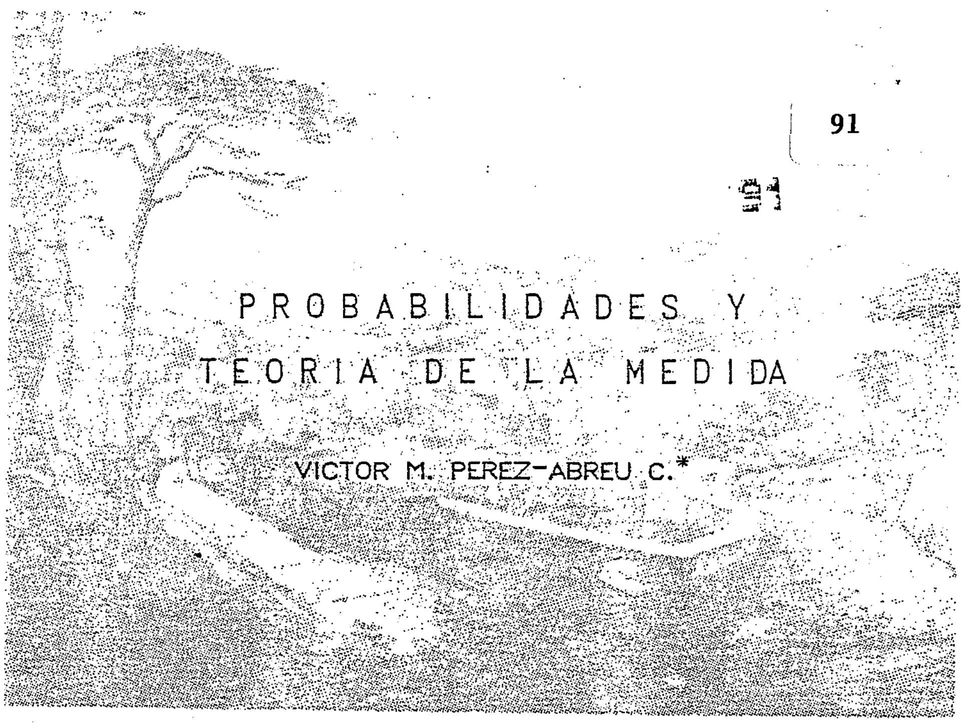
COMUNICACIONES DEL CIMAT

91



PROBABILIDADES Y TEORIA DE LA MEDIDA

VICTOR M. PEREZ-ABREU C.*

Notas del Curso "Probabilidades y Teoría de la Medida", impartido en la Universidad de El Salvador, San Salvador, del 15 al 19 de febrero de 1988, dentro del VII Curso Centro Americano y del Caribe de Matemáticas.

* Investigador de Tiempo Completo

CENTRO DE INVESTIGACION EN MATEMATICAS

Apartado Postal 402
Guanajuato, Gto.
México
Tels. (473) 2-25-50
2-02-58

INTRODUCCION

Estas notas corresponden al curso de Probabilidades y Teoría de la Medida impartido dentro del Séptimo Curso Centro Americano y del Caribe de Matemáticas del 15 al 19 de febrero de 1988. Este evento tuvo como sede organizadora a la Universidad de El Salvador y el patrocinio de esta Universidad y la Organización de Estados Americanos.

El desarrollo del curso se vio modificado y enriquecido no sólo por la crítica y los comentarios de los asistentes, sino por el entusiasmo e interés de un gran número de alumnos y colegas participantes y en especial de los organizadores. Mi iniciativa de escribir estas notas surgió de su entusiasmo y espíritu.

El material presentado está dirigido a profesores de Matemáticas, egresados de una licenciatura en Matemáticas, así como a estudiantes de los últimos semestres de la misma. Se supone que el lector tiene conocimientos de teoría de conjuntos, análisis real e introducción a la probabilidad. El objetivo es presentar la relación entre la teoría moderna de la probabilidad y la teoría de la medida, tratando de cubrir ideas y resultados básicos sin perderse en detalles o contra ejemplos patológicos. He tratado de cubrir sólo los ejemplos y demostraciones de resultados que no son comunes en textos usuales de Teoría de la Medida. Otras veces he preferido indicar la idea o técnica empleada en la demostración de un resultado, con el doble objetivo, primero, de alentar al lector a hacerla y, por otro lado, el tratar de cubrir el mayor material posible y llegar de una manera rápida a temas que por falta de tiempo no se ven en un curso común de Probabilidades y Teoría de la Medida.

La elección del material presentado, así como el orden, reflejan mi interés en los Procesos Estocásticos y la Estadística Matemática, ramas de la Matemática que se apoyan en la Teoría de Probabilidad. Esta es la razón de la presentación temprana de los ejemplos que se dan de espacios de probabilidad. Lo mismo es cierto para temas como teoremas límites en Estadística, esperanza condicional y distribuciones de extremos. Por otro lado, la rapidez del cuso y la presentación conjunta de temas de Probabilidad y Teoría de la Medida han hecho que muchos conceptos se introduzcan en un orden no usual y que muchas veces se hable de ellos sin un estudio previo y detallado de los mismos.

El curso fue pensado inicialmente para llevarse a cabo en siete horas y media. La versión final de las notas incluye temas no vistos en el curso como esperanza condicional, dominios de atracción y distribuciones de extremos. Pienso que catorce horas de exposición sería un tiempo razonable para cubrir todo el material.

Estas notas no contienen motivaciones ni cubren aspectos filosóficos o históricos del tema, en los cuales se hizo énfasis durante el desarrollo del curso. La presentación no pretende ser más que el de unas notas complementarias de un curso rápido de introducción a los fundamentos de la Teoría Moderna de Probabilidad, desde un punto de vista de la Teoría de la Medida. Espero que este trabajo motive a los asistentes a un estudio profundo y detallado del tema.

El curso y la elaboración de las notas fueron posibles gracias al esfuerzo de los organizadores, en especial de los Profesores Rolando Lemus Gómez, Francisco Marroquín y Juan Agustín Cuadra, a quienes expreso mi sincero agradecimiento.

Guanajuato, Gto., mayo de 1988

Víctor M. Pérez Abreu C.

1. AXIOMAS DE KOLMOGOROV

Un **espacio de Probabilidad** es una terna $(\Omega, \mathcal{F}, P)$ en donde

i) Ω es un conjunto no vacío (que representa los posibles resultados de un experimento aleatorio).

ii) $\mathcal{F}$ es una **σ-álgebra de subconjuntos** de Ω.

iii) P es una **medida de Probabilidad** sobre $\mathcal{F}$.

En las siguientes secciones estudiaremos y precisaremos los conceptos involucrados en estos axiomas.

2. CLASES DE CONJUNTOS

Definición 1. Una **σ-álgebra $\mathcal{F}$ de Ω** es una colección de subconjuntos de Ω que cumple las siguientes propiedades:

i) $\Omega \in \mathcal{F}$.

ii) Ai $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.

iii) Si $\left\{A_n\right\}_{n\geq 1}$ es una colección de subconjuntos de $\mathcal{F}$ entonces $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Otros sistemas de conjuntos útiles en la construcción de medidas de probabilidad son los siguientes:

Definición 2. Una colección $\mathcal{A}$ de subconjuntos de Ω se llama **álgebra** si satisface las siguientes propiedades:

i) $\Omega \in \mathcal{A}$

ii) Si $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.

iii) Si $A_1, \ldots, A_n \in \mathcal{A}$ entonces $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$

Observación: Toda σ-álgebra es un álgebra.

Definición 3. Una colección $\mathcal{S}$ de subconjuntos de Ω se llama **semi-álgebra** si cumple las siguientes propiedades:

i) $\Omega \in \mathscr{S}$.

ii) Si $A, B \in \mathscr{S} \Rightarrow A \cap B \in \mathscr{S}$

iii) Si $A \in \mathscr{S} \Rightarrow$ existen $A_1, \ldots A_n \in \mathscr{S}$, mutuamente ajenos tal que $A^c = \bigcup_{i=1}^{n} A_i$.

Observación: **Todo álgebra es una semi-álgebra**

Proposición 1. Sea $\mathscr{S}$ una semi-álgebra de subconjuntos de Ω y definamos

$$\mathscr{A}(\mathscr{S}) = \left\{ \sum_{i=1}^{k} A_i \; : \; A_1, \ldots A_n \in \mathscr{S} \text{ ajenos } k \geq 0 \right\}$$

Entonces $\mathscr{A}(\mathscr{S})$ es un álgebra llamada el **álgebra generada por $\mathscr{S}$.**

Notación: $\sum_{i=1}^{k} A_i$ denota unión de conjuntos mutuamente ajenos.

Proposición 2. Sea $\mathscr{C}$ una colección arbitraria de subconjuntos de Ω. Entonces siempre existe una mínima σ-algebra de Ω que contiene a $\mathscr{C}$, llamada la **σ-álgebra generada por $\mathscr{C}$** y denotada por $\sigma(\mathscr{C})$.

Demostración. Consideremos la colección Λ de todas las σ-álgebras $\mathscr{F}_\lambda$ que contienen a $\mathscr{C}$. Esta colección es no vacía pues 2^Ω la σ-álgebra potencia (de todos los subconjuntos de Ω) está en Λ. Es fácil probar que la intersección arbitraria de σ-álgebras es una σ-álgebra. Entonces

$$\sigma(\mathscr{C}) = \bigcap_{\lambda \in \Lambda} \mathscr{F}_\lambda .$$

∎

Si S es un espacio métrico $\mathcal{O}$ y la colección de todos los conjuntos abiertos, entonces $\sigma(\mathcal{O})$ (la cual siempre existe por la Proposición 2) se llama la σ-álgebra de Borel de S y será denotada por $\mathscr{B}(S)$.

3. MEDIDAS

En esta sección Ω denotará un conjunto no vacío.

Definición 4. Sea $\mathcal{A}$ un álgebra de subconjuntos de Ω. Una función $\mu : \mathcal{A} \to [0,\infty]$ se llama **medida finitamente aditiva** en $\mathcal{A}$ si

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

para todo par de conjuntos ajenos A y B en $\mathcal{A}$. La función de conjunto μ se llama **medida σ-aditiva** o simplemente **medida** si para toda colección de conjuntos mutuamente ajenos $A_1, A_2, \ldots,$ de $\mathcal{A}$ tal que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$(*) se tiene que

$$\mu\left[\sum_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} \mu(A_n).$$

Definición 5. Una medida P sobre una σ-álgebra se llama **Probabilidad** si $P(\Omega) = 1$.

El siguiente concepto es importante para probar la unicidad en algunos resultados de teoría de la medida.

Definición 6. Una medida μ sobre un álgebra $\mathcal{A}$ se dice **σ-finita** si existe una sucesión $\left\{A_n\right\}_{n\geq 1}$ de elementos en $\mathcal{A}$ mutuamente ajenos tal que $\mu(A_n) < \infty \ \forall \ n \geq 1$ y $\sum_{n=1}^{\infty} A_n = \Omega$. Toda medida de probabilidad es σ-finita.

Teorema 1. (Criterio útil para probar σ-aditividad de una medida, criterio también conocido como **continuidad en el vacío**).

