1} \sum X_i$, como en el capítulo anterior si se remplaza este término por la función de distribución empírica se obtiene

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

que M. D. Donsker prueba que converge en distribución para el caso de variables aleatrias independientes e idénticamente distribuías a un elemento Gaussiano Y llamado el Puente Browniano.

Ahora el reto es estudiar la convergencia de

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(X_i) - Ef(X_1))$$

a un elemento Gaussiano, con X_i variables aleatorias independientes y f en una familia $\mathscr{F}$ de funciones, como el el capítulo anterior.

Los dos Teoremas de Donsker como la sección de funciones unifofmemente acotadas estan basadas en el libro van der Vaart y Wellner [1996], los Teoremas de Donsker para los espacio C y D son del libro Billingsley [1994].

4.1 El Teorema de Donsker en C

De nueva cuenta vamos a trabajar con la función de distribución empírica, para ello consideremos $\xi_1, \xi_2, \dots$ variables aleatorias independientes con distribución uniforme en el intervalo $[0,1]$, entonces $F(t) \equiv t$. Sea $Y_n(\omega)$ unaa función con valor en t

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

un incoveniente que presenta la función $Y_n(t,\omega)$ es que no es continua; en la siguiente sección probaremos que el elemento aleatorio Y_n converge débilmente a un elemento Gaussiano, en la presente sección se sustituirá el elemento aleatorio Y_n por un elemento continuo Z_n similar a Y_n, para probar la convergencia débil de Z_n a W° en C.

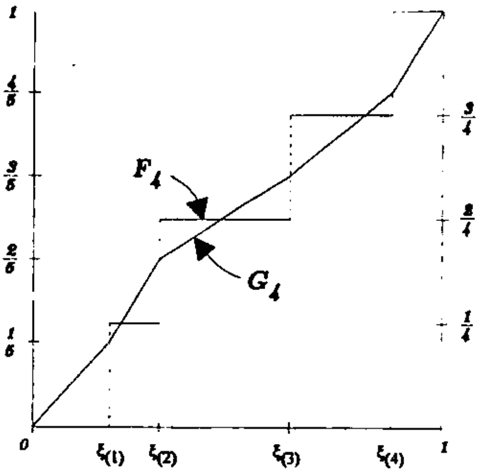

Figura 4.1: Gráfica de F_n, para $n = 4$.

Sea $G_n(t,\omega)$ como función de t, la función de distribución correspondiente a la distribución uniforme de masa $(n+1)^{-1}$ en cada uno de los $n+1$ intervalos $[\xi_{(i-1)}(\omega) - \xi_{(i)}(\omega)]$, donde $\xi_{(0)} = 0$, $\xi_{(n+1)} = 1$ y $\xi_{(1)}, \dots, \xi_{(n)}$ son los valores de $\xi_1, \dots, \xi_n$ en orden creciente. Notemos que

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

para toda $t \in [0,1]$. Definamos $Z_n(t,\omega) = \sqrt{n}(G_n(t,\omega) - t)$ que es un elemento aleatorio en C y además satisface

$$\sup_t |Y_n(t,\omega) - Z_n(t,\omega)| \leq \frac{1}{\sqrt{n}} \tag{4.1}$$

es decir, se ha sustituido la distribución empírica por G_n para obtener un elemento Z_n continuo.

Teorema 4.1 (Teorema de Donsker) *Sean ξ_n variables aleatorias independientes, uniformemente distribuídas en $[0,1]$, y Z_n un elemento aleatorio como se definió anteriormente, entonces*

$$Z_n \Rightarrow W^\circ$$

donde W° es el Puente Browniano.

DEMOSTRACIÓN: Sea $U_n(t,\omega) = nF_n(t,\omega)$, esto es, el número de puntos $\xi_1(\omega), \ldots, \xi_n(\omega)$ tal que $\xi_i(\omega) \leq t$. Sea $t_0, \ldots, t_k$ una partición del intervalo $[0,1]$, entonces las variables aleatorias

$$U_n(t_i) - U_n(t_{i-1}) \quad i = 1, \ldots, k.$$

tiene distribución binomial con parámetros n y $p_i = t_i - t_{i-1}$, ya que ξ_i se distribuye uniformemente en $[0,1]$. Por el Teorema del Límite Central se tiene que

$$Y_n(t_i) - Y_n(t_{i-1}) = \frac{1}{\sqrt{n}}\left((U_n(t_i) - U_n(t_{i-1})) - np_i\right)$$

converge en distribución a $W^\circ_{t_i} - W^\circ_{t_{i-1}}$ para $i = 1, \ldots, k$, por (4.1) se puede establecer lo mismo para $Z_n(t_i) - Z_n(t_{i-1})$, obteniendo así que las distribuciones de dimensión finita de Z_n convergen a las correspondientes de W°, sólo resta probar que $\{Z_n\}$ es tensa. Para ello es suficiente probar que para cada ε y η positivos existe δ, $0 < \delta < 1$ y n_0, tal que para $n \geq n_0$

$$P\left\{\sup_{t \leq s \leq t+\delta} |Z_n(s) - Z_n(t)| \geq \varepsilon\right\} \leq \delta\eta$$

para toda t, que como ya se hizo notar, podemos remplazar $Z_n(s) - Z_n(t)$ por $Y_n(s) - Y_n(t)$ sin nungún inconveniente. Dado que los incrementos de Y_n son estacionarios, no hay problema en considerar $t = 0$ para llegar a

$$P\left\{\sup_{s \leq \delta} |Y_n(s)| \geq \varepsilon\right\} \leq \delta\eta \tag{4.2}$$

Para p_1, p_2 probaremos que

$$E\{|Y_n(s+p_1) - Y_n(s)|^2 \cdot |Y_n(s+p_1+p_2) - Y_n(s+p_1)|^2\} \leq 6p_1p_2 \tag{4.3}$$

para despúes utilizar el Teorema 2.7, y poder acotar (4.2). Para $1 \leq i \leq n$ sea $\alpha_i = 1 - p_1$ si ξ_i esta en el intervalo $(s, s+p_1]$, de lo contrario $\alpha_i = -p_1$,

$\beta_i = 1 - p_2$ si ξ_i esta en $(s+p_1, s+p_1+p_2]$ o, $\beta_i = -p_2$ en otro caso. De esta manera (4.3) se puede reescribir como

$$E\left\{\left(\sum_{i=1}^{n}\alpha_i\right)^2\left(\sum_{i=1}^{n}\beta_1\right)^2\right\} \leq n^2 6p_1p_2$$

Ahora bien, dado que las variables aleatorias ξ_i son independientes, lo mismo se pude decir de los vectores (α_i, β_i), como las variables aleatorias ξ_i son uniformemente distribuidas, el vector (α_i, β_i) toma los valores de $(1-p_1, -p_2)$, $(-p_1, 1-p_2)$ y $(-p_1, -p_2)$ con probabilidad p_1, p_2 y $p_3 = 1 - p_1 - p_2$ respectivamente. Como $E\{\alpha_i\} = E\{\beta_i\} = 0$ se tiene

$$\begin{aligned} E\left\{\left(\sum_{i=1}^{n}\alpha_i\right)^2\left(\sum_{i=1}^{n}\beta_1\right)^2\right\} &= nE\{\alpha_1^2\beta_1^2\} + n(n-1)E\{\alpha_1^2\}E\{\beta_2^2\} + \\ & \quad 2n(n-1)E\{\alpha_1\beta_1\}E\{\alpha_2\beta_2\} \end{aligned}$$

pero además contamos con

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

probando así la desigualdad (4.3)

Para δ fija consideremos las variables

$$Y_n\left(\frac{i}{m}\delta\right) - Y_n\left(\frac{i-1}{m}\delta\right) \tag{4.4}$$

tomando $\gamma = 2$, $\alpha = 1$, $u_i = \sqrt{6}\delta/m$ para $1 \leq i \leq m$, ξ_i la variable (4.4) y junto con (4.3) empleamos el Teorema 2.7, esto es, si

$$M'_m = \max_{1\leq i\leq m} \min\left\{\left|Y_n\left(\frac{i}{m}\delta\right)\right|, \left|Y_n(\delta) - Y_n\left(\frac{i}{m}\delta\right)\right|\right\}$$

entonces $P\{M'_m \geq \varepsilon\} \leq 6K\delta^2/\varepsilon^4$ con $K = K_{2,1}$.