Sea μ una medida finitamente aditiva sobre un álgebra $\mathcal{A}$. Entonces μ es σ-aditiva en $\mathcal{A}$ si y sólo si μ es continua en el

(*) Si $\mathcal{A}$ es una σ-álgebra no es necesario pedir esta condición.

vacío, es decir si $A_n \downarrow \phi$ y $\mu(A_n) < \infty$(**) para algún n implican que $\lim_{n\infty} \mu(A_n) = 0$.

Proposición 3. (Propiedades elementales de una medida de Probabilidad)

Sea P una medida de probabilidad sobre un álgebra $\mathcal{A}$ de subconjuntos de Ω. Entonces

a) $P(\phi) = 0$ y $P(A^c) = 1 - P(A)$ para $A \in \mathcal{A}$.

b) Si $A,B \in \mathcal{A} \Rightarrow P(A\cup B) = P(A) + P(B) - P(A\cap B)$.

c) Si $A,B \in \mathcal{A}$ y $A\subseteq B \Rightarrow$

$$P(A) \leq P(B).$$

d) Si $A_n \in \mathcal{A}$ $n \geq 1$ y $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A} \Rightarrow$

$$P\left[\bigcup_{n=1}^{\infty} A_n\right] \leq \sum_{n=1}^{\infty} P(A_n)$$

e) Si P es una medida de Probabilidad entonces

$A_n \uparrow A \Rightarrow \lim_{n\infty} P(A_n) = P(A)$ (Continuidad por abajo)

$A_n \downarrow A \Rightarrow \lim_{n\infty} P(A_n) = P(A)$ (Continuidad por arriba)

Demostración de e): Sea $B_1 = A_1, B_2 = A_2\setminus A_1, \ldots, B_n = A_n\setminus A_{n-1}$. Entonces $A = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$ y B_n son mutuamente ajenos. Por lo tanto $P\left[\bigcup_{n=1}^{\infty} A_n\right] = P\left[\bigcup_{n=1}^{\infty} B_n\right] = \sum_{n=1}^{\infty} P(B_n) = \lim_{n\infty} P(B_n) = \lim_{n\infty} P(A_n)$. ∎

4. INDEPENDENCIA

Sea $(\Omega,\mathcal{F},P)$ un espacio de probabilidad. Decimos que dos conjuntos A y B en $\mathcal{F}$ son independientes si

(**) Observese que si $\mu(\Omega)$ es finita entonces esta condición no es necesaria.

$$P(A \cap B) = P(A)P(B).$$

La definición de independencia se puede extender a clases de eventos. Así, decimos que A es **una clase de eventos independientes** si para toda colección finita de eventos $A_1, \ldots A_n$ de A

$$P\left[\bigcap_{i=1}^{\infty} A_i\right] = \prod_{i=1}^{n} P(A_i)$$

Observación: Si $(\Omega, \mathcal{F}, P)$ es un espacio de probabilidad a los elementos de $\mathcal{F}$ es común llamarles **eventos**.

Si A_λ es una clase de eventos para cada λ en algún conjunto de índices Λ, decimos que $\{A_\lambda : \lambda \in \Lambda\}$ es una **familia de clases independientes** si para cada elección de un miembro A_λ de A_λ, la clase $\{A_\lambda : \lambda \in \Lambda\}$ es una clase de eventos independientes.

Las clases de una familia independiente muchas veces se pueden aumentar sin perder independencia, tal como lo indica el seguiente resultado.

Teorema 2. Sea $\{A_\lambda : \in \Lambda\}$ una familia de clases independientes tal que A_λ es cerrada bajo intersecciones finitas. Sea $\mathcal{C}_\lambda = \sigma(A_\lambda)$. Entonces $\{\mathcal{C}_\lambda : \lambda \in \Lambda\}$ es una familia de clases independientes.

Corolario 1. Si $\mathcal{A}$ es una clase de eventos independientes y si alguno de los eventos de $\mathcal{A}$ son reemplazados por sus complementos, entonces la clase resultante es una clase eventos independientes.

Observación. Si A y B son independientes, entonces A^c y B son independientes ya que

$$P(B) = P(A^c \cap B) + P(A \cap B) = P(A^c \cap B) + P(A)P(B) \therefore$$
$$P(A^c \cap B) = P(B) - P(A)P(B) = P(B)(1-P(A)) = P(B)P(A^c).$$

La demostración del Teorema 2 se basa en los siguientes resultados, los cuales son muy útiles en la Teoría de Probabilidad. Decimos que $\mathcal{M}$ es una **clase monótona** si para $A_n \in \mathcal{M}$ y $A_n \uparrow A$ o $A_n \downarrow A$ se tiene que $A \in \mathcal{M}$. Siempre existe una clase monótona mínima que contiene una colección de conjuntos $\mathcal{C}$ dada, denotada por $\mathcal{M}(\mathcal{C})$. La demostración de esta última aseveración es similar a la de la Proposición 2.

Lema 1. (De las clases monótonas)

$$\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$$

si y sólo si $\mathcal{C}$ es un álgebra.

Una colección de conjuntos $\mathcal{D}$ es un **D-sistema** si

i) $\Omega \in \mathcal{D}$

ii) $A,B \in \mathcal{D}$ $A \subseteq B \Rightarrow B\backslash A \in \mathcal{D}$

iii) $A_n \in \mathcal{D}$ $n\geq 1$ $A_n\uparrow \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$.

Siempre existe un mínimo D-sistema que contiene a una clase arbitraria $\mathcal{C}$ denotada por $\mathcal{D}(\mathcal{C})$, y llamado el **mínimo D-sistema que contiene a $\mathcal{C}$.**

Lema 2. Sea $\mathcal{E}$ una clase cerrada bajo intersección. Entonces

$$D(\mathcal{E}) = \sigma(\mathcal{E}).$$

La demostración de los dos Lemas anteriores usa el principio de los "Conjuntos Buenos", el cual se explicará en la demostración de la Proposición 4.

5. CONSTRUCCION DE MEDIDAS Y EJEMPOS DE ESPACIOS DE PROBABILIDAD

La construcción de los ejemplos de espacios de probabilidad que daremos se basa en el siguiente teorema de la Teoría de la Medida

Teorema 3. (Teorema de Carathéodory). Sea Ω un conjunto no vacío, $\mathcal{A}$ un álgebra de subconjuntos de Ω y $\mathcal{F} = \sigma(\mathcal{A})$. Sea μ_0 una medida σ-finita en $(\Omega, \mathcal{A})$. Entonces hay una única medida μ en $(\Omega, \mathcal{F})$ la cual es una extensión de μ_0, es decir

$$\mu(A) = \mu_o(A) \quad \forall \quad A \in \mathcal{A}.$$

La medida μ se llama la **extensión** de μ_0.

Observación: El teorema de Carathéodory requiere que se tenga una medida en el álgebra $\mathcal{A}$ la cual es σ-finita. Si se quita esta condición hay ejemplos "patológicos" de la no-unicidad de μ. Esta condición siempre se cumple en el caso de medidas de probabilidad puesto que $P(\Omega) = 1$.

Ejemplo 1. (Medidas en $\mathbb{R}$).

Sea $\mathbb{R}$ el conjunto de los números reales y $\mathcal{B}(\mathbb{R})$ sus conjuntos de Borel. Recordemos que $\mathcal{B}(\mathbb{R})$ es la σ-álgebra generada por los conjuntos abiertos de $\mathbb{R}$. Definamos

$$(a,b] = \{x \in \mathbb{R} : a < x \leq b\}$$

para todo $-\infty \leq a < b < \infty$ en donde $(a,\infty]$ lo tomamos como (a,∞). Sea $\mathcal{S}$ la colección de intervalos de la forma anterior. Se puede demostrar que $\mathcal{S}$ es una semi-álgebra, llamada la **semi-álgebra** de semi-intervalos de $\mathbb{R}$. Sea $\mathcal{A} = \mathcal{A}(\mathcal{S})$ (véase la Proposición 1) o sea

$$A \in \mathcal{A} \quad \text{si} \quad A = \sum_{i=1}^{n} (a_i, b_i]$$

Observe que $\mathcal{A}$ no es una σ-álgebra pues si $A_n = (0, 1-\frac{1}{n}] \in \mathcal{A}$

$$\bigcup_{n=1}^{\infty} A_n = (0,1) \notin \mathcal{A}.$$

Lema 3. $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$.

Definición 7. Una función $F : \mathbb{R} \quad \mathbb{R}$ se llama función de distribución si satisface las siguientes propiedades:

i) $F(x)$ es no decreciente.

ii) $F(-\infty) = 0$ y $F(+\infty) = 1$, en donde

$$F(-\infty) = \lim_{x \ -\infty} F(x) \quad \text{y} \quad F(+\infty) = \lim_{x \ \infty} F(x).$$

iii) $F(x)$ es continua por la derecha y tiene límite por la izquierda.

Como ejemplos de funciones de distribución tenemos:

(1) $$F(x) = \begin{cases} 0 & x < 0. \\ x & 0 \leq x \leq 1. \\ 1 & x \geq 1. \end{cases}$$ (Uniforme)

(2) $$F(x) = \begin{cases} 0 & x < 0 \\ 1-e^{-\theta x} & x \geq 0 \quad \theta > 0 \end{cases}$$ (Exponencial)

(3) $$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^{2/2}} dt$$ (Normal)

(4) $$F(x) = \begin{cases} \sum_{k=0}^{[x]} \frac{\lambda e^{-\lambda x}}{k!} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad \lambda > 0$$ (Poisson)

en donde $[x]$ denota la parte entera de x.

Teorema 4. Sea F una función de distribución en $\mathbb{R}$. Entonces existe una única medida de probabilidad P_F en $(\mathbb{R}, B(\mathbb{R}))$ tal que

$$P_F((a,b]) = F(b) - F(a)$$

para todo $a, b, -\infty \leq a < b < \infty$.

Demostración. Observe que P_F esta definida en $\mathcal{S}$. Si $A \in \mathcal{A}(\mathcal{S})$ o sea

$$A = \sum_{i=1}^{n} (a_i, b_i],$$

definimos

$$P_F(A) = \sum_{i=1}^{n} P_F(a_i, b_i] = \sum_{i=1}^{n} \{F(b_i)-F(a_i)\} .$$

Claramente P_F es finitamente aditiva en $\mathcal{A}$ y $P_F(\mathbb{R}) = 1$. Para demostrar que P_F es σ-aditiva se usa el Teorema de Continuidad en el Vacío (Teorema 1) así como la propiedad de Heine-Borel: "Un conjunto A en $\mathbb{R}$ es compacto si y sólo si es cerrado y acotado". Así dada una medida P_F en el álgebra $\mathcal{A}, P_F(\Omega) = 1 < \infty$ y la demostración termina aplicando el Teorema de Carathéodory. ∎

En forma enteramente análoga a la demostración del Teorema anterior se puede construir la llamada **medida de Lebesgue** λ en $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ que es la única medida que cumple que para cualquier $-\infty < a < b < \infty$

$$\lambda((a,b]) = b-a.$$

Esta medida es claramente σ-finita en $\mathcal{A}(\mathcal{S})$ tomando $A_n=(n,n+1]$ $n = 0,1,2,\ldots,$ al igual que $A_{-n} = (-n-1,-n]$.

Ejemplo 2. (Medidas en $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$).