Si $M_m = \max_{1\leq i\leq m}|Y_n(i\delta/m)|$, se tiene que $M_m \leq M'_m + |Y_n(\delta)|$, con esto se llega a la desigualdad

$$P\{M_m \geq \varepsilon\} \leq \frac{2^4 \cdot 6K}{\varepsilon^4}\delta^2 + P\left\{|Y_n(\delta)| \geq \frac{\varepsilon}{2}\right\}$$

Para cada ω, $Y_n(t,\omega)$ es continua por la derecha en t, de manera que si $m \to \infty$, M_m converge a $\sup_{t\leq\delta}|Y_n(t,\omega)|$ para cada ω, obteniendo así

$$P\left\{\sup_{t\leq\delta}|Y_n(t,\omega)| > \varepsilon\right\} \leq \frac{2^4\cdot 6K}{\varepsilon^4} + P\left\{|Y_n(\delta)| \geq \frac{\varepsilon}{2}\right\}. \tag{4.5}$$

Por tanto se tiene $Y_n(\delta) \Rightarrow \sqrt{\delta(1-\delta)}N$, cuando $n \to \infty$, con δ fija. Por lo cual

$$P\left\{|Y_n(\delta)| > \frac{\varepsilon}{2}\right\} \to P\left\{N \geq \frac{\varepsilon/2}{\sqrt{\delta(1-\delta)}}\right\} \leq \frac{2^4\delta^2}{\varepsilon^4}E\{N^4\} = \frac{3\cdot 2^4}{\varepsilon^4}\delta^2$$

Para n mayor que un n_δ contamos con

$$P\left\{|Y_n(\delta)| > \frac{\varepsilon}{2}\right\} < \frac{6\cdot 2^4}{\varepsilon^4}\delta^2$$

y juntos con (4.5) se tiene

$$P\left\{\sup_{t\leq\delta}|Y_n(t,\omega)| > \varepsilon\right\} \leq \frac{6\cdot 2^4(K+1)}{\varepsilon^4}\delta^2$$

De esta manera para ε, η dados se elige δ de manera que $6\cdot 2^4(K+1)/\varepsilon^4 < \delta/\eta$, de aquí para para $n > n_\delta$ se tiene (4.2), es decir $\{Z_n\}$ en tensa y converge débilemte a W°. □

4.2 El Teorema de Donsker en D

Sean $\xi_1, \xi_2, \ldots$ variables aleatorias independientes e idénticamente distribuídas tales que $0 \leq \xi_i \leq 1$, sea $Y_n(\omega)$ la función dada por

$$Y_n(t,\omega) = \sqrt{n}(F_n(t,\omega) - F(t)) \tag{4.6}$$

con t en $[0,1]$, donde $F(t)$ es la función de distribución común de las variables aleatorias ξ_i. Como $Y_n(t)$ es una variable aleatoria para cada t, y $Y_n(\omega)$ pertenece a D para cada ω, Y_n es un elemento aleatorio en D.

Ahora si se probará la convergencia débil de Y_n a un elemento Gaussiano en D, y además se descarta la hipótesis de que las variables aleatorias ξ_i estén uniformemente distribuídas en el intervalo $[0,1]$. El proceso estocástico $\{Y_n(t) : 0 \leq t \leq 1\}$ es llamado *proceso empírico*.

Teorema 4.2 (Teorema de Donsker) *Sean $\xi_1, \xi_2, \ldots$ variables aleatorias independientes con distribución común $F(t)$. Si los elementos aleatorios Y_n están definidos como en (4.6), entonces*

$$Y_n \Rightarrow Y$$

donde Y es un elemento Gaussiano en D determinado por

$$E\{Y(t)\} = 0 \tag{4.7}$$

$$E\{Y(s)Y(t)\} = F(s)(1 - F(t)) \qquad \text{con } s \leq t \tag{4.8}$$

DEMOSTRACIÓN: De la misma manera en que se extendió la medida de Wiener en el espacio $(D, \mathcal{B}(D))$, se puede extender el Puente Browniano $W°$. También $W°$ denotará un elemento aleatorio en D, cuya distribución es la del Puente Browniano.

Supongamos que las variables aleatorias ξ_i se distribuyen uniformemente en el intervalo $[0, 1]$. Sea U_n el número de puntos $\xi_1(\omega), \ldots, \xi_n(\omega)$ que no exceden un punto t. Sea $\{t_i\}$ una partición del intervalo $[0, 1]$, entonces las variables aleatorias $U_n(t_i) - U_n(t_{i-1})$ se distribuyen de manera binomial con parámetros n y $p = t_i - t_{i-1}$, por el Teorema del Límite Central las distribuciones de finita dimensión de Y_n convergen débilmente a las respectivas de $W°$.

Por el Teorema 2.20 es suficiente demostrar

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

para $t_1 \leq t \leq t_2$, la cual se probó en la sección anterior justamente en la demostración del Teorema de Donsker, por tanto $Y_n \Rightarrow W°$.

Supongamos que las variables aleatorias ξ_i tienen una distribución arbitraria F sobre $[0, 1]$; definamos la función φ como

$$\varphi(s) = \inf\{t : s \leq F(t)\}$$

Entonces $s \leq F(t)$ si y sólo si $\varphi(s) \leq t$, si η_n es una variable aleatoria distribuida uniformemente en $[0, 1]$ se tiene que $P\{\varphi(\eta_n) \leq t\} = F(t)$, definamos $\xi_i = \varphi(\eta_i)$ con η_i variables aleatorias independientes distribuídas uniformemente en $[0, 1]$.

Sea $G_n(\cdot, \omega)$ la distribución empírica de las variables aleatorias $\eta_1, \ldots, \eta_n$ y $Z_n(t, \omega) = \sqrt{n}(G_n(t, \omega) - t)$ entonces $Z_n \to W°$ según el caso que se probó. Pero la distribución empirica de $\xi_1(\omega), \ldots, \xi_n(\omega)$ es justamente $F_n(t, \omega) = G_n(F(t), \omega)$, de donde $Y_n(t, \omega) = Z_n(F(t), \omega)$. Sea $\psi : D \to D$ una función dada por $(\psi x)(t) = x(F(t))$, si x_n converge a x en C en la topología de

Skorohod entonces la convergencia es uniforme, por lo que ψx_n converge uniformemente a ψx, y por consigueinte en la topología de Skorohod . Como $Z_n \Rightarrow W^\circ$, entonces $Y_n = \psi(Z_n) \Rightarrow \psi(W^\circ)$, y definamos $Y = \psi(W^\circ)$ que es un elemento Gaussiano que satisface (4.7). □

4.3 Funciones Uniformemente Acotadas

Sea T un conjunto arbitrario, denotemos por $\ell^\infty(T)$ al conjunto de las funciones $z : T \to \mathbb{R}$ que son uniformemente acotadas, esto es,

$$\|z\|_T = \sup_{t \in T} |z(t)| < \infty$$

definiendo $d(x, y) = \|x - y\|_T$, $(\ell^\infty(T), d)$ es un espacio métrico.

El papel que desempeña el espacio $\ell^\infty(T)$ es importante para los procesos estocásticos cuyas trayectorias son acotados. Tal es el caso del espacio C, donde se trabaja sobre un subconjunto de $\ell^\infty([0, 1])$, el de las funciones continuas.

Para probar tensión de $\{X_n\}$ en $\ell^\infty(T)$, se pedirá que las variaciones de X_n sean casi acotadas en subconjuntos de T, formalmente esto es, dado $\varepsilon, \eta > 0$ existe una partición finita $\{T_i\}$ de T tal que

$$\limsup_n P^* \left(\sup_i \sup_{s,t \in T_i} |X_n(s) - X_n(t)| > \varepsilon \right) < \eta \tag{4.9}$$

Diremos que una sucesión $\{X_n\}$ es *asintóticamente tensa* si para cualquier ε positivo existe un conjunto compacto K tal que

$$\liminf_n P_*(X_n \in K^\delta) \geq 1 - \varepsilon$$

para toda $\delta > 0$.

Teorema 4.3 *Una sucesión $\{X_n\}$ con $X_n : \Omega_n \to \ell^\infty(T)$ es asintóticamente tensa si y sólo si $X_n(t)$ es asintóticamente tensa en $\mathbb{R}$ para cada t y para todo $\varepsilon, \eta > 0$ existe una partición finita $\{T_i\}$ de T que satisfece (4.9).*

DEMOSTRACIÓN: Probemos la condición de suficiencia. Para una partición arbitraria $T = \cup_{i=1}^k T_i$ se tiene

$$\|X_n\|_T \leq \max_i |X_n(t_i)| + \varepsilon$$

con probabilidad $1 - \eta$, de donde $\|X_n\|_T$ es asintóticamente tensa en $\mathbb{R}$.