Sea $\mathbb{R}^n = \mathbb{R}\times\ldots\times\mathbb{R}$ el producto cartesiano de $\mathbb{R}$ i.e.

$$\mathbb{R}^n = \{(x_1,\ldots,x_n) : x_i \in \mathbb{R} \quad i = 1,\ldots,n\}.$$

Definamos

$$\mathcal{S}^n = \{A_1\times\ldots\times A_n : A_i \in \mathcal{B}(\mathbb{R}) \quad i = 1,\ldots,n\}.$$

Se puede probar que $\mathcal{S}^n$ es una semi-álgebra de subconjuntos de $\mathbb{R}^n$ y que como en el Lema 3 $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{S}^n)$. Sean $\mu_1,\dots,\mu_n$ n medidas (no necesarimente de probabilidad) en $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, las cuales son σ-finitas.

Teorema 5. (Medida Producto).

Existe una única medida μ^n en $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$ tal que

(5) $$\mu^n(A_1 x\dots xA_n)=\mu_1(A_1)\dots\mu_n(A_n)\ \forall\ A_i\in\mathcal{B}(\mathbb{R})\ i=1,\dots,n.$$

Algunas veces en lugar de μ^n se escribe $\mu_1 x\dots x\mu_n$.

Demostración. Para un elemento de $\mathcal{S}^n$ definamos μ^n como en (5). Extendamos μ^n a $\mathcal{A}^n = \mathcal{A}(\mathcal{S}^n)$ en la forma usual, es decir para

$$A \in \mathcal{A}^n \quad A = \sum_{i=1}^{k} A_1^i x\dots xA_n^i \quad A_j^i \in \mathcal{B}(\mathbb{R}) \quad \text{algún } k \geq 0$$

$$\mu^n(A) = \sum_{i=1}^{k} \mu_1(A_j^i)\dots\mu_n(A_n^i).$$

Esta μ^n es finitamente aditiva en $\mathcal{A}^n$. Para probar que es σ-aditiva hay que usar el Teorema 1 de Continuidad en el Vacío y la propiedad de Heine-Borel en $\mathbb{R}^n$. μ^n es σ-finita en $\mathcal{A}^n$ puesto que cada μ_i lo es en $\mathcal{B}(\mathbb{R})$. La demostración termina aplicando el Teorema de Extensión de Carathéodory. ∎

Si $\mu_i = \lambda\ \ \forall\ \ i = 1,\dots,n$ entonces la medida producto μ^n se llama medida de Lebesgue en $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$ y se denota por λ^n.

Ejemplo 3. (Medidas en $\mathbb{R}^\infty$).

Sea $\mathbb{R}^\infty$ el espacio de todas las sucesiones ordenadas

$$x = (x_1,x_2,\dots) \quad -\infty < x_k < \infty \quad k = 1,2,\dots$$

Sea $\mathscr{S}^{\infty}$ la colección de todos los subconjuntos de $\mathbb{R}^{\infty}$ de la forma

$$C_{B^n} = \{(x_1, x_2, \dots) : (x_1, \dots x_n) \in B^n \quad \text{algún } n \geq 1$$
$$B^n \in \mathcal{B}(\mathbb{R}^n)\}.$$

Entonces $\mathscr{S}^{\infty}$ es una semi-álgebra, llamada **semi-álgebra de cilindros de** $\mathbb{R}^{\infty}$.

Teorema 6. (Teorema de Kolmogorov). Sea $\left\{P_n\right\}_{n\geq 1}$ una sucesión de medidas de probabilidad en $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Entonces existe una única medida de probabilidad P^{∞} en $(\mathbb{R}^{\infty}, \sigma(\mathscr{S}^{\infty}))$ tal que

$$P^{\infty}(C_{B^n}) = P^n(B^n) \quad \forall \quad C_{B^n} \in \mathscr{S}^{\infty} \tag{6}$$

en donde $P^n = P_1 x \dots x P_n$.

La demostración es similar en idea a las demostraciones de los Teoremas 5 y 6. Es importante observar que P^{∞} está bien definida y para esto se usa que las P_n son medidas de Probabilidad.

Ejemplo 4. (Medidas en C[0,1]).

Sea C(0,1] el espacio de todas las funciones continuas en el intervalo [0,1]. C[0,1] es un espacio métrico con la métrica

$$d(f,g) = \sup_{0\leq t\leq 1} |f(t) - g(t)| .$$

Considere los subconjuntos de C[0,1] de la forma

$$A(t_1, \dots t_k, I_1, \dots, I_k) =$$
$$\{f \in C[0,1] : (f(t_1), \dots, f(t_k)) \in I_1 x \dots x I_k\}$$

para $t_i \in [0,1]$, $I_i = (a_i, b_i]$ $-\infty \leq a_i < b_i \leq \infty$ $i = 1, \dots, k$ algún $k \geq 0$. Sea $\mathscr{S}_C$ la colección de todos los subconjuntos de esta forma (llamados **cilindros**). $\mathscr{S}_C$ es una semi-álgebra llamada la **semi-álgebra de los cilindros de C[0,1]**. Sea $\mathcal{B}(C)$ la σ-álgebra de

Borel de C[0,1] o sea la σ-álgebra generada por los conjuntos abiertos de C[0,1] con la métrica d. Se demuestra que $\sigma(\mathscr{S}_C) = \mathscr{B}(C)$.

La demostración del siguiente resultado es similar en idea a la del Teorema 1.

Teorema 7. (Medida de Wiener). Existe una única medida de probabilidad μ_W en $(C[0,1],\mathscr{B}(C))$ tal que para todo $t_1,\ldots,t_k\in[0,1]$, $a_i < b_i\in \mathbb{R}$, $i = 1,\ldots,k$, y $k \geq 0$ (tomemos $t_1<t_2<\ldots<t_k$ sin perder generalidad)

$$\mu_W(A(t_1,\ldots,t_k;I_1,\ldots,I_k)) =$$

$$\frac{1}{(\sqrt{2\pi})^k\sqrt{t_1(t_2-t_1)\ldots(t_k-t_{k-1})}}\int_{a_1}^{b_1}\cdots\int_{a_k}^{b_k}\cdot$$

$$\left\{e^{-\frac{1}{2}\left[\frac{x_1^2}{t_1}+\frac{(x_2-x_1)^2}{t_2-t_1}+\ldots+\frac{(x_n-x_{n-1})^2}{t_n-t_{n-1}}\right]}dx_1\ldots dx_n\right\}.$$

La medida de probabilidd μ_W se llama la medida de Wiener.

6. VARIABLES ALEATORIAS

A menos que se indique lo contrario, de ahora en adelante $(\Omega,\mathscr{F},P)$ denotará un espacio de probabilidad arbitrario y $(\mathbb{R},\mathscr{B}(\mathbb{R}))$ la recta real con sus conjuntos de Borel.

Definición 8. Una función real $X : \Omega - \mathbb{R}$ es una función **$\mathscr{F}$-medible** o una variable aleatoria si

$$\{\omega : X(\omega) \in B\} \in \mathscr{F} \quad \forall \quad B \in \mathscr{B}(\mathbb{R}) \tag{7}$$

o de manera equivalente, la imagen inversa

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \quad \in \mathcal{F} \quad \forall \quad B \in \mathcal{B}(\mathbb{R}) .$$

Si $(\Omega,\mathcal{F}) = (\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$, las funciones $\mathcal{B}(\mathbb{R}^n)$-medibles se llaman **funciones de Borel**.

Al igual que en los cursos de probabilidad elemental, una variable aleatoria la interpretamos como una propiedad numérica de un experimento, con un valor que "depende del azar". La condición (7) es importante puesto que si P es una medida de probabilidad en $(\Omega,\mathcal{F})$, sólo tiene sentido hablar de la probabilidad del conjunto $\{\omega : X(\omega) \in B\}$ cuando este último es un evento. Por conveniencia el evento $\{X(\omega) \in B\}$ será escrito simplemente como $\{X \in B\}$, pero es importante hacer notar que este último es un elemento de la σ-álgebra $\mathcal{F}$.

Definición 9. Sea X una variable aleatoria sobre un espacio de probabilidad $(\Omega,\mathcal{F},P)$. La medida P_X en $(\mathbb{R},\mathcal{B}(\mathbb{R})$ definida por (algunas veces P_X se escribe como PX^{-1})

$$P_X(B) = P(\{\omega \in \Omega : X(\omega) \in B\}) \qquad B \in \mathcal{B}(\mathbb{R})$$

se llama la **distribución de X**. La función $F_X : \mathbb{R} \quad \mathbb{R}$ definida por

$$F_X(x) = P_X((-\infty,x]) = P(\{\omega \in \Omega : X(\omega) \in (-\infty,x]\})$$

se llama la **función de distribucion de X** y cumple con las propiedades de la Definición 7. Obsérvese que por la unicidad del Teorema 4 $P_X = P_F$.

Los siguientes resultados son criterios útiles para decidir si una función es variable aleatoria.

Proposición 4. Sea $\mathcal{E}$ una colección de subconjuntos de $\mathbb{R}$ tal que $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$. Una condición necesaria y suficiente para que una función real $X = X(\omega)$ sea $\mathcal{F}$-medible (variable aleatoria) es que

(8) $$X^{-1}(E) = \{\omega \in \Omega : X(w) \in E\} \in \mathcal{F} \quad \forall \quad E \in \mathcal{E}$$

Demostración. La necesidad de la condición es clara. Para probar la condición suficiente usaremos el "Principio de los Conjuntos Buenos". Sea $\mathcal{D}$ la colección de conjuntos para los cuales la condición (8) es verdad, o sea

$$\mathcal{D} = \{E \subseteq \mathcal{B}(\mathbb{R}) : X^{-1}(E) \in \mathcal{F}\}.$$

Usaremos las siguientes propiedades de la operación "imagen inversa", las cuales son fáciles de probar:

$$X^{-1}(\bigcup_{\alpha} B_\alpha) = \bigcup_{\alpha} X^{-1}(B_\alpha)$$

$$X^{-1}(\bigcap_{\alpha} B_\alpha) = \bigcap_{\alpha} X^{-1}(B_\alpha)$$

$$(X^{-1}(B))^c = X^{-1}(B^c) .$$

Usando estas propiedades se prueba que $\mathcal{D}$ es una σ-álgebra. Por lo tanto

$$\mathcal{E} \subseteq \mathcal{D} \subseteq \mathcal{B}(\mathbb{R})$$

y por consiguiente

$$\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{D}) = \mathcal{D} \subseteq \mathcal{B}(\mathbb{R}).$$

Pero por hipótesis $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$, por lo tanto $\mathcal{D} = \mathcal{B}(\mathbb{R})$.

Esta técnica de los "Conjuntos Buenos" es bastante usada en Teoría de la Medida. En la demostración anterior los "conjuntos buenos" son los elementos de la colección $\mathcal{D}$.

Corolario 2. Una condición necesaria y suficiente para que una función real $X = X(\omega)$ sea una variable aleatoria es que

$$\{\omega : X(\omega) < x\} \in \mathcal{F} \quad \forall\, x \in \mathbb{R}$$

o que

$$\{\omega : X(\omega) \leq x\} \in \mathcal{F} \quad \forall\, x \in \mathbb{R}$$

La demostración es inmediata aplicando la Proposición anterior y el hecho que si $\mathcal{S}$ es la semi-álgebra de semi-intervalos de $\mathbb{R}$, entonces $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$. (Lema 3).