Sea ζ positivo, $\{\varepsilon_m\}$ una sucesión tal que $\varepsilon_m \to 0^+$ y $M \in \mathbb{R}$ tal que $\limsup P^*(\|X_n\| > M) < \zeta$; para cada $\varepsilon = \varepsilon_m$ y $\eta = 2^{-m}\zeta$ se elige una partición $\{T_i\}$ de T de manera que

$$\limsup P^*\left(\sup_i \sup_{s,t\in T_i} |X_n(s) - X_n(t)| > \varepsilon\right) < \eta$$

Sean $z_1, \ldots, z_n$ elementos de $\ell^\infty(T)$ constantes en los conjuntos T_i y que sólo tomen los valores de $0, \pm\varepsilon_m, \ldots, \pm[M/\varepsilon_m]\varepsilon_m$. Llamemos K_m a la unión de las bolas cerradas con centro en z_i y radio ε_m, entonces para m fija se tiene

$$\|X_n\|_T \leq M \quad \text{y} \quad \sup_i \sup_{s,t\in T_i} |X_n(s) - X_n(t)| \leq \varepsilon_m$$

de donde $X_n \in K_m$. Sea $K = \cap_{m=1}^{\infty} K_m$, por ser cerrado y totalmente acotado K es compacto en $\ell^\infty(T)$, y además para $\delta > 0$ existe m tal que $\cap_{i=1}^{m} K_i \subset K^\delta$, por tanto

$$\limsup P^*(X_n \notin K^\delta) \leq \limsup P^*\left(X_n \notin \bigcap_{i-1}^{m} K_i\right) \leq \zeta + \sum_{i=1}^{m} \zeta 2^{-m} < 2\zeta$$

La condición necesaria es consecuencia del siguiente teorema. □

El segundo y último resultado es más complicado y además es necesario considerar la siguiente definición. Una sucesión $\{X_n\}$ con $X_n : \Omega_n \to \ell^\infty(T)$ es *asintóticamente uniformemente ρ-continua en probabilidad* si para cada $\varepsilon, \eta > 0$ existe $\delta > 0$ tal que

$$\limsup_n P^*\left(\sup_{\rho(s,t)<\delta} |X_n(s) - X_n(t)| > \varepsilon\right) < \eta$$

con ρ una métrica en T.

Teorema 4.4 *Una sucesión $\{X_n\}$ con $X_n : \Omega_n \to \ell^\infty(T)$ es asintóticamente tensa si y sólo si para cada t, $X_n(t)$ es asintóticamente tensa en $\mathbb{R}$ y si existe una semimétrica ρ en T tal que el espacio (T, ρ) es totalmente acotado y X_n es asintóticamente uniformemente ρ-equicontinua en probabilidad.*

DEMOSTRACIÓN: Sea $\delta > 0$ tal que

$$\limsup_n P^*\left(\sup_{\rho(s,t)>\delta} |X_n(s) - X_n(t)| > \varepsilon\right) < \eta$$

como T es totalmente acotado, T puede ser cubierto por un número finito de bolas de radio menor que δ, obteniendo así una partición de T y la prueba de la condición de suficiencia.

Sea $\{X_n\}$ una sucesión asintóticamente tensa, y $K_1 \subset K_2 \subset \cdots$ conjuntos compactos tales que para cada $\varepsilon > 0$ $\liminf P_*(X_n \in K_m^\varepsilon) \geq 1 - 1/m$. Definamos la semimétrica ρ_m en T como

$$\rho_m(s,t) = \sup_{z \in K_m} |z(s) - z(t)|$$

con s,t en T y m fija; así T es totalmente acotado, ya que K_m puede ser cubierto por un número finito de bolas de radio η y centro en z_i.

Ahora definamos la semimétrica

$$\rho(s,t) = \sum_{m=1}^{\infty} \frac{1}{2^m}(\rho_m(s,t) \wedge 1)$$

Para $\eta > 0$ fijo se elige m entero positivo tal que $2^{-m} < \eta$, veamos que T es totalmente acotado bajo ρ. Para cada m se cubre T con un número finito de ρ_m-bolas con centros en los puntos $t_1, \ldots, t_p$, como $\rho_1 \leq \rho_2 \leq \cdots$ para cada t existe t_i tal que $\rho(t,t_i) \leq \sum_{k=1}^m 2^{-k}\rho_k(t,t_i) + 2^{-m} < 2\eta$, de donde (T,ρ) es totalmente acotado.

Sea z en K_m, entonces $|z(s) - z(t)| \leq \rho_m(s,t)$ y $\rho_m(s,t) \wedge 1 \leq 2^m\rho(s,t)$, si $\|z_0 - z\|_T < \varepsilon$ con z en K_m, entonces $|z_0(s) - z_0(t)| < 2\varepsilon + |z(s) + z(t)|$ y

$$K_m^\varepsilon \subset \left\{ z : \sup_{\rho(s,t)<2^{-m}\varepsilon} |X_n(s) - X_n(t)| \leq 3\varepsilon \right\}$$

y por tanto

$$\liminf P_* \left(\sup_{\rho(s,t)<\delta} |X_n(s) - X_n(t)| < 3\varepsilon \right) \geq 1 - \frac{1}{m}$$

□

4.4 Dos Teoremas de Donsker

Sea P la distribución común de las variables aleatorias X_i, con $i \geq 1$, el *proceso empírico* $\mathbb{G}_n$ se define como

$$f \to \mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P)f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(X_i) - Pf)$$

con índices en el conjunto de funciones $\mathcal{F}$ y las variables aletorias X_i independientes e idénticamente distribuídas.

Dada una función medible f tal que $Pf^2 < \infty$, entonces

$$\mathbb{G}_n f \Rightarrow N(0, P(f - Pf)^2)$$

según el Teorema del Límite Central. Al igual que el capítulo anterior, respecto a los teoremas de Glivenko-Cantelli, daremos condiciones necesarias para obtener esta convergencia uniformemente sobre una familia $\mathcal{F}$ de funciones.

Para poder hablar de una versión uniforme del Teorema del Límite Central de una familia $\mathcal{F}$ tenemos que suponer que

$$\sup_{f\in\mathcal{F}} |f(x) - Pf| < \infty$$

para toda x, de esta manera el proceso empírico $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ tiene trayectorias acotadas, es decir, en $\ell^\infty(\mathcal{F})$. Diremos que una clase $\mathcal{F}$ de funciones es una *clase Donsker* si existe un elemento tenso $\mathbb{G}$ en $\ell^\infty(\mathcal{F})$ tal que

$$\mathbb{G}_n \Rightarrow \mathbb{G}$$

Si $\mathbb{G}_n \Rightarrow \mathbb{G}$, entonces las distribuciones de dimensión finita de $\mathbb{G}_n$ convergen a las respectivas distribuciones de $\mathbb{G}$, por tanto el proceso $\{\mathbb{G}f : f \in \mathcal{F}\}$ es un proceso Gaussiano con media cero y además para f_1, f_2 en $\mathcal{F}$

$$E\mathbb{G}f_1\mathbb{G}f_2 = P(f_1 - Pf_1)(f_2 - Pf_2) = Pf_1f_2 - Pf_1Pf_2$$

esta propiedad junto con la de tensión de $\mathbb{G}$ determinan la distribución de $\mathbb{G}$, que es el Puente Browniano. Por ejemplo una clase finita de funciones cuadrado integrable es Donsker. Para determinar si $\mathcal{F}$ es una clase Donsker, dependerá del tamaño de $\mathcal{F}$, para hacer esto se utilizarán lo números de entropía definidos anteriormente.

Por el Teorema 4.4 una clase $\mathcal{F}$ es Donsker si y sólo si $\mathcal{F}$ es totalmente acotada en $L_2(P)$ y asintóticamente equicontinua. La definición de asintóticamente equicontinua se puede reescribir como: Si $\mathcal{F}_{\delta_n} = \{f-g : f, g \in \mathcal{F}, \|f-g\|_{P,2} < \delta_n\}$, entonces

$$\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \to 0$$

en probabilidad exterior, para cualquier sucesión $\{\delta_n\}$ tal que $\delta_n \to 0^+$.

Probaremos dos teoremas referentes al Teorema del Límite Central Empírico, para ello supondremos que la función envolvente es cuadrado integrable y la cota uniforme de entropía

$$\int_0^\infty \sup_Q \sqrt{\log N(\varepsilon\|F\|_{Q,2}, \mathscr{F}, L_2(Q))}d\varepsilon < \infty \tag{4.10}$$

donde el supremo es sobre todas las medidas Q de probabilidad discretas en $(\Omega, \mathscr{B})$ y $\|F\|_{Q,2}^2 = \int F^2 dQ > 0$. A (4.10) se le llama la *condición uniforme de entropía.*

En los dos Teoremas de Donsker se involucran dos integrales, que miden el cambio que hay en los números de entropía de la clase $\mathscr{F}$ y determinarán si es una clase Donsker.