Corolario 3. Sea $(\Omega,\mathcal{F}) = (\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$. Si una función $f : \mathbb{R}^n \; \mathbb{R}$ es continua, entonces f es $\mathcal{B}(\mathbb{R}^n)$)-medible.

La demostración es trivial puesto que si una función $f\colon \mathbb{R}^n \; \mathbb{R}$ es continua, la imagen inversa f^{-1} manda abiertos de $\mathbb{R}$ en abiertos de $\mathbb{R}^n$ y los conjuntos abiertos generan las σ-álgebras de Borel respectivas.

Corolario 4. Suponga que la función $f : \mathbb{R} \quad \mathbb{R}$ es monótona. Enfonces f es Borel-medible.

Demostración. Supongamos que f es no-decreciente. entonces para cualquier número real α

$$\{y : f(y) \le \alpha\} = \begin{cases} \phi & \\ \Omega & \\ (-\infty,a) & \text{para algún } a \in \mathbb{R} \\ (-\infty,a] & \text{para algún } a \in \mathbb{R} \end{cases}$$

y por lo tanto f es Borel-medible. ■

Proposición 5. Sea $\varphi = \varphi(x)$ una función de Borel y $X = X(\omega)$ una variable aleatoria. Entonces la composición $Y = \varphi\circ X$ (o sea $Y(\omega) = \varphi(X(\omega))$ es también una variable aleatoria.

Demostración. Sea $B \in \mathcal{B}(\mathbb{R})$, entonces $\varphi^{-1}(B) \in \mathcal{B}(\mathbb{R})$ puesto que φ es una función Borel-medible. La demostración se completa verificando que

$$\{\omega : Y(\omega)\in B\} = \{\omega : \varphi(X(\omega)\in B\} = \{\omega : X(\omega)\in\varphi^{-1}(B)\} \in \mathcal{F}$$ ■

Corolario 5. Sea X una variable aleatoria. Entonces X^P, $X^+ = \max(X,0)$, $X^- = -\min(X,0)$ y $|X|$ son también variables aleatorias.

Demostración. Como las funciones reales x^P, x^+, x^- y $|x|$ son continuas, por el Corolario 3 y la Proposición 5 se obtiene el resultado. ∎

Comentario. Se puede probar que si X y Y son variables aleatorias, entonces X+Y, XY, max(X,Y) y min(X,Y) son también variables aleatorias. Igualmente si $\{X_n\}$ es una sucesión de variables aleatorias entonces $\sup X_n$, $\inf X_n$, $\limsup X_n$ y $\liminf X_n$ son también variables aleatorias (las cuales pueden ser $+\infty$ ó $-\infty$).

El siguiente es un resultado importante en Probabilidad y Estadística pues da la existencia de una variable aleatoria con una distribución dada.

Teorema 8. Sea F una función de distribución arbitraria. Entonces existe un espacio de probabilidad $(\Omega,\mathcal{F},P)$ y una variable aleatoria X definida en $(\Omega,\mathcal{F},P)$ tal que X tiene función de distribución F.

Demostración. Sea $(\Omega,\mathcal{F},P)$ el espacio de Probabilidad asociado a F dado por el Teorema 4, o sea

$$(\Omega,\mathcal{F},P) = (\mathbb{R},\mathcal{B}(\mathbb{R}),\ P_F).$$

Definamos $X : \mathbb{R} \to \mathbb{R}$ como

$$X(x) = x \qquad x \in \mathbb{R}$$

Entonces por el Corolario 2, X es variable aleatoria y obviamente

$$P_X(B) = P_F(\{x : X(x) \in B\}) \quad \forall \ B \in \mathcal{B}(\mathbb{R})$$

o sea X tiene distribución F. ∎

Sea X una variable aleatoria definida en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y definamos

$$\sigma(X) = \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\} \in \mathcal{F} .$$

Entonces $\sigma(X)$ es una σ-álgebra que es la mínima σ-álgebra contenida en $\mathcal{F}$ con respecto a la cual X es medible, es decir X es $\sigma(X)$-medible.

Decimos que dos variables aleatorias X y Y **son independientes** si $\sigma(X)$ y $\sigma(Y)$ son independientes. En general si $\left\{X_n\right\}_{n\geq 1}$ es una sucesión de variables aleatorias, decimos que son independientes si $\{\sigma(X_n) : n \geq 1\}$ es una familia de σ-álgebras independientes (véase el apartado 4).

El siguiente resultado es de importancia en Probabilidad y Estadística pues asegura la existencia de una sucesión de variables aleatorias independientes con distribuciones dadas.

Teorema 9. (Kolmogorov). Dada una sucesión $\left\{F_n\right\}_{n\geq 1}$ de funciones de distribución en $\mathbb{R}$, existe un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ en donde se encuentra definida una sucesión $\left\{X_n\right\}_{n\geq 1}$ de variables aleatorias independientes y tal que F_n es la función de distribución de X_n para $n = 1,2,\ldots$

Demostración. Sea $P_i = P_{F_i}$ $i = 1,2,\ldots$ en donde P_{F_i} es la medida en $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dada por el Teorema 4. Sea $(\mathbb{R}^{\infty}, \sigma(\mathcal{G}^{\infty}))$ el espacio de probabilidad dado por el Teorema 6. Definamos para $n = 1,2,\ldots$ $X_n : \mathbb{R}^{\infty} \to \mathbb{R}$ como $X_n((x_1, x_2, \ldots)) = x_n$. Por el Corolario 2, X_n es variable aleatoria ya que $\{(x_1, x_2, \ldots) : x_n \leq x\}$ es un cilindro. Por el Teorema 6

$$P^{\infty}(X_n \le x) = P_{F_n}((-\infty, x]) = F_n(x)$$

o sea X_n tiene distribución F_n. Para probar independencia de la sucesión $\{X_n\}$ aplicamos el siguiente resultado cuya demostración usa el Lema 2.

Teorema 10. Una condición necesaria y suficiente para que las variables aleatorias $Y_1, \dots, Y_n$ sean independientes es que $\forall\ (x_1, \dots, x_n) \in \mathbb{R}^n$

$$P(Y_1 \le x_1, \dots, Y_n \le x_n) = P(Y_1 \le x_1) \dots P(Y_n \le x_n).$$

La demostración de que dadas n funciones de distribución $F_1, \dots, F_n$ existe un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ en donde se pueden definir variables aleatorias $X_1, \dots, X_n$ tales que sean independientes y X_i tenga distribución F_i, es análoga (de hecho más sencilla) a la del Teorema 9, usando también el Teorema 10.

Un **proceso estocástico** $\left\{X_t\right\}_{t \ge 0}$ es una colección de variables aleatorias en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, es decir, para cada t $X_t = \Omega \quad \mathbb{R}$ es una variable aleatoria. Entre los procesos estocásticos uno de los más importantes es el siguiente:

Un proceso estocástico $\{W_t \ 0 \le t \le 1\}$ en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ se llama **movimiento Browniano** si

a) $W_0 = 0$

b) Existe un conjunto $A \subset \Omega$ tal que $P(A) = 1$ y para cada $\omega \in \Omega$ la función $\{W_t(\omega) : 0 \le t \le 1\}$ es continua.

c) Si $0 \le t_1 \le t_2 \le \dots \le t_k \le 1 \quad k \ge 1$, las variables aleatorias $W_{t_k} - W_{t_{k-1}}, \dots, W_{t_2} - W_{t_1}, W_{t_1}$ son independientes.

d) Para cada $0 \leq s \leq t$ $W_t - W_s$ tiene una distribución normal con media cero y varianza t-s.

Teorema 10b. Existe un movimiento Browniano.

Demostración. Sea $(\Omega, \mathcal{F}, P) = (C[0,1], \mathcal{B}(C), \mu_W)$ el espacio de Wiener del Teorema 7. Para $\omega \in \Omega$ (ω es una función continua en el intervalo $[0,1]$), defina $W_t(\omega) = \omega(t)-\omega(0)$. Se puede probar que W_t satisface las condiciones (a), (b), (c) y (d) del movimiento Browniano. ∎

7 TEORIA DE INTEGRACION DE LEBESGUE

La integral de Riemann no es suficiente en el estudio de la Teoría moderna de la Probabilidad y por lo tanto hay que desarrollar otro tipo de integral sobre espacios abstractos, conocida como integral de Lebesgue. Un resultado básico de esta teoría es el siguiente. En toda esta sección supondremos que $(\Omega, \mathcal{F}, \mu)$ es un espacio de medida, es decir μ es una medida que no necesariamente es de probabilidad.

Teorema 11. Sea X una función medible no-negativa. Entonces existe una sucesión no decreciente $\{X_n\}_{n\geq 1}$ de funciones simples no negativas tal que $X_n \uparrow X$ o sea $X_n(\omega) \uparrow X(\omega)$ $\forall\ \omega \in \Omega$.

Demostración. Defina para $n \geq 1$

$$X_n(\omega) = \sum_{k=0}^{n^2} \frac{k}{n} 1_{A_k^n}(\omega)$$

en donde $A_k^n = \left\{\omega : \frac{k}{n} \leq X(\omega) < \frac{k+1}{n}\right\}$ y 1_A es la función indicadora de un conjunto A (o sea $1_A(\omega) = 1$ si $\omega \in A$ e igual a 0 si $\omega \notin A$).

Obsérvese que puesto que X es medible entonces $A_k \in \mathcal{F}\ \forall\ k \geq 1$. Probar que $X_n \uparrow X$ es fácil pero tedioso. ∎

Si X es una función simple medible, es decir para algún n

$$X(\omega) = \sum_{k=1}^{n} \sigma_k 1_{A_k}(\omega)$$

para $A_k \in \mathcal{F}$, $\sigma_k \in \mathbb{R}$ $k = 1,\dots,n$, definimos la integral de **Lebesgue de X con respecto a** μ como

$$\int_\Omega f d\mu = \sum_{k=1}^{n} \sigma_k \mu(A_k).$$

Es fácil ver que $\int_\Omega f d\mu$ está bien definida y que si $X \geq Y$ y X y Y son funciones simples medibles, entonces $\int_\Omega x d\mu \geq \int_\Omega Y d\mu$.

Si X es una función medible no negativa definimos $\int_\Omega X d\mu$ como $\lim_{n\ \infty} \int_\Omega X_n d\mu$ en donde $\left\{X_n\right\}_{n\geq 1}$ es la sucesión de funciones dadas por el Teorema 11, o sea $0 \leq X_n \uparrow X$. Obsérvese que el límite de $\int_\Omega X_n d\mu$ siempre existe puesto que se trata de una sucesión monótona de números reales, pero el límite podría ser infinito. El valor de este límite es el mismo para todas las sucesiones de funciones simples no negativas que convergen a X.

Definición 10. Para cualquier función $\mathcal{F}$-medible decimos que existe su **Integral de Lebesgue** con respecto a la medida μ si $\int_\Omega X^+ d\mu < \infty$ ó $\int_\Omega X^- d\mu < \infty$ en cuyo caso definimos

$$\int_\Omega X d\mu = \int_\Omega X^+ d\mu - \int_\Omega X^- d\mu.$$

Algunas veces se usará la notación $\int_\Omega X(\omega)\mu(d\omega)$.