Teorema 4.5 *Sea $\mathscr{F}$ una clase de funciones medibles que satisfacen la cota uniforme de entropía. Para cada $\delta > 0$, sea $\mathscr{F}_\delta = \{f - g : f, g \in \mathscr{F}, \|f - g\|_{P,2} < \delta\}$ y $\mathscr{F}_\infty{}^2$ clases P-medibles. Si $P^*F^2 < \infty$, entonces $\mathscr{F}$ es. P-Donsker.*

DEMOSTRACIÓN: Sea $\{\delta_n\}$ una sucesión tal que $\delta_n \to 0^+$, por la desigualdad de Chebyshev, el Lema de Simetrización y por ser cada función f medible se tiene

$$\begin{aligned} P^*(\|\mathbb{G}_n\|_{\mathscr{F}_{\delta_n}} > x) &\le \frac{1}{x}E^*\|\mathbb{G}_n\|_{\mathscr{F}_{\delta_n}} \\ &\le \frac{2}{x}E^*\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathscr{F}_{\delta_n}} \\ &= \frac{2}{x}E_X E_\varepsilon \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathscr{F}_{\delta_n}} \end{aligned}$$

Tomando las variables aleatorias $X_1, \ldots, X_n$ fijas, consideremos el proceso estocástico

$$f \to \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i f(X_i)\right\}$$

que por la desigualdad de Hoeffding resulta ser un proceso sub-Gaussiano bajo la norma $\|f\|_n = \sqrt{n^{-1}\sum f^2(X_i)}$, entonces por el Corolario 3.1

$$E_\varepsilon \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{\mathscr{F}_{\delta_n}} \le K\int_0^\infty \sqrt{\log N(\varepsilon, \mathscr{F}_{\delta_n}, L_2(\mathbb{P}_n))}d\varepsilon \tag{4.11}$$

Para valores de ε grandes la familia $\mathcal{F}_{\delta_n}$ esta contenida en la bola de radio ε y centro en el origen, llamemos

$$\theta_n^2 = \sup_{f \in \mathcal{F}_{\delta_n}} \|f\|_n^2 = \left\| \frac{1}{n} \sum_{i=1}^{n} f^2(X_i) \right\|_{\mathcal{F}_{\delta_n}}$$

si $\varepsilon > \theta_n$, $\log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)) = 0$, y así en (4.11) basta con considerar la integral sobre el intervalo $[0, \theta_n]$.

Dado que los números cubrientes de $\mathcal{F}_{\delta_n}$ están acotados por los números cubrientes de $\mathcal{F}_\infty$ y $N(\varepsilon, \mathcal{F}_\infty, L_2(Q)) \leq N^2(\varepsilon/2, \mathcal{F}, L_2(Q))$ para cualquier medida de probabilidad Q, por un cambio de variable la integral (4.11) se puede acotar por

$$\int_0^{\theta_n/\|F\|_n} \sup_Q \sqrt{\log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))} \|F\|_n d\varepsilon$$

donde el supremo es sobre todas las medidas de probabilidad discretas.

Como $\sup\{Pf^2 : f \in \mathcal{F}_{\delta_n}\} \to 0$ y $\mathcal{F}_{\delta_n} \subset \mathcal{F}_\infty$ hay que demostar que

$$\|\mathbb{P}_n f^2 - Pf^2\|_{\mathcal{F}_\infty} \to 0$$

en probabilidad, es decir, la Ley de los Grandes Números para la clase $\mathcal{F}_\infty^2$, esta clase tiene como una envolvente a la función $(2F)^2$ que es medible; sean f, g en $\mathcal{F}_\infty$ como $|f| \leq 2F$, entonces

$$\mathbb{P}_n|f-g|^2 = \mathbb{P}_n|f-g|2F \leq \|f-g\|_n \|4F\|_n.$$

Por tanto $N(\varepsilon\|2F\|_n^2, \mathcal{F}_\infty{}^2, L_1(\mathbb{P}_n)) \leq N(\varepsilon\|F\|_n, \mathcal{F}_\infty, L_2(\mathbb{P}_n))$, pero este último número esta acotado, entonces $\log N(\varepsilon\|F\|_n, \mathcal{F}_\infty, L_2(\mathbb{P}_n)) = o^*(n)$, por el Teorema 3.5 se tiene la equicontinuidad asintótica.

Veamos que $\mathcal{F}$ es totalmente acotado en $L_2(P)$, por lo anterior existe una sucesión de medidas discretas $\{P_n\}$ tal que

$$\|(P_n - P)f^2\|_{\mathcal{F}_\infty} \to 0$$

Sea n tal que $\|(P_n - P)f^2\|_{\mathcal{F}_\infty} < \varepsilon^2$, como $N(\varepsilon, \mathcal{F}, L_2(P_n))$ es finito, cualquier ε-red de $\mathcal{F}$ en $L_2(P_n)$ es una $\sqrt{2}\varepsilon$-red en $L_2(P)$, por tanto $\mathcal{F}$ es totalmente acotado. $\square$

Consideremos la colección $\mathcal{F}$ de las funciones $1\{(-\infty, t]\}$ con t en $\mathbb{R}$, entonces

$$N(\varepsilon, \mathcal{F}, L_2(Q)) \leq N_{[]}(\varepsilon^2, \mathcal{F}, L_1(Q)) \leq \frac{2}{\varepsilon^2}$$

para $\varepsilon \leq 1$ y cualquier medida Q. Como $\int_0^1 \log(1/\varepsilon) < \infty$, $\mathscr{F}$ es Donsker.

Para el siguiente teorema sera necesaria la "*norma*" $L_{2,\infty}$ dada por

$$\|f\|_{P,2,\infty} = \sup_{x>0}(x^2 P(|f| > x))^{1/2}$$

que no satisface la desigualdad del triángulo, sin embargo es posible encontrar una norma equivalente a esta salvo por un factor 2; para nuestros propósitos no habra ningún inconveniente trabajar con ella. Además por la desigualdad de Chebyshevse se tiene $\|f\|_{P,2,\infty} \leq \|f\|_{P,2}$.

También sera necesario utilizar la desigualdad de Bernstein que asegura para una función f acotada cuadrado integrable

$$P(|\mathbb{G}_n f| > x) \leq 2\exp\left\{-\frac{1}{2}\frac{x^2}{Pf^2 - 1/3\|f\|_\infty x/\sqrt{n}}\right\}$$

por la desigualdad anterior y por el Lemma 3.4 se tiene

$$E\|\mathbb{G}_n\|_{\mathscr{F}} \leq K\left[\max_f \frac{\|f\|_\infty}{\sqrt{n}} \log|\mathscr{F}| + \max_f \|f\|_{P,2}\sqrt{\log|\mathscr{F}|}\right] \tag{4.12}$$

para una clase $\mathscr{F}$ finita.

Diremos que X tiene segundo momento débil si $P(|X| > t) = o(t^{-2})$ cuando $t \to \infty$ y por tanto $E|X|1\{|X| > t\} \leq o(t^{-1})$ si $t \to \infty$.

Teorema 4.6 *Sea $\mathscr{F}$ una clase de funciones medibles tal que*

$$\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathscr{F}, L_{2,\infty}(P))}d\varepsilon + \int_0^\infty \sqrt{\log N(\varepsilon, \mathscr{F}, L_2(P))}d\varepsilon < \infty$$

y supongamos que la función envolvente F de $\mathscr{F}$ tiene segundo momento débil. Entonces $\mathscr{F}$ es P-Donsker.

DEMOSTRACIÓN: Sea q un entero positivo, entonces $\mathscr{F}$ puede ser cubierto con N_q^1 $L_2(P)$-bolas ajenas de radio 2^{-q}, y con N_q^2 $L_{2,\infty}(P)$-brackets ajenos de radio 2^{-q}, si consideremos la intersección de estas dos familias se tienen $N_q \leq N_q^1 N_q^2$ bolas ajenas de donde $\mathscr{F} = \cup_{i=1}^{N_q} \mathscr{F}_{qi}$, $\log N_q \leq \log N_q^1 + \log N_q^2$ y

$$\sum 2^{-q}\sqrt{\log N_q} < \infty$$

$$\|(\sup_{f,g\in\mathscr{F}_{qi}} |f-g|)^*\|_{P,2,\infty} < 2^{-q}$$

$$\sup_{f,g\in\mathscr{F}_{qi}} \|f-g\|_{P,2} < 2^{-q}$$

y las particiones subsecuentes se pueden tomar como un refinamiento de la partición anterior.

Para cada q se elige una función f_{qi} en $\mathscr{F}_{qi}$ para definir las funciones

$$\begin{aligned} \pi_q f &= f_{qi} \quad f \in \mathscr{F}_{qi} \\ \Delta_q f &= \sup_{f,g \in \mathscr{F}_{iq}} |f-g|^* \quad f \in \mathscr{F}_{qi} \end{aligned}$$

Basta demostrar que $\|\mathbb{G}_n(f-\pi_{q0}f)\|_{\mathscr{F}} \to 0$ en probabilidad si q_0 y n tienden a infinito. Para $q \geq q_0$ y n fijos se definen

$$\begin{aligned} a_q &= 2^{-q}/\sqrt{\log N_{q+1}} \\ A_{q-1}f &= 1\{\Delta_{q_0}f \leq \sqrt{n}a_{q_0}, \ldots, \Delta_{q-1}f \leq \sqrt{n}a_{q-1}\} \\ B_q f &= 1\{\Delta_{q_0}f \leq \sqrt{n}a_{q_0}, \ldots, \Delta_{q-1}f \leq \sqrt{n}a_{q-1}, \Delta_q f > \sqrt{n}a_q\} \\ B_{q_0}f &= 1\{\Delta_{q_0}f > \sqrt{n}a_{q_0}\} \end{aligned}$$

hay que notar que las funciones $A_q f$ y $B_q f$ son constantes en $\mathscr{F}_{qi}$, veamos que

$$\begin{aligned} f - \pi_{q_0}f &= (f-\pi_{q_0}f)B_{q_0}f + \sum_{q=q_0+1}^{\infty}(f-\pi_q f)B_q f \\ &\quad + \sum_{q=q_0+1}^{\infty}(\pi_q f - \pi_{q-1}f)A_{q-1}f \end{aligned} \tag{4.13}$$

ya que la función $B_q f$ es cero para toda q o bién existe un sólo punto q_1 tal que $B_{q_1}f = 1$. En el primer caso los primeros dos sumandos son cero, $A_q = 1$ para toda q y la serie converge a $f - \pi_{q_0}f$, en el segundo caso $A_{q-1}f = 1$ si y sólo si $q \leq q_1$, obteniendo $f - \pi_{q_0}f$.