Si $A \in \mathcal{F}$ definimos

$$\int_A A d\mu = \int_\Omega X 1_A d\mu$$

Las siguientes son propiedades elementales de la Integral de Lebesgue, cuya demostración se sigue a partir de la definición, primero probando para funciones simples medibles, luego para medibles no negativas y finalmente para funciones medibles arbitrarias.

Proposición 6. Si la integral de Lebesgue de las funciones medibles X y Y existen, entonces:

1) Si $X \geq Y \Rightarrow \int_\Omega X d\mu \geq \int_\Omega Y d\mu$.

2) Para $a, b \in \mathbb{R}$

$$\int_\Omega (aX+bY) d\mu = a\int_\Omega X d\mu + b\int_\Omega Y d\mu$$

3) Si X es no-negativa y $\int_\Omega X d\mu < \infty$ entonces para $A \in \mathcal{F}$

$$\nu(A) = \int_A X d\mu = \int_\Omega X \cdot 1_A d\mu$$

define una medida sobre $(\Omega, \mathcal{F})$. (La demostración de esto requiere del Terema 14 que aparece más adelante).

4) Si $A, B \in \mathcal{F}$ y $A \cap B = \phi \Rightarrow$

$$\int_{A \cup B} X d\mu = \int_A X d\mu + \int_B X d\mu$$

5) Si $\mu(A) = 0$ entonces $\int_A X d\mu = 0$.

Los siguientes tres resultados son básicos en la Teoría de Integración de Lebesgue y nos dicen bajo qué condiciones se permite intercambiar un límite con una integral.

Teorema 12. (Teorema de Convergencia Monótona). Sea $\left\{X_n\right\}_{n\geq 1}$ una sucesión monótona de funciones medibles no-negativas. Entonces

$$\lim_{n\to\infty} \int_\Omega X_n d\mu = \int_{\Omega} \lim_{n \; \infty} X_n d\mu$$

Teorema 13. (Lema de Fatou). Sea $\left\{X_n\right\}_{n\geq 0}$ una sucesión de funciones medibles no-negativas, entonces

$$\int_\Omega \underline{\lim}\, X_n d\mu \leq \underline{\lim} \int_\Omega X_n d\mu.$$

Teorema 14. (Teorema de Convergencia Dominada). Sea $\left\{X_n\right\}_{n\geq 1}$ una sucesión de funciones medibles tal que para alguna función medible X $\mu(X_n(\omega) \quad X(\omega)) = 0$. Suponga que existe una función medible Y tal que $|X_n| \leq Y$ y $\int_\Omega Y d\mu < \infty$. Entonces

$$\lim \int_\Omega X_n d\mu = \int_\Omega X d\mu .$$

El Ejemplo 2 de estas notas se puede generalizar de la siguiente manera (lo haremos para $n = 2$). Dados dos espacios de medida arbitrarios $(\Omega_1, \mathcal{F}_1, \mu_1)$ y $(\Omega_2, \mathcal{F}_2, \mu_2)$ definamos la **semi-álgebra de rectángulos** de $\Omega_1 \times \Omega_2$ como

$$\mathcal{S}^2 = \{A_1 \times A_2 : A_1 \in \mathcal{F}_1 \text{ y } A_2 \in \mathcal{F}_2\}$$

En esta semi-álgebra definimos (si μ_1 y μ_2 son σ-finitas)

$$\mu_1 \times \mu_2 (A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

y la extendemos en la forma usual a $\sigma(\mathcal{S}^2)$. La σ-álgebra $\sigma(\mathcal{S}^2)$ se llama σ-álgebra producto y se escribe $\mathcal{F}_1 \times \mathcal{F}_2$. Entonces tenemos el espacio de medida producto $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mu_1 \times \mu_2)$. Para este espacio se tiene el siguiente resultado.

Teorema 15. (Teorema de Fubini).

Sea $X : \Omega_1 \times \Omega_2 \quad \mathbb{R}$ una función $\mathcal{F}_1 \times \mathcal{F}_2$-medible y supongamos que

$$\int_{\Omega_1 \times \Omega_2} |X| d(\mu_1 \times \mu_2) < \infty.$$

Entonces

$$\int_{\Omega_1} X(\omega_1, \omega_2)\mu_1(d\omega_1) \quad \text{y} \quad \int_{\Omega_2} X(\omega_1, \omega_2)\mu_2(d\omega_2)$$

son funciones $\mathcal{F}_2$-medible y $\mathcal{F}_1$-medible respectivamente que son integrables con respecto a μ_2 y μ_1 respectivamente. Además

$$(9) \qquad \int_{\Omega_1 \times \Omega_2} X(\omega_1, \omega_2) d(\mu_1 \times \mu_2) = \int_{\Omega_1} \left[\int_{\Omega_2} X d\mu_2 \right] d\mu_1 = \int_{\Omega_2} \left[\int_{\Omega_1} X d\mu_1 \right] d\mu_2$$

Comentario. Si X es una variable no-negativa, el Teorema de Fubini siempre es válido a pesar de que $\int_{\Omega_1 \times \Omega_2} X d(\mu_1, \mu_2)$ puede ser infinito.

Teorema 15.b. (Cambio de Variable). Sean $(\Omega, \mathcal{F})$ y $(E, \mathcal{E})$ dos espacios medibles y $X = X(\omega)$ una transformación $\mathcal{F}/\mathcal{E}$-medible con valores en E. Sea P una probabilidad en $(\Omega, \mathcal{F})$ y P_X la probabilidad en $(E, \mathcal{E})$ inducida por X, o sea

$$P_X(A) = P(\{\omega : X(\omega) \in A\}) \qquad A \in \mathcal{E}$$

Entonces

$$\int_A g(x) P_X(dx) = \int_{X^{-1}(A)} g(X(\omega)) P(d\omega) \qquad A \in \mathcal{E}$$

para toda función $g = g(x)$ $\mathcal{E}$-medible (en el sentido que si una integral existe, la otra está bien definida, y las dos son iguales).

Teorema 15c. Sea $\bar{\lambda}$ la medida de Lebesgue completa[(***)]. Si una función es integrable en el sentido de Riemann, entonces es $\bar{\lambda}$-integrable y las integrales coinciden.

8. VALOR ESPERADO Y MOMENTOS

Sea $(\Omega,\mathcal{F},P)$ un espacio de probabilidad arbitrario y X una variable aleatoria. Si $\int_\Omega |X|dP < \infty$ definimos el **valor esperado de** X, que se escribe E(X), como

$$E(X) = \int_\Omega XdP.$$

En general si $r\geq 0$ y $\int_\Omega |X|^r dP < \infty$ definimos el **r-ésimo momento de** X como

$$EX^r = \int_\Omega X^r dP.$$

Sea P_X la distribución de X y F su función de distribución (vea la Definición 9). Entonces usando el Teorema 15.b se demuestra que

$$\int_\Omega XdP = \int_{\mathbb{R}} x\, dP_F(x) \tag{10}$$

y

$$\int_\Omega X^r dP = \int_{\mathbb{R}} x^r dP_F(x).$$

En general si $g : \mathbb{R} \quad \mathbb{R}$ es una función medible entonces

$$E(g(X)) = \int_{\mathbb{R}} g(x)dP_F(x). \tag{11}$$

Veremos ahora cómo se pueden calcular las expresiones (10) y (11) en algunos casos particulares. Para esto necesitamos

(***) Una medida en $(\Omega,\mathcal{F})$ se llama completa si $\forall \quad A \in \mathcal{F}$ con $\mu(A)=0$ se tiene que $B \in \mathcal{F} \;\; \forall \;\; B \subset A$.

introducir los siguientes conceptos: Sea F una función de distribución y definamos

$$D_F = \{x \in \mathbb{R} : F(x)-F(x^-) > 0\}$$

en donde $F(x^-)$ es el límite por la izquierda de F en x (el cual siempre existe pues F es función de distribución). El conjunto D_F es a lo más numerable (véase la demostración más adelante). Si $D_F = \phi$ decimos que la **función de distribución F es continua.** Decimos que F es una **función de distribución discreta** si $P_F(D_F)=1$ o sea el conjunto de puntos de discontinuidad de F tiene probabilidad uno. En general usaremos la notación

$$P_x = F(x)-F(x^-) = P(X=x) = P_F(\{x\})$$

lo cual implica que si x es un punto de continuidad de F entonces $P_x = 0$. Si F es una distribución discreta la expresión (10) se convierte en

$$E(X) = \int_\Omega XdP = \int_{\mathbb{R}} x dP_F(x) = \int_{D_F} x dP_F(x) = \sum_{x\in D_F} xP_x = \sum_{x\in D_F} xP(X=x) \tag{12}$$

en donde la suma es sobre un número a lo más numerable de puntos. Si $P_x > 0$ decimos que x es un **átomo** de F.

Decimos que una función de distribución continua F es **absolutamente continua** si existe una función no-negativa f, llamada la **densidad** de F, tal que para todo $x \in \mathbb{R}$

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Para una función de distribución absolutamente continua la expresión (10) se convierte en

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

Para la expresión (11) se tiene que si F es discreta

$$E(g(X)) = \sum_{x \in D_F} g(x)P(X=x)$$

y si F es absolutamente continua con densidad f

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx .$$

Comentario sobre la Expresión (12):

Explicación de por qué $P_F(\{x\}) = F(x)-F(x^-)$

Observe que $\forall\ x \in \mathbb{R}$

$$\{x\} = \bigcap_{n=1}^{\infty} (x-\tfrac{1}{n}, x] \text{ y los conjuntos } A_n = (x-\tfrac{1}{n}, x] \quad n \geq 1$$

decrecen. Entonces por el Teorema de continuidad de la probabilidad $P_F(\{x\}) = \lim P_F(A_n) = \lim (P_F((x-\frac{1}{n}, x])) = \lim (F(x)-F(x-\frac{1}{n})) = F(x) - F(x^-)$.

Demostración de que el conjunto D_F es a lo más numerable:

D_F se llama el **conjunto de átomos de F**.

Para cada entero n = 1,2,... sea

$$D_n = \{x : F(x)-F(x^-) > 1/n\}.$$

Entonces D_n no tiene más que n puntos porque si los tuviera $P_F(D_n) > 1$. Pero

$$D_F = \bigcup_{n=1}^{\infty} D_n$$

el cual es a lo más numerable ya que cada D_n es finito ∎

El siguiente resultado es un resumen de las propiedades más importantes del valor esperado, aparte de la Proposición 6(2) y los Teoremas 12, 13 y 14.

Proposición 7. Suponga que X y Y son variables aleatorias en un espacio de probabilidad $(\Omega,\mathcal{F},P)$

1) Si $E|X|^p<\infty$, $E|Y|^q<\infty$ para $p > 1$, $q > 1$ $1/p + 1/q = 1$ entonces $E|XY| < \infty$ y

$$|E(XY)| \le E|XY| \le (E|X|^p)^{1/p} (E|Y|^q)^{1/q} \quad \text{(Desigualdad de Holder)}$$

2) Si $E|X|^p<\infty$, $E|Y|^p\le\infty$ para algún $p \ge 1$ entonces $E|X+Y|^p<\infty$ y

$$E|XY|^p)^{1/p} \le (E|X|^p)^{1/p} + (E|Y|^p)^{1/p} \quad \text{(Desigualdad de Minkowski)}$$

3) Si $0 < p < 1$ y $E|X|^p<\infty$, $E|Y|^p<\infty$, entonces $E|X+Y|^p<\infty$ y

$$(E|X+Y|^p)^{1/p} \le (E|X|^p)^{1/p} +(E|Y|^p)^{1/p}$$

4) (Desigualdad de Markov) Si $0 < p < \infty$

$$P(|X| \ge a) \le E|X|^p/a^p \quad a > 0$$

5) (Desigualdad de Chebyshev) Si $E|X|^2<\infty$ entonces

$$P(|X-E(X)| \ge a) \le \frac{VAR(X)}{a^2} \quad \forall\, a > 0$$

en donde $VAR(X) = EX^2-(EX)^2$.