Aplicaremos $\mathbb{G}_n$ en (4.13), tomaremos el supremo sobre todas las funciones f de $\mathscr{F}$, y probar que cada uno de los sumandos converge a cero en probabilidad.

Como $|f-\pi_{q_0}f|B_{q_0}f \leq 2F1\{2F > \sqrt{n}a_{q_0}\}$ entonces

$$E^*\|\mathbb{G}_n(f-\pi_{q_0}f)B_{q_0}f\|_{\mathscr{F}} \leq 4\sqrt{n}P^*F1\{2F > a_{q_0}\}$$

dado que F tiene segundo momento débil, el lado derecho de esta desigualdad converge a cero con q_0 fija cuando $n \to \infty$, con esto se tiene la convergencia del primer sumando.

Dado que las particiones estan refinadas $\Delta_q f B_q f \leq \Delta_{q-1} f B_q f$ se tiene que

$$\sqrt{n}a_q P(\Delta_q f B_q f) \leq 2\|\Delta_q f\|^2_{P,2,\infty} \leq 2 \cdot 2^{-q}$$

dado que $\Delta_{q-1}fB_qf \leq \sqrt{n}a_{q-1}$ para $q \geq q_0$ se tiene

$$\begin{aligned} P(\Delta_{q-1}fB_qf)^2 &\leq \sqrt{n}a_{q-1}P(\Delta_q f 1\{\Delta_q > \sqrt{n}a_q\}) \\ &\leq 2\frac{a_{q-1}}{a_q}2^{-2q} \end{aligned}$$

Por (4.12) se llega a

$$\begin{aligned} E^*\|\sum_{q=q_0+1}^{\infty} \mathbb{G}_n(f-\pi_q f)B_qf\|_{\mathscr{F}} &\leq \sum_{q=q_0+1}^{\infty} E^*\|\mathbb{G}_n\Delta_q fB_qf\|_{\mathscr{F}}+ \\ &\quad \sum_{q=q_0+1}^{\infty} 2\sqrt{n}\|P\Delta_q fB_qf\|_{\mathscr{F}} \\ &\leq K\left[\sum_{q=q_0+1}^{\infty} a_{q-1}\log N_q + \right. \\ &\quad \left. \sqrt{\frac{a_{q-1}}{a_q}}2^{-q}\sqrt{\log N_q} + \frac{4}{a_q}2^{-2q}\right] \\ &\leq K\left[\sum_{q=q_0+1}^{\infty} 2^{-q}\sqrt{\log N_q}\right] \end{aligned}$$

que converge a cero cuando $q_0 \to \infty$.

Para el tercer sumando, notemos que a lo más hay N_q funciones $\pi_q f - \pi_{q-1}f$ y N_{q-1} funciones $A_{q-1}f$, entonces $|\pi_q f - \pi_{q-1}f|A_{q-1}f \leq \Delta_{q-1}fA_{q-1}f \leq \sqrt{n}a_{q-1}$, de nuevo por (4.12)

$$\begin{aligned} E^*\left\|\sum_{q=q_0+1}^{\infty} \mathbb{G}(\pi_q f - \pi_{q-1}f)A_{q-1}f\right\|_{\mathscr{F}} &\leq \\ K\left[\sum_{q=q_0+1}^{\infty} a_{q-1}\log N_q + 2^{-q}\sqrt{\log N_q}\right] \end{aligned}$$

□

Existen varios autores pioneros en el campo de extender a clases más generales el Teorema de Donsker, entre ellos podemos citar a Alexander [1987], Dudley [1978],[1984], Gaenssler [1983] , Giné y Zinn [1984], Ledoux y Talagrand [1990], Ossiander [1987], Pollard [1984] y Talagrand [1988] ente otros.

Apéndice A Medida y Probabilidad

A.1 Teoría de la Medida

Al hablar de medida nos referimos a una función μ definida en una clase $\mathscr{B}$ de subconjuntos de Ω, pero esta clase de subconjuntos no es tan arbitraria, le pediremos que satisfaga las siguientes propiedades:

- $\omega \in \mathscr{B}$
- Si $A \in \mathscr{B}$, entonces $A^c \in \mathscr{B}$
- Si $A_1, A_2, \ldots \in \mathscr{B}$, entonces $\bigcup_{i=1}^{\infty} A_i \in \mathscr{B}$

A la colección de subconjuntos $\mathscr{B}$ que satisface estas tres condiciones se le llama *σ-álgebra*. Si $\mathscr{B}$ satisface las primeras dos condiciones, y la última condición la cumple sólo para un número finito de conjuntos A_i, entonces $\mathscr{B}$ recibe el nombre de *álgebra de conjuntos*. Una σ-álgebra importante es la *σ-álgebra de Borel* de S, la generada por los subconjuntos abiertos de S, donde S es un espacio métrico. El par $(\Omega, \mathscr{B})$ se le llama *espacio medible*, y los elementos de $\mathscr{B}$ reciben el nombre de conjuntos $\mathscr{B}$-mebibles.

Una colección $\mathscr{M}$ de subconjuntos de Ω es una *clase monótona* si para cualquier sucesión monótona creciente $\{E_n\}$ de $\mathscr{M}$ y cualquier sucesión monótona decresiente $\{F_n\}$ de $\mathscr{M}$, los conjuntos

$$\bigcup_{n=1}^{\infty} E_n \quad \text{y} \quad \bigcap_{n=1}^{\infty} F_n$$

pertenecen a $\mathscr{M}$. La importancia de las clases monótonas esta dada por el siguiente teorma, y por que generalmente es más facil el manejo de una clase monótona que una σ-álgebra.

Teorema A.1 (Clases Monótonas) *Si $\mathscr{A}$ es una álgebra de conjuntos, entonces la σ-álgebra $\mathscr{F}$ generada por $\mathscr{A}$ coincide con la clase monótona $\mathscr{M}$ generada por $\mathscr{A}$.*

Se tiene un espacio medible $(\Omega, \mathscr{B})$ y una función f en Ω, las funciones que serán de nuestro interés son las siguientes.

Definición. *Una función $f : \Omega \to \mathbb{R}$ es una función medible si para cada a en $\mathbb{R}$ el conjunto*

$$\{x : \Omega : f(x) > a\}$$

pertenece a $\mathscr{B}$.

La *función indicadora* 1_A de un conjuto $A \in \mathscr{B}$, se define como

$$1_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

que por supuesto es una función medible, simpre y cuando A sea miembro de $\mathscr{B}$, en ocasiones escribiremos la función indicadora como $1\{A\}$.

Como se dijo en un principio vamos a considerar una función μ sobre $\mathscr{B}$, la cual le llamaremos *medida* que debe satisfacer:

- $\mu(\emptyset) = 0$
- $\mu(A) \geq 0$ para todo $A \in \mathscr{B}$
- Si $A_1, A_2, \ldots \in \mathscr{B}$ son conjuntos ajenos, entonces

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

La terna $(\Omega, \mathscr{B}, \mu)$ se le conoce como *espacio de medida*. Si $\mu(\Omega) < \infty$ diremos que μ es una *medida finita*, ahora si existe una sucesión de conjuntos $\{A_i\}$ de $\mathscr{B}$ tal que $\Omega = \cup A_i$ y $\mu(A_i) < \infty$ para toda i, entonces diremos que μ es una *medida σ-finita.*

Denotemos por $M(\Omega, \mathscr{B})$ a la colección de todas las funciones $\mathscr{B}$-medibles de Ω a $\mathbb{R}^*$, y a la colección de las funciones no negativas de Ω a $\mathbb{R}^*$ $\mathscr{B}$-medibles por $M^+(\Omega, \mathscr{B})$, donde $\mathbb{R}^*$ es el conjunto de los reales extendidos. Veamos dos teoremas referentes a convergencia que serán de gran utilidad en este trabajo.

Teorema A.2 (Covergencia Monótona) *Si $\{f_n\}$ es una sucesión monótona creciente de funciones en $M^+(\Omega, \mathscr{B})$ que convege a f, entonces*

$$\int f d\mu = \lim_n \int f_n d\mu.$$

Lema A.1 (Fatou) *Si* $\{f_n\}$ *es una sucesión de funciones de* $M^+(\Omega, \mathcal{B})$, *entonces*

$$\int (\liminf_n f_n) d\mu \leq \liminf_n \int f_n d\mu$$

Si $\{f_n\}$ es una sucesión de funciones tales que $f_n : \Omega \to \mathbb{R}$ para toda n, se dice que $\{f_n\}$ *converge puntualmente* a f si $f_n(x) \to f(x)$ para cada x en Ω; esta convergencia se pude debilitar a *convergencia casi segura* (c.s.), es decir, $f_n(x) \to f(x)$ para toda x en $\Omega \setminus A$ con $\mu(A) = 0$.