6) (Desigualdad de Jensen). Si $E|X| < \infty$ y g es una función convexa en $\mathbb{R}$ tal que $E|g(X)|<\infty$, entonces

$$g(E(X)) \le E(g(X)).$$

7) Si X y Y son variables aleatorias independientes con $E|X|<\infty$ y $E|Y|<\infty$ entonces $E|XY|<\infty$ y

$$E(XY) = (EX)(EY).$$

9. LA FUNCION CARACTERISTICA (TRANSFORMADA DE FOURIER)

Sea $(\Omega,\mathcal{F},P)$ un espacio de probabilidad y $\mathbb{C}$ el conjunto de los números complejos. Una función $Z : \Omega\ \mathbb{C}$ $(Z(\omega) = (X(\omega)+iY(\omega))$ se llama **variable aleatoria compleja** si X y Y son variables aleatorias reales. Si $E|X|<\infty$ y $E|Y|<\infty$ definimos

$$E(Z) = E(X) + iE(Y).$$

Sea X una variable aleatoria real con función de distribución F. Definimos la función característica de F como

$$\hat{\mu}(t) = E(e^{itX}) = \int_{\Omega} e^{itX} dP = \int_{\mathbb{R}} e^{itx} dP_F(x) \quad t \in \mathbb{R}$$

Observe que puesto que $|e^{itx}| \leq 1 \quad \forall \quad t \in \mathbb{R},\ x \in \mathbb{R}$ entonces $\hat{\mu}(t)$ siempre está definida para todo $t \in \mathbb{R}$. $\hat{\mu}(t)$ se conoce también como la **Transformada de Fourier** de la medida de probabilidad P_F, o de la distribución F. También se dice que la variable aleatoria X tiene función característica $\hat{\mu}$ y algunas veces se escribe $\hat{\mu}_X$ ó $\hat{\mu}_F$.

Teorema 16 (Unicidad). Sean F y G dos funciones de distribución con la misma función característica. Entonces $F \equiv G$, o sea $F(x) = G(x) \quad \forall \quad x \in \mathbb{R}$.

El siguiente resultado es de bastante utilidad en el estudio de sumas de variables aleatorias independientes.

Proposición 8. Sean $X_1, \ldots X_n$ variables aleatorias independientes y $S_n = X_1 + \ldots + X_n$. Entonces

$$\hat{\mu}_{S_n}(t) = \hat{\mu}_{X_1}(t) \ldots \hat{\mu}_{X_n}(t) .$$

Proposición 9. Sea F una función de distribución y $\hat{\mu}$ su función característica, o sea

$$\hat{\mu}(t) = \int_{\mathbb{R}} e^{itx} dP_F(x) \qquad t \in \mathbb{R}. \tag{13}$$

Entonces la función $\hat{\mu} : \mathbb{R} \quad \mathbb{C}$ tiene las siguientes propiedades.

i) $\hat{\mu}(0) = 1$

ii) $\hat{\mu}$ es positiva definida, o sea $\forall \; t_1,\dots t_k \in \mathbb{R}$, $a_1,\dots,a_k \in \mathbb{C}$ y $k > 0$

$$\sum_{i,j} a_i \bar{a}_j \hat{\mu}(t_i - t_j) \geq 0$$

iii) $\hat{\mu}$ es continua.

El siguiente resultado es el converso de la Proposición 9 y nos permite construir medidas de probabilidad en $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ de una manera diferente a como se hizo en los ejemplos de la Sección 4, en los que se usó el Teorema de Caratheodory.

Teorema 17 (Teorema de Bochner). Sea $\hat{\mu} : \mathbb{R} \quad \mathbb{C}$ una función que satisface las propiedades (i), (ii) y (iii) de la Proposición 9. Entonces existe una única medida de Probabilidad μ en $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ tal que $\hat{\mu}$ es su función característica, o sea

$$\hat{\mu}(t) = \int_{\mathbb{R}^x} e^{itx} \, d\mu(x)$$

La función característica tiene muchas otras propiedades que no mencionaremos aquí. Sólo diremos que si $E|X|^k<\infty$ entonces

$$EX^k = \frac{\hat{\mu}^{(k)}(0)}{k!}$$

en donde $\mu^{(k)}(t)$ es la k-ésima derivada de $\hat{\mu}(t)$

Teorema 18. Una condición necesaria y suficiente para que las variables aleatorias $(X_1,\dots X_n)$ sean independientes es que

$$E\left[e^{i(t_1X_1+\dots+t_nX_n)}\right] = \prod_{j=1}^{n} E\left[e^{itx_j}\right] \quad \forall \quad (t_1,\dots t_n)\in\mathbb{R}^n.$$

Las funciones características de las distribuciones más usadas en probabilidad y estadística son las siguientes:

Bernouilli: (parámetro p) (B(P))

$$\hat{\mu}(t) = pe^{it}+(1-p)$$

Normal: $(N(\mu,\sigma^2))$

$$\hat{\mu}(t) = e^{it\mu-\frac{\sigma^2}{2}t^2}$$

Poisson: $\hat{\mu}(t) = e^{\lambda(e^{it}-1)}$ $(P(\lambda))$

Cauchy: $\hat{\mu}(t) = e^{-|t|}$ (CA)

Estable: $\hat{\mu}(t) = e^{-|t|^{\alpha}}$ $0 < \alpha \leq 2$ $(ES(\alpha))$

Las secciones 10, 11, 12 y 13 contienen resultados de Probabilidad que son bastante útiles en el estudio de distribuciones límites en Estadística.

10. MODOS DE CONVERGENCIA

Sean $X, X_1, X_2, \ldots$ variables aleatorias con funciones de distribución $F, F_1, F_2, \ldots$ y funciones características $\phi, \phi_1, \phi_2, \ldots$ respectivamente. En esta sección estudiaremos cuatro diferentes tipos de convergencia de X_n a X, a saber:

a) Convergencia con Probabilidad 1 o casi segura (c.s. ó c.p.1)

b) Convergencia en Probabilidad (p)

c) Convergencia en media r (r)

d) Convergencia Débil o en Distribución (D) las cuales denotaremos respectivamente por

$$X_n \xrightarrow{c.s.} X$$
$$X_n \xrightarrow{p} X$$
$$X_n \xrightarrow{r} X$$
$$X_n \xrightarrow{\mathcal{D}} X$$

Dada una función de distribución F, denotamos por C_F el conjunto de puntos de continuidad en F.

Definición 11. Sea $\{X_n\}_{n\geq 1}$ y X variables aleatorias definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Decimos que X_n converge **casi seguramente** a X si

$$P(\lim_{n \to \infty} X_n \text{ exista y es } X) = 1$$

también llamada **convergencia con probabilidad 1.**

Decimos que X_n **converge en probabilidad** a X si $\forall \ \epsilon > 0$

$$P(|X_n - X| > \varepsilon) \to 0 \text{ cuando } n \to \infty$$

y decimos que X_n **converge en media r** a X si

$$E|X_n - X|^r \to 0 \text{ cuando } n \to \infty.$$

Definición 12. Sea $\{F_n\}$ una sucesión de funciones de distribución. Si existe una función de distribución F tal que

$$F_n(x) \to F(x) \quad n \to \infty$$

para todo x punto de continuidad de F, decimos que F_n **converge en distribución o débilmente a F.** Decimos que una sucesión de variables aleatorias X_n converge en distribución a X si las correspondientes funciones de distribución F_n convergen débilmente a F la función de distribución de X.

Observación. Se puede tener una sucesión de funciones de distribución F_n que no converge a una función de distribución: Sea

$$F_n(x) = \begin{cases} 0 & \text{si } x < n \\ 1 & \text{si } x \geq n \end{cases}$$

entonces $F_n(x) \quad F(x) = 0 \quad \forall \quad x \in \mathbb{R}$ y F no es una función de distribución.

Ejemplo 5. Sea $X_n \sim N(0,\frac{1}{n})$, $X = 0$ y Φ la función de distribución normal estándar. Entonces

$$F_n(x) = \Phi(\sqrt{n}x)$$

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } X \geq 0 \end{cases} \qquad C_F = \mathbb{R}\backslash\{0\}$$

$$\lim_{n \quad \infty} F_n(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/2 & \text{si } x = 0 \\ 1 & \text{si } x > 0 \end{cases}$$

es decir $F_n(x) \quad F(x) \quad \forall \quad x \neq 0$ o sea $\forall \quad x \in C_F$, por lo tanto $F_n \overset{\mathcal{D}}{-} F$.

Obsérvese también que $\forall \in > 0$

$$P(|X_n-X| > \varepsilon) = P(|X_n| > \varepsilon) = 1-F_n(-\varepsilon)+F_n(\varepsilon)$$

$$= 2\Phi(-\sqrt{n}\varepsilon) \underset{n \quad \infty}{\quad} 0$$

o sea $X_n \xrightarrow[n \quad \infty]{p} 0$ (en probabilidad).

Ejemplo 6. Este ejemplo muestra qué variables aleatorias discretas pueden converger a una continua.

Sea $X \sim \cup(0,1)$ y X_n uniformemente distribuida en $j = 1,\ldots,2^n$, o sea

$$P(X_n=j) = \frac{1}{2^n} \qquad j = 1,2,\ldots,2^n .$$

Por lo tanto

$$F_n(x) = \begin{cases} 0 & x < 0 \\ \dfrac{j-1}{2^n} & \dfrac{j-1}{2^n} \leq x < \dfrac{j}{2^n} \\ 1 & x \geq 1 \end{cases}$$

Así, si x es tal que $\frac{j-1}{2^n} \le x < \frac{j}{2^n}$

$$|F(x)-F_n(x)| = \frac{j-1}{2^n} - x \le \frac{1}{2^n} \quad 0 \quad n \quad \infty$$

y por lo tanto $\lim\limits_{n \quad \infty} F_n(x) = F(x) \quad \forall \quad x \in [0,1]$, o sea

$$F_n \overset{\mathcal{D}}{-} F$$

siendo F_n distribuciones discretas y F una distribución continua.