Teorema A.3 (Convergencia Dominada de Lebesgue) *Sea* $\{f_n\}$ *una sucesión de funciones integrables que convergen casi seguramente a una función* f *real-valuada. Si existe una función* g *integrable tal que* $|f_n| \leq g$ *para toda* n, *entonces* f *es integrable y*

$$\int f d\mu = \lim_n \int f_n d\mu$$

A.2 Teoría de Probabilidad

Una *medida de probabilidad* P es una medida finita en $(\Omega, \mathcal{B})$ tal que $P(\Omega) = 1$ y una *variable aleatoria* es simplemente una función $\mathcal{B}$-medible.

Lema A.2 (Borel-Cantelli) *Sea* $(\Omega, \mathcal{B}, P)$ *un espacio de probabilidad y sean* $A_1, A_2 \ldots \in \mathcal{B}$ *tal que* $\sum_{n=1}^{\infty} P(A_n) < \infty$, *entonces* $P(\limsup_n A_n) = 0$.

Una función de distribución es una función $F(x) = F(x_1, \ldots, x_k)$ en $\mathbb{R}^k$ que satisface las siguientes tres propiedades:

(*i*) F es continua por abajo.

(*ii*) F es no decreciente, $0 \leq F(x) \leq 1$ para toda x, y para cada rectángulo $(a, b]$ de dimensión k

$$\sum \pm F(a_1 + \theta_1 d_1, \ldots, a_k + \theta_k d_k) \geq 0$$

donde $d_i = b_i - a_i$, la suma corre sobre todas los vectores $(\theta_1, \ldots, \theta_k)$ con componentes 0 y 1 y el signo + o − depende si el vector tiene un número par de entradas cero o impar.

(*iii*) $F(x) \to 0$, cuando cualquier componente de x tiende a $-\infty$, y $F(x) \to 1$ si todas las componentes de x tienden a ∞.

Si P es una medida de probabilidad en $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$, entonces

$$F(x) = P\{y : y \leq x\}$$

es una función de distribución.

Teorema A.4 (Helly) *Sea $\{F_n\}$ una sucesión de funciones de distribución en $\mathbb{R}^k$, entonces existe una subsucesión $\{F_{n'}\}$ y una función F que satisface las condiciones (i) y (ii) tal que*

$$\lim_{n' \to \infty} F_{n'}(x) = F(x)$$

en todo punto x de continuidad de F.

Definición. *Sea $(\Omega, \mathscr{B}, P)$ un espacio de probabilidad, $\{X_n\}$ una sucesión de variables aleatorias definidas en $(\Omega, \mathscr{B}, P)$ y $\mathscr{F}_1 \subset \mathscr{F}_2 \cdots$ una sucesión creciente de sub σ-álgebras de $\mathscr{F}$, supngamos que X_n es $\mathscr{F}_n$-medible. Entonces $\{X_n\}$ es una martingala relativa a $\mathscr{F}_n$ si $E(X_{n+1} \mid \mathscr{F}_n) = X_n$ c.s. para toda n; es una submartingala si $E(X_{n+1} \mid \mathscr{F}_n) \geq X_n$ c.s. y es una supermartingala si $E(X_{n+1} \mid \mathscr{F}_n) \leq X_n$ c.s.*

Si $\{\mathscr{F}_n\}$ una sucesión decreciente de sub σ-álgebras de $\mathscr{F}$, y $\{X_n\}$ es una sucesión de variables aleatorias tal que $E(X_n \mid \mathscr{F}_{n+1}) = X_{n+1}$, entonces se dice que $\{X_n\}$ es una martingala decreciente respecto a $\{\mathscr{F}_n\}$.

Presentamos dos desigualdades importantes en la teoría de probabilidad; la primera de ellas, la Desigualdad de Chebyshev que es usada a menudo en este trabajo por cual conviene tenerla en mente.

Teorema A.5 (Desigualdad de Chebyshev) *Sea X una variable aleatoria no negativa, $0 < p < \infty$ y $0 < \varepsilon < \infty$, entonces*

$$P\{X \geq \varepsilon\} \leq \frac{E(X^p)}{\varepsilon^p}$$

Una función $g : \mathbb{R} \to \mathbb{R}$ es *convexa* si

$$g(ax + (1-a)y) \leq ag(x) + (1-a)g(y)$$

para todo $x, y \in \mathbb{R}$ y $a \in [0,1]$.

Teorema A.6 (Desigualdad de Jensen) *Sea $g : (a,b) \to \mathbb{R}$ una función convexa, X una variable aleatoria en $(\Omega, \mathscr{F}, P)$ tal que $X(\omega) \in (a,b)$ para todo ω y $E(X)$ es finita. Entonces $Eg(X) \geq g(E(X))$.*

A.3 Minorantes Medibles

Sea $(\Omega, \mathcal{B}, P)$ un espacio de probabilidad y $T : \Omega \to \mathbb{R}^*$ una función arbitraria, la *integral exterior* de T respecto a P se define como

$$E^*T = \inf \{EU : U \geq T, U : \Omega \to \mathbb{R}^* \text{ medible y } EU \text{ existe}\}$$

La *probabilidad exterior* de un subconjunto B de Ω se define como

$$P^*(B) = \inf \{P(A) : B \subset A, A \in \mathcal{B}\}$$

Análogamente se pueden definir la *integral interior* de T y la *probabilidad interior* de B, y obtener las relaciones $E_*T = -E^*(-T)$ y $P_*(B) = 1 - P^*(B^c)$.

Un hecho importante es, que en la definición de integral exterior se alcanza el ínfimo que es de gran ayuda y probaremos en el sigueinte lema, a la función para la cual se logra el ínfimo la denotaremos por T^*. Lo mismo se puede decir para la probabilidad exterior como veremos más adelante.

Lema A.3 *Dada una función $T : \Omega \to \mathbb{R}^*$, existe un función medible $T^* : \Omega \to \mathbb{R}^*$ tal que:*

(i) $T^* \geq T$

(ii) $T^* \leq U$ *c.s., para toda función mebible* $U : \Omega \to \mathbb{R}^*$ *con* $U \geq T$ *c.s.*

Además para cualquier T^* que satisfaga estas condiciones se tiene $E^*T = ET^*$ simpre y cuando ET^* exista, es decir si $E^*T < \infty$.

DEMOSTRACIÓN: Sea $\{U_m\}$ una sucesión de fuciones medibles tal que $U_m \geq T$ y $E \arctan U_m \to E^* \arctan T$, llamemos

$$T^*(\omega) = \lim_{m \to \infty} \inf_{1 \leq k \leq m} U_k(\omega)$$

de esta manera T^* es una función medible y $T^* \geq T$. Como la función arctan es creciente por el Teorema de la Convergencia Monótona $E \arctan T^* = E^* \arctan T$.

Sea U una función medible con $U \geq T$, entonces $\arctan U \wedge T^* \geq \arctan T$ de donde $E \arctan U \wedge T^* \geq E^* \arctan T = E \arctan T^*$. Pero $\arctan U \wedge T^* \leq \arctan T^*$ entonces

$$E \arctan U \wedge T^* = E \arctan T^*$$

por lo que $\arctan U \wedge T^* = \arctan T^*$ c.s. lo que es mejor, $T^* \leq U$ c.s.

Si ET^* existe, por (i) $ET^* \geq E^*T$ y por (ii) $ET^* \leq E^*T$, por lo cual $ET^* = E^*T$. Si $E^*T < \infty$, exite U función medible tal que $U \geq T$ y $EU^+ < \infty$, por tanto $E(T^*)^+ \leq EU^+$ es decir ET^* existe. □

La función T^* es llamada *mínimo mejorante medible* de T, de igual manera se define la función T_* como *máximo minorante medible* de T como $T_* = -(-T^*)$. Ambas funciones son únicas salvo conjuntos de medida cero.

Lema A.4 *Sean $S, T : \Omega \to \mathbb{R}^*$ funciones arbitrarias, entonces se satisfacen las sigueintes relaciones c.s.*

(i) $(S+T)^* \leq S^* + T^*$, *si S es medible se tiene igualdad.*

(ii) $(S-T)^* \geq S^* - T^*$.

(iii) $|S^* - T^*| \leq |S+T|^*$.

(iv) Si S es medible, entonces $(ST)^* = S1\{S>0\}T^* + S1\{S<0\}T_*$.

(v) $(ST)^* \leq S^*T^*1\{S^*>0, T^*>0\} + S^*T_*1\{S^*<0, T_*>0\} + S_*T^*1\{S_*>0, T^*<0\} + S_*T_*1\{S_*<0, T_*<0\}$.

(vi) $(1\{T>c\})^* = 1\{T^*>c\}$ *para* $c \in \mathbb{R}$.

(vii) $|T|^* = T^* \vee (-T)^* = T^* \vee (-T)_* = |T^*| \vee |T_*|$.

(viii) $(S \vee T)^* = S^* \vee T^*$.

(ix) $(S \wedge T)^* \leq S^* \wedge T^*$, *si S es medible se tiene igualdad.*

Además $P^*(T>c) = P(T^*>c)$ *para* $c \in \mathbb{R}$.

Demostración: De la relación $S^* \geq S$ y $T^* \geq T$ se obtiene (i); para el caso en que S es medible $T+S \leq (T+S)^*$, por el lema anterior $T^* \leq (T+S)^* - S$ esto es $T^* + S \leq (T+S)^*$. Para la desigualdad (iii) se tiene

$$S^* - T^* \leq (S - \overline{T})^* \leq |S-T|^*$$

Para (iv) sea U una función medible tal que $U \geq ST$, entonces por ser S una función medible

$$\begin{aligned}
U1\{S>0\} &\geq ST1\{S>0\} \\
U/S1\{S>0\} &\geq T1\{S>0\} \\
U/S1\{S>0\}+T^*1\{S\leq 0\} &\geq T \\
U/S1\{S>0\}+T^*1\{S\leq 0\} &\geq T^* \\
U/S1\{S>0\} &\geq T^*1\{S>0\} \\
U1\{S>0\} &\geq ST^*1\{S>0\}
\end{aligned}$$

como $U \geq (-S)(-T)$, también se tiene $U1\{-S>0\} \geq (-S)(-T)^*1\{-S>0\}$. De ambas desigualdades llegamos a $U \geq S1\{S>0\}T^* + S1\{S<0\}T_*$, ya que $U1\{S=0\} \geq 0$, de donde $(ST)^* \geq S1\{S>0\}T^* + S1\{S<0\}T_*$. La desigualdad $(ST)^* \leq S1\{S>0\}T^* + S1\{S<0\}T_*$ no presenta mayor problema.

Dado que $ST \leq S^*T1\{T>0\} + S_*T1\{T<0\}$, considerando que S^* y S_* son funciones medibles junto con (iv) y el hecho que $(T1\{T>0\})^* \leq T^*1\{T^*>0\}$ resulta la desigualdad en (v)

Sea U una función medible tal que $U \geq 1\{T>c\}$, entonces la función $S = T^*1\{U \geq 1\} + (T^* \wedge c)1\{U<1\}$ es medible y $S \geq T$, por tanto $S \geq T^*$, y $T^* \leq c$ si $U < 1$, de donde $1\{T^*>c\} \leq U$ lo cual implica que $(1\{T>c\})^* \geq 1\{T^*>c\}$

Es cierto que $|T|^* \leq T^* \vee (-T)^* \leq T^* \vee (-T)_* \leq |T^*| \vee |T_*|$, sólo veamos que $|T^*| \vee |T_*| \leq |T|^*$. Por (iii) se tiene $|T^*| \leq |T|^*$ y también $|T_*| = |(-T)^*| \leq |-T|^* = |T|^*$, probando $|T^*| \vee |T_*| = |T|^*$.

$S^* \vee T^*$ es un mejorante medible de $S \vee T$, probemos la desigualdad restante. Sea U una función medible con $U \geq S$ y $U \geq T$, por tanto $U \geq S^*$ y $U \geq T^*$ de donde $U \geq S^* \vee T^*$.

La desigualdad $(S \wedge T)^* \leq S^* \wedge T^*$ se sigue del hecho que $S \wedge T \leq S^* \wedge T^*$. Supongamos que S es una función medible, entonces $(S \wedge T)^* \leq S \wedge T^*$, sea U una función medible tal que $U \geq S \wedge T$, entonces $U1\{U \leq S\} \geq T1\{U<S\}$, por tanto $U1\{U \leq S\} \geq (T1\{U<s\})^* = T^*1\{U<S\}$ de donde $U \geq T^* \wedge S$.

Por último, desarrollando $P(T^* > c) \geq P^*(T>c) \geq E^*1\{T>c\} = E1^*\{T>c\} = P(T^*>c)$ por tanto $P^*(T>c) = P(T^*>c)$ □

Al igual que en el caso de integral exterior, la probabilidad exterior también logra su ínfimo, el conjunto medible en cual se logra lo denotaremos por B^*. El siguiente lema análogo al Lema A.3 respecto probabilidad exterior, y además muestra que la probabilidad exterior es una caso particular de la integral exterior.

Lema A.5 *Para cualquier subconjunto B de Ω se tiene:*

(i) $P^*(B) = E^*1\{B\}$ *y* $P_*(B) = E(1\{B\})_*$

(ii) Existe un conjunto medible B^*, tal que $B \subset B^*$ y $P(B^*) = P^*(B)$; además para dicho conjunto B^* se tiene $1\{B^*\} = (1\{B\})^*$

(iii) $(1\{B\})^* + (1\{\Omega - B\})_* = 1$

DEMOSTRACIÓN: Por definición de integral exterior $P^*(B) \geq E^*1\{B\}$; ahora consideremos el conjunto $A = \{1^*\{B\} \geq 1\}$ entonces $B \subset A$ y

$$E^*1\{B\} = E1^*\{B\} \geq P(A) \geq P^*(B)$$

obteniendo la primera parte de (i) y (ii) con $B^* = A$.

Dado que $P_*(B) = 1 - P^*(\Omega - B) = 1 - E(1 - 1\{B\})^* = 1 - E(1 - (1\{B\})_*)$ por tanto $P_*(B) = E(1\{B\})_*$. Si $B \subset B^*$ entonces $1\{B^*\} \leq (1\{B\})^*$, por otro lado $E1\{B^*\} = P(B^*) = E(1\{B\})^*$. Términado (i) y (ii) por completo.

Para probar (iii), $(1\{\Omega - B\})_* = (1 - 1\{B\})_* = 1 - (1\{B\})^*$. □

El mejorante T^* depende de la medida de probabilidad P, y hasta el momento sólo hemos logrado resultados para una sóla medida de probabilidad; en el sigueinte resultado romperemos este esquema trabajando con una familia de medidas de probabilidad. Dimremos que una familia de medidas de probabilidad $\mathscr{P}$ es *dominada* si existe una medida P que domina a cada una de las medidas de $\mathscr{P}$.

Antes veamos que si P y P_0 medidas de probabilidad en $(\Omega, \mathscr{B})$ con densidades p y p_0 respectivamente, definamos los conjuntos $\Omega^a = \{p > 0, p_0 > 0\}$ y $\Omega^\perp = \{p = 0 \text{ ó } p_0 = 0\}$, que son conjuntos medibles ajenos y cuya unión es Ω, además satisfacen que $P(\Omega^\perp) = 0$ y P_0 es absolutamente continua respecto a P en Ω^a.

Lema A.6 *Sea $\mathscr{P}$ una clase de medidas de probabilidad en $(\Omega, \mathscr{B})$ dominada y $T : \Omega \to \mathbb{R}^*$ una función. Entonces existe una función $T^* : \Omega \to \mathbb{R}^*$ tal que*

(i) $T^* \geq T$

(ii) $T^* \leq U$ *P-c.s. para toda función $U : \Omega \to \mathbb{R}^*$ con $U \geq T$ P-c.s., para toda P en $\mathscr{P}$.*

DEMOSTRACIÓN: Sea P_0 una medida de probabilidad que domina a $\mathscr{P}$, y T^* una función medible que satisface (i) y (ii) para $\mathscr{P} = \{P_0\}$.

Sea P en $\mathscr{P}$ $(P \ll P_0)$, entonces Ω es la unión de los conjuntos medibles ajenos Ω^a, $\Omega^\perp$ tal que $P(\Omega^\perp) = 0$ y $P_0 \ll P$ en Ω^a. Sea U una función medible tal que $U \geq T$ P-c.s., entonces $U1\{\Omega^a\} \geq T1\{\Omega^a\}$ P_0-c.s. y además $U1\{\Omega^a\}$ es medible. Por tanto $U1\{\Omega^a\} \geq (T1\{\Omega^a\})^{*P_0} = T^*1\{\Omega^a\}$ P_0-c.s. Como $\mathcal{P}$ es dominada por P_0 esta última desigualdad es P-c.s, dado que $P(\Omega^{\perp}) = 0$, $U \geq T^*$ P-c.s. □

El Teorema de Fubini no es valido para integrales exteriores, que es una gran desventaja, pero a cambio se tiene el siguiente lema como un sustituto del Teorema de Fubini.

Sea T una función real valuada definida en $(\Omega_1 \times \Omega_2, \mathcal{B}_1 \times \mathcal{B}_2, P_1 \times P_2)$, por E^*T se entendera la esperanza exterior como se ha estado trabajando y por $E_1^*E_2^*T$ la esperanza exterior definida como: para cada ω_1 se tiene

$$(E_2^*T)(\omega_1) = \inf \int U(\omega_2)dP_2(\omega_2)$$

donde el ínfimo es sobre todas las funciones medibles $U : \Omega_2 \to \mathbb{R}^*$, con $U(\omega_2) \geq T(\omega_1, \omega_2)$ para cada ω_2 simpre y caundo $\int UdP_2$ exista, dando paso a $E_1^*E_2^*T$ que es la integral exterior de la función $E_1^*T : \Omega_2 \to \mathbb{R}^*$. Análogamente se defina la esperanza interior.