Ejemplo 7. Sean $X_1, X_2, \ldots$ variables aleatorias independientes con igual función de densidad uniforme en $(0,\theta)$, es decir

$$f(x) = \begin{cases} \frac{1}{\theta} & 0 < x < \theta \\ 0 & \text{de otra forma} \end{cases} \qquad \theta > 0$$

Defina $M_n = max(X_1,\ldots,X_n)$. Es fácil probar que la distribución de $F_{M_n}(x)$ es

$$F_{M_n}(x) = \begin{cases} 0 & x < 0 \\ \left(\frac{x}{\theta}\right)^n & 0 \le x < \theta \\ 1 & x \ge \theta \end{cases}$$

y por lo tanto

$$\lim_{n \quad \infty} F_{M_n}(x) = \begin{cases} 0 & x < \theta \\ 1 & x \ge \theta \end{cases}$$

o sea, si

$$F(x) = \begin{cases} 0 & x < \theta \\ 1 & x \ge \theta \end{cases}$$

es la función de distribución de la variable aleatoria que toma el valor θ con probabilidad 1 (degenerada en θ), entonces

$$\lim_{n \quad \infty} F_{M_n}(x) = F(x) \quad \forall \quad x \neq \theta \quad x \in C_F$$

y por lo tanto $F_{M_n} \overset{\mathcal{D}}{-} F$. Además

$$P(|M_n-\theta|>\varepsilon) = 1-P(|M_n-\theta|\le\varepsilon) = 1-P(-\varepsilon\le M_n-\theta\le\varepsilon)$$
$$= 1-F_{M_n}(\varepsilon+\theta) + F_{M_n}(\theta-\varepsilon)$$

Puesto que $\varepsilon+\theta > \theta$ entonces $F_{M_n}(\varepsilon+\theta) = 1$ y por lo tanto

$$P(|M_n-\theta|>\varepsilon) = F_{M_n}(\theta-\varepsilon)$$

Si $\varepsilon < \theta$ entonces $F_{M_n}(\theta-\varepsilon) = 0$ y si $0 < \varepsilon < \theta$ entonces

$$F_{M_n}(\theta-\varepsilon) = \left(\frac{\theta-\varepsilon}{\varepsilon}\right)^n = \left(1-\frac{\varepsilon}{\theta}\right)^n \longrightarrow 0 \quad n \to \infty \quad y$$

por lo tanto $P(|M_n-\theta|>\varepsilon) \xrightarrow[n \to \infty]{} 0$ es decir $M_n \xrightarrow{p} \theta$.

Esto demuestra que el "máximo" M_n es un "**estimador consistente de θ**".

Teorema 19 (Teorema de Scheffé) Sean $\{X_n\}_{n\ge1}$ y X variables aleatorias absolutamente continuas con funciones de densidad $\{f_n\}_{n\ge1}$ y f respectivamente y tal que

$$f_n(x) \to f(x) \quad n \to \infty \quad \forall \; x \in \mathbb{R}.$$

Entonces

$$\int_{-\infty}^{\infty} |f_n(x)-f(x)|\,dx \xrightarrow[n \to \infty]{} 0$$

y $X_n \xrightarrow{\mathcal{D}} X$ o sea

$$F_n(x) \to F(x) \quad \forall \; x \in C_F$$

en donde $\{F_n\}_{n\ge1}$ y F son las funciones de distribución de $\{X_n\}_{n\ge1}$ y X respectivamente.

Teorema 20. (Teorema de Slutsky). Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias y sea X una variable aleatoria tal que $X_n \xrightarrow{\mathcal{D}} X$. Sea $\{Y_n\}_{n\geq 1}$ una sucesión de variables aleatorias tal que $Y_n \xrightarrow{p} c$ en donde c es una constante. Entonces

1) $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$

2) $X_n Y_n \xrightarrow{\mathcal{D}} cX$ si $c \neq 0$ y $X_n Y_n \xrightarrow{p} 0$ si $c = 0$

3) $\dfrac{X_n}{Y_n} \xrightarrow{\mathcal{D}} \dfrac{X}{c}$ si $c \neq 0$.

El siguiente resultado, cuya demostración es trivial, es el análogo para el caso discreto del Teorema de Scheffé.

Teorema 21. Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias enteras positivas y X otra variable aleatoria positiva. Entonces

$$P(X_n=k) \xrightarrow[n\,\infty]{} P(X=k) \quad \forall \ k \in \mathbb{R} \quad \Leftrightarrow \quad X_n \xrightarrow{\mathcal{D}} X.$$

Proposición 10. Si $X_n \xrightarrow{p} X$ y g es una función continua entonces $g(X_n) \xrightarrow{p} g(X)$.

El siguiente resultado es de suma importancia en Estadística y Probabilidad.

Teorema 22. (Cramer-Levy). Sea $\{X_n\}_{n\geq 1}$ y X variables aleatorias definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{F}, P)$ con correspondientes funciones características $\{\phi_n(t)\}_{n\geq 1}$ y $\phi(t)$. Si $X_n \xrightarrow{\mathcal{D}} X$ entonces $\phi_n(t) \xrightarrow[n\,\infty]{} \phi(t)$ uniformemente en t sobre intervalos finitos. Suponga ahora que $g(t) = \lim_{n \to \infty} \phi_n(t)$ existe para todo real t y que g es una función continua en cero. Entonces g es la función característica de una variable aleatoria X y $X_n \xrightarrow{\mathcal{D}} X$.

Como aplicación del Teorema 22 probaremos la Ley Débil de los Grandes Números y el Teorema Central del Límite. El siguiente resultado será usado también.

Lema 4. Para cualquier numero complejo z

$$\left(1+\frac{z}{n} + o\left(\frac{1}{n}\right)\right)^n \to e^z \quad \text{cuando } n \to \infty$$

en donde $o\left(\frac{1}{n}\right) \to 0$ cuando $n \to \infty$

11. LEY DE LOS GRANDES NUMEROS Y EL TEOREMA DEL LIMITE CENTRAL

Teorema 23. (Ley Débil de los Grandes Números) Sean $X_1, X_2, \ldots$ variables aleatorias independientes igualmente distribuidas con media finita μ y sea $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Entonces $X_n \xrightarrow{P} \mu$.

Teorema 24. (Ley Fuerte de los Grandes Números) Bajo las hipótesis del Teorema 24

$$P(\bar{X}_n \xrightarrow[n \to \infty]{} \mu) = 1$$

Demostraremos el Teorema 23 usando el siguiente resultado:

Lema 5. Sea X una variable aleatoria tal que $E|X|^n < \infty$ para algún $n \geq 1$. Entonces

$$\hat{\mu}_X(t) = 1 + it(EX) + \ldots + \frac{(it)^n}{n!} EX^n + o(t^n)$$

en donde $\lim_{t \to \infty} \dfrac{o(t^n)}{t^n} = 0.$

Demostración del Teorema 23. Sea ϕ la función característica de X_1. Entonces por el Lema 5

$$\phi(t) = 1+it(EX) + o(t)$$

en donde $\dfrac{o(t)}{t} \underset{t \to 0}{\longrightarrow} 0$. Por el Teorema 18 la función característica de $\hat{\mu}_{\bar{X}_n}(t)$ es

$$\hat{\mu}_{\bar{X}_n}(t) = \phi\left(\frac{t}{n}\right)^n$$

por lo tanto

$$\hat{\mu}_{\bar{X}_n}(t) = \left(1+i\frac{(EX)t}{n} + o\left(\frac{t}{n}\right)\right)^n.$$

Entonces para t fijo, aplicando el Lema 4

$$\phi_{\bar{X}_n}(t) \to e^{i\mu t}.$$

Pero $e^{i\mu t}$ es la función característica de la distribución degenerada en μ y aplicando el Teorema 22

$$\bar{X}_n \xrightarrow{\mathcal{D}} \mu$$

y por lo tanto puesto que μ es una constante $\bar{X}_n \xrightarrow{p} \mu$.

Terema 25. (Teorema del Límite Central). Sean $X_1, X_2, \ldots$ variables aleatorias independientes igualmente distribuidas con media μ y varianza $\sigma^2>0$ y sea $S_n = X_1+\ldots+X_n$. Entonces

$$\frac{S_n-n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathcal{D}} Z$$

donde Z tiene una distribución normal estándar N(0,1), o sea

$$P\left(\frac{S_n-n\mu}{\sigma\sqrt{n}} \le x\right) \underset{n \to \infty}{\longrightarrow} \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}dy \quad \forall \quad x \in \mathbb{R}$$

Demostración. Basta demostrarlo cuando $\mu = 0$ y $\sigma = 1$.

Sea ϕ la función característica de X_1. Entonces aplicando el Lema 4

$$\hat{\mu}(t) = 1-\frac{t^2}{2}+o(t^2)$$

La variable aleatoria S_n tiene función característica

$$\hat{\mu}_{S_n}(t) = \left[1-\frac{t^2}{2}+o(t^2)\right]^n$$

y por lo tanto la v.a. $\frac{S_n}{\sqrt{n}}$ tiene función característica

$$\hat{\mu}_{\frac{S_n}{\sqrt{n}}}(t) = \left[1-\frac{t^2}{2n}+o(t^2/n)\right]^n$$

y aplicando el Lema 4 $\hat{\mu}_{\frac{S_n}{\sqrt{n}}}(t) \underset{n\ \infty}{\longrightarrow} e^{-t^2/2} \quad \forall\ t \in \mathbb{R}$.

Pero $e^{-t^2/2}$ es la función característica de una distribución normal estándar, por lo que aplicando los Teoremas 16 y 22 se tiene el resultado. ∎

Ejemplo 8. (Este ejemplo es una aplicación del Teorema del Límite Central al Cálculo).

$$\lim_{n\ \infty} \sum_{k=0}^{n} \frac{n^k e^{-n}}{k!} = \frac{1}{2}.$$

Demostración. Sean $X_1, X_2, \ldots$ variables aleatorias independientes con distribución de Poisson $P(1)$. Entonces S_n tiene distribución $P(n)$. Por lo tanto aplicando el Teorema 25 con $x = 0$ se obtiene

$$\lim_{n\to\infty} P\left[\frac{S_n-n}{\sqrt{n}} \le 0\right] = \Phi(0) = \frac{1}{2}.$$

Pero

$$P\left[\frac{S_n-n}{\sqrt{n}} \leq 0\right] = P(S_n \leq n) = \sum_{k=0}^{n} \frac{n^k e^{-n}}{k!}$$

y por lo tanto

$$\lim_{n \to \infty} \sum_{k=0}^{n} \frac{n^k e^{-n}}{k!} = \frac{1}{2}$$

12. TEOREMA DE GLIVENKO-CANTELLI

Sea X una variable aleatoria con distribución F y $X_1,\ldots,X_n$ variables aleatorias independientes con igual distribución F. Sea $X_{(1)},\ldots X_{(n)}$ las estadísticas de orden de $X_1,\ldots,X_n$, o sea $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$. Definimos la **función de distribución empírica**

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_j \leq x\}} = \begin{cases} 0 & \text{si} \quad x < X_{(1)} \\ \frac{k}{n} & \text{si} \quad X_{(k)} \leq x < X_{(k+1)} \qquad k=1,\ldots n-1 \\ 1 & \text{si} \quad x \geq X_{(n)} \end{cases}$$

Teorema 26. Para $x \in \mathbb{R}$ fijo pero arbitrario la variable aleatoria $F_n^*(x)$ tiene función de probabilidad

$$P\left[F_n^*(x) = \frac{j}{n}\right] = \binom{n}{j}[F(x)]^j[1-F(x)]^{n-j} \qquad j = 0,1,..,n$$

con media

$$E(F_n^*(x)) = F(x)$$

y varianza

$$VAR(F_n^*(x)) = \frac{1}{n}F(x)(1-F(x))$$

o sea $nF_n^*(x)$ tiene una distribución binomial con parámetros n y $F(x)$.

Demostración. Sea $Y_x^i = 1_{\{X_i \le x\}}$. Entonces Y_x^i es una variable aleatoria Bernoulli con parámetro $P_x = P(X_i \le x) = F(x)$ o sea

$$P(Y_x^i=1) = F(x)$$

$$P(Y_x^i=0) = 1-F(x)$$

Puesto que las variables aleatorias $X_1, \ldots, X_n$ son independientes entonces las variables aleatorias $Y_x^1, \ldots, Y_x^n$ son independientes y por lo tanto

$$n\, F_n^*(x) = \sum_{i=1}^{n} Y_X^i.$$

tiene una distribucióon binomial con parámetros n y $P_x=F(x)$.

Corolario 6. Para cada $x \in \mathbb{R}$

$$F_n^*(x) \xrightarrow{P} F(x)$$

Demostración. Se sigue fácilmente utilizando la Ley Débil de los grandes números, aplicada a las variables aleatorias $\left\{Y_x^i\right\}_{i\ge1}$.

Corolario 7. Para cada $x \in \mathbb{R}$

$$\frac{\sqrt{n}\,(F_n^*(x)-F(x))}{\sqrt{F(x)(1-F(x))}} \xrightarrow{D} Z$$

en donde Z tiene una distribución N(0,1).

Demostración. Se sigue fácilmente aplicando el Teorema del Límite Central a la sucesión de variables aleatorias $\left\{Y_x^i\right\}_{i\ge1}$.

Teorema 27. (Glivenko-Cantelli). Para cada $\varepsilon > 0$

$$\lim_{n\ \infty} P\left(\sup_{-\infty<x<\infty} |F_n(x)-F(x)| > \varepsilon\right) = 0$$

o sea $F_n^*(x)$ converge uniformemente a $F(x)$, en probabilidad.

Además

$$P(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n^*(x)-F(x)| = 0) = 1$$

13. DOMINIOS DE ATRACCION Y LEYES ESTABLES

Sea $X_1, X_2, \ldots,$ una sucesión de variables aleatorias independientes igualmente distribuidas con función de distribución F. Si existen sucesiones de constantes a_n, b_n, $n \geq 1$ tal que la variable aleatoria

$$(\sum_{i=1}^{n} X_i - a_n)/b_n$$

converge en distribución a una distribución G decimos que **F es atraída a G.** El conjunto de todas las funciones de distribución que son atraídas a G se llama el **dominio de atracción de G.** El teorema del límite central (Teorema 25) nos muestra que la función de distribución normal tiene un dominio de atracción no vacío. Este es un caso especial de una clase de distribuciones con esta propiedad llamadas **distribuciones estables.** Así, decimos que una distribución es estable si tiene un dominio de atracción no vacío, es decir, es una distribución límite para sumas de variables aleatorias independientes e igualmente distribuidas. EL siguiente resultado es una representación canónica para las distribuciones estables.

Teorema 28. (P. Levy). Para que una función de distribución F sea estable es necesario y suficiente que el logaritmo de su función característica se pueda expresar de la siguiente manera

$$\log \phi(t) = i\gamma t - c|t|^{\alpha} [1 + i\beta\frac{t}{|t|} w(t,\alpha)]$$

donde α, β, γ y c son constantes; γ es cualquier número real, $-1\leq\beta\leq1$, $0<\alpha\leq2$, $c>0$ y

$$
w(t,\alpha) = \begin{cases} \operatorname{tg} \frac{\pi}{2}\,\alpha & \text{si} \quad \alpha \neq 1 \\ \frac{2}{\pi} \log |t| & \text{si} \quad \alpha \neq 1 \end{cases}
$$

En tal caso decimos que (α,β,γ,c) es el vector de parámetros de la distribución estable F. El parámetro α se conoce como el **exponente** o **índice característico** de F; para el caso $\alpha\notin(0,2]$ se demuestra que $\phi(t)$ no es una función característica. El parámetro β es el **sesgo**, si $\beta=0$ entonces la distribución es simétrica alrededor del **parámetro de localización** γ; es sesgada a la derecha para $0<\beta\leq1$ y sesgada a la izquierda para $-1\leq\beta<0$, además la variable aleatoria es positiva si y solamente si $\beta=-1$, $\alpha<1$ y $\gamma>0$; es negativa si y sólo si $\beta=1$, $\alpha<1$ y $\gamma>0$. La constante c es el **parámetro de escala**. Observe que el caso $\alpha=2$, $\beta=0$, $\gamma=\mu$ y $c=\sigma^2/2$ corresponde a la función característica de la distribución normal con media μ y varianza σ^2; y que el caso $\alpha=1$, $\beta=0$ corresponde a la distribución Cauchy con parámetros de localización y escala γ y c respectivamente.

Uno de los problemas importantes en la teoría de la leyes estables es la determinación de sus dominios de atracción. Es claro que toda distribución estable pertenece a su propio dominio de atracción y que todas las distribuciones con segundo momento finito pertenecen al dominio de atracción de la distribución normal (Teorema 25). El resultado que se presenta a continuación caracteriza los dominos de atracción de una distribución estable con índice α $(0<\alpha<2)$ a través del comportamiento de las colas de la distribución.

Teorema 29. (P. Levy). Una función de distribución F pertenece al dominio de atracción de una distribución estable con índice $\alpha(0<\alpha<2)$ si y solamente si cumple las siguentes dos condiciones:

$$\text{i)} \quad \frac{F(-x)}{1-F(x)} \to \frac{c_1}{c_2} \quad \text{cuando } x \to \infty$$

y

ii) Para cualquier constante $k > 0$

$$\frac{1-F(x) + F(-x)}{1-F(kx) + F(-kx)} \to k^{\alpha} \text{ cuando } x \to \infty$$

En donde las constantes c_1 y c_2 so no negativas y $c_1 + c_2 > 0$.

La convergencia establecida en i) indica que las colas de la distribución F, y por lo tanto las de una distribución estable, están balanceadas y ii) nos dice que éstas varían lentamente cuando $x \to \infty$, lo que significa que el peso en las colas $1-F(kx)+F(-kx)$ para cualesquiera $x>0$ y $k>0$ es del mismo orden que $1-F(x)+F(-x)$, es decir las distribuciones estables con índice $0<\alpha<2$ tiene **"colas pesadas"**. Lo anterior no es cierto para la distribución normal ya que de acuerdo al **teorema del límite central de Lindeberg-Feller** se tiene que una distribución F pertenece al dominio de atracción de una distribución normal si y solamente si la siguiente condición se cumple

$$\int_{|x|\geq t} dF(x) = o\left(t^2 \int_{|x|<t} x^2 dF(x)\right) \quad \text{cuando } t \to \infty$$

o equivalentemente

$$\lim_{t\to\infty} \frac{\int_{|x|\geq t} dF(x)}{t^2 \int_{|x|<t} x^2 dF(x)} = 0$$

lo cual significa que las colas de la distribución F son despreciables.

14. DISTRIBUCIONES DE EXTREMOS

Sean $X_1,\dots X_n$ variables aleatorias independientes con la misma función de distribución F y sea

$$M_n = max(X_1,\dots,X_n).$$

Ya que $P(M_n \le z) = P(X_1 \le z,\dots X_n \le z) = P(X_1 \le z)\dots P(X_n \le z)$ entonces

$$F_{M_n}(z) = (F(z))^n \quad z \in \mathbb{R}.$$

Se puede probar (resultado análogo a los Teoremas 25 y 28) que si existen sucesiones de constantes $\{a_n\}_{n\ge1}$ y $\{b_n\}_{n\ge1}$ $b_n > 0$ $n \ge 1$ tal que

$$\frac{M_n - a_n}{b_n} \xrightarrow{\mathcal{D}} M$$

entonces la distribución de la variable aleatoria M tiene que ser de alguno de los siguientes tres tipos:

i) $G_M(x) = exp\ (-e^{-x}) \quad -\infty < x < \infty$

ii) $G_M(x) = \begin{cases} 0 & x \le 0 \\ exp(-x^{\alpha}) & x > 0 \end{cases}$ para algún $\alpha > 0$

iii) $G_M(x) = \begin{cases} exp(-(-x)^{\alpha}) & \alpha > 0 \quad x \le 0 \\ 1 & \quad\quad\quad x > 0 \end{cases}$

llamadas **distribuciones de Extremos.**

15. TEOREMA DE RADON-NIKODYM Y ESPERANZA CONDICIONAL

Sean $(\Omega,\mathcal{F},\mu_1)$ y $(\Omega,\mathcal{F},\mu_2)$ dos espacios de medida y supongamos que $\mu_1(A)=0$ implica que $\mu_2(A) = 0$, en este caso decimos que μ_2 es **absolutamente continua con respecto a** μ, y escribimos $\mu_2 \ll \mu_1$. Si $\mu_1(N^c)=0$ para algún conjunto $N \in \mathcal{F}$ tal que $\mu_2(N)=0$, entonces decimos que μ_1 es μ_2 **singular**.

Teorema 30 (Radon-Nikodyn). Sean ν_1, ν_2 y μ medidas σ-finitas en un espacio medible $(\Omega, \mathcal{F})$ tal que $\nu_1 \ll \mu$, $\nu_2 \ll \mu_2$ y suponga que $\nu=\nu_1-\nu_2$ (****) está bien definida en $\mathcal{F}$ (o sea $\nu_1(\Omega)$ y $\nu_2(\Omega)$ no son al mismo tiempo ∞). Entonces existe una función $\mathcal{F}$-medible g, finita casi seguramente μ tal que

$$\nu(A) = \int_A g d\mu \quad \forall\ A \in \mathcal{F} \tag{13}$$

Además g es única c.s. $[\mu]$.

La función g definida por (13) se llama la **derivada de Radon Nikodyn de ν con respecto a μ** y se denota de manera sugestiva por **$d\nu/d\mu$**,

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabiidad arbitrario y sean ξ una variable aleatoria tal que $E|\xi|<\infty$ y $\mathcal{G}$ una sub σ-álgebra de $\mathcal{F}$. Consideremos la función de conjunto

$$\nu(E) = \int_E \xi dP \quad E \in \mathcal{G}$$

La función ν es una medida con signo finita y $\nu \ll P_{\mathcal{G}}$ en donde $P_{\mathcal{G}}$ es la restricción de P de $\mathcal{F}$ a $\mathcal{G}$. Por el teorema de Radon-Nikodyn existe una función $\mathcal{G}$-medible y P-integrable f en Ω determinada únicamente c.s. $[P_{\mathcal{G}}]$ tal que

$$\nu(F) = \int_E f dP_{\mathcal{G}} = \int_E f dP \quad \forall\ E \in \mathcal{G}.$$

Escribiremoes $f=\underline{E}(\xi|\mathcal{G})$ y la llamaremos la **esperanza condicional** de ξ dada la σ-álgebra $\mathcal{G}$. Entonces la esperanza condicional $\underline{E}(\xi|\mathcal{G})$ de ξ dado $\mathcal{G}$ es una función $\mathcal{G}$-medible y P-integrable la cual queda determinada en forma única por la igualdad

$$\int_E \xi dP = \int_E \underline{E}(\xi|\mathcal{G}) dP \quad \forall\ E \in \mathcal{G}$$

(****) ν definida en esta forma se llama medida con signo

Las propiedades más sencillas de esperanza condicional son las siguientes