Lema A.7 (Teorema de Fubini) *Sea T una función definida en el producto de dos espacios de probabilidad. Entonces $E_*T \leq E_{1*}E_{2*}T \leq E_1^*E_2^*T \leq E^*T$.*

Demostración: Supongamos que $E^*T < \infty$, por tanto $E^*T = ET^*$ y $(T^*)^+$ es integrable.

Dado que la función T^* es medible respecto a la σálgebra producto, la función dada por $\omega_2 \to T^*(\omega_1, \omega_2)$ es una función mejorante medible de la función $\omega_2 \to T(\omega_1, \omega_2)$ para cada ω_1 por tanto

$$\int T^*(\omega_1, \omega_2)dP_2(\omega_2) \geq (E_2^*T)(\omega_1)$$

para cada ω_1, pero

$$\begin{aligned}\int T^*(\omega_1,\omega_2) \quad & dP_2(\omega_2) = \qquad\qquad (A.1)\\ & \int (T^*)^+(\omega_1, \omega_2)dP_2(\omega_2) - \int (T^*)^-(\omega_1, \omega_2)dP_2(\omega_2)\end{aligned}$$

Por el Teorema de Fubini los dos términonis de la izquierda son funciones medibles de ω_1, de donde (A.1) es una función mejorante de $\omega_1 \to (E_2^*T)(\omega_1)$, integrando respectoa P_1 resulta ser una cota superior de $E_1^*E_2^*T$.

Nuevamente por el Teorema de Fubini, la doble intragral puede ser remplazada por la integral respecto a la medida producto, de donde $E(T^*)^+ - E(T^*)^- \geq E_1^* E_2^* T$, esto es $E^* \geq E_1^* E_2^* T$.

Considerando ahora $-T$ se llega a $E_* T \leq E_{1*} E_{2*} T$ y la desigualdad restante es consecuancia de la definción. $\square$

tiene punto límite lo cual es una contradicción. Sea $\{x_n\}$ una sucesión fundamental con punto límite x, entonces $\{x_n\}$ converge a x.

$d \Rightarrow a$ Sea $\{G_\alpha\}$ un cubierta abierta de A tal que no contiene ninguna subcubierta finita que cubra al conjunto A. Como A es totalmente acotado para cada n existen un número finito de bolas abiertas $B_{n1}, \ldots, B_{nk_n}$ de radio 2^{-n} que cubren al conjunto A. $\square$

Recordemos que $C_0(S)$ es el conjunto de las funciones acotadas reales definidas en el espacio métrico S; una *funcional lineal* en $C_0(S)$ es una función Λ con valores reales tal que $\Lambda(\alpha f + \beta g) = \alpha\Lambda(f) + \beta\Lambda(g)$ para toda $f, g \in C_0(S)$ y $\alpha, \beta \in \mathbb{R}$, una funcional Λ es *acotada* si existe una constante M tal que $|\Lambda(f)| \leq M\|f\|$ para toda f en X, la norma de una funcional se define como

$$\|\Lambda\| = \sup\{|\Lambda(f)| : \|f\| le1\}$$

sobre toda f en $C_0(S)$ no nulo.

Por último diremos que una funcional Λ es positiva si $\Lambda(f) \geq 0$ para toda $f \geq 0$.

Sea $J = [a, b]$, consideremos al conjunto $C(J)$ como la colección de todas las funciones continuas del intervalo J a $\mathbb{R}$ con la métrica del supremo, entonces veamos el siguiente teorema.

Teorema B.2 (Representación de Riesz) *Sea Λ una funcional positiva en $C(J)$, entonces existe una medida μ en $\mathcal{B}(\mathbb{R})$ tal que*

$$\Lambda(f) = \int_J f d\mu$$

para toda f en $C(J)$. Además $\|\Lambda\| = \mu(J)$.

Apéndice B Miscelánea

Sea (S, d) un espacio métrico, una cubierta abierta es una colección de conjuntos abiertos de S, y una cubierta abierta de $A \subset S$ es una colección de conjuntos abiertos $\{A_\alpha\}$ tal que $A \subset \cup_\alpha A_\alpha$. Se dice que un conjunto $A \subset S$ es *compacto* en S si para toda cubierta abierta de A contiene una subcubierta finita de A.

Una ε-red de un conjunto A es un conjunto de puntos $\{x_k\}$ con la propiedad de que para todo punto x de A existe un punto x_k tal que $d(x, x_k) < \varepsilon$; los puntos x_k no necesariamente tienen que pertenecer al conjunto A. Un conjunto A es *totalmente acotado* si tiene un ε-red finita para toda ε positiva.

Un conjunto A es *completo* si toda sucesión en A converge a un punto en A.

Teorema B.1 *Sea (S, d) un espacio métrico y $A \subset S$, entonces las siguientes condiciones son equivalentes:*

(a) *$\overline{A}$ es compacto.*

(b) *Toda cubierta numerable de $\overline{A}$ tiene un subcubierta finita.*

(c) *Toda sucesión en A tiene un punto límite.*

(d) *A es totalmente acotado y $\overline{A}$ es completo.*

DEMOSTRACIÓN:

$a \Rightarrow b$ Por definición de compacidad.

$b \Rightarrow c$ Sea $\{x_n\}$ una sucesuón en A y F_n la cerradura del conjunto $\{x_k : k \geq n\}$, si $\cap_n F_n =$ entonces los conjuntos F_n^c forman una cubierta abierta numerable de A, por tanto $A \subset F_1^c \cup \cdots F_n^c$ para alún n, lo cual implica $F_n \cap A = 0$. Entonces $\cap_n F_n$ contiene un punto x que es límite de $\{x_n\}$.

$c \Rightarrow d$ Supongamos que A no es totalmente acotado, entonces existe $\varepsilon > 0$ y una sucesión $\{x_n\}$ tal que $d(x_n, x_m) \geq \varepsilon$ para $n \neq m$ esto es $\{x_n\}$ no

Referencias

ALEXANDER, K.

[1987] *The central limit theorem for empirical processes on Vapnik-Červonenkis classes*, Ann. Probab. **17**, 737-759.

ASH, R.

[1972] *Real analisis and probability*, Academic Press, New York.

BARTLE, R. G.

[1995] *The elements of integration and lebesgue measure*, John Wiley & Sons, New York.

BILLINGSLEY P.

[1994] *Probability and measure*, John Wiley & Sons, New York.

[1968] *Convergence of probability measures*, John Wiley & Sons, New York.

DEHARDT, J.

[1971]

it Generalizations of the Glivenko-Cantelli therem, Ann. Math. Statist., **6**, 2050-2055.

DUDLEY, R.

[1978] *Central limit theorems for empirical measures*, Ann. Probab. **6**, 899-929.

[1984] *A course on empirical processes*, Ecole d'Eté de Probabilités de St-Flour 1982, Lecture Notes in Mathematics, vol. 1097, Spinger, Berlin Heidelberg, 2-142.

ETHIER, S. N., KURTZ, T. G.

[1986] *Markov processes: characterization and convergence*, John Wiley & Sons, New York.

GAENSSLER, P.
[1983] *Empirical processes*, Inst. Math. Statist., Lecture Notes Monograph Series, vol 3.

GINÉ, E., ZINN, J.
[1984] *Some limit theorems for empirical processes*, Ann. Probab., **12**, 929-989.

IRIBARREN, I. L.
[1973] *Topología de espacios métricos*, Limusa-Wiley, México.

LAHA, R. G., ROHATGI, V. K.
[1979] *Probability theory*, John Wiley & Sons, New York.

LEDOUX, M., TALAGRAND., M
[1990] *Probability in Banach Spaces*, Springer-Verlag, Berlin.

OSSIANDER, M.
[1987] *A central limit theorem under metric entropy with L_2- bracketing*, Ann. Probab., **15**, 897-919.

PARTHASARATHY, K. R.
[1967] *Probability measure on metric spaces*, Academic Press, New York.

POLLARD, D.
[1984] *Convergence of sthocastic processes*, Spinger, Berlin Heidelberg.

PYKE, R
[1984] *Asymptotic results for empirical and partial sums processes: a review*, Canadian Journal of Statistics, **12**, 241-264

TALAGRAD, M.
[1988] *Donsker classes of sets.* Probab. Theor. Rel. Fields, **78**, 169-191.

WELLNER, J. A.
[1992] *Empirical processes in action: a review*, International Statistics Review, **60**, 247-269

VAN DER VAART, A. W., WELLNER, J. A.
[1996] *Weak convergence and empirical processes*, Springer-Verlag, New York.

Índice Analítico

Ojos en que rebervera
la estrella crespuscular,
ojos verdes como el mar,
como el mar por la ribera;
ojos de lumbre hechicera
que ignoráis lo que es llorar,
glorificad mi pesar.
¡No me desoléis así!
¡Tened compasión de mí,
ojos verdes como el mar!

SALVADOR DÍAZ MIRÓN, 1895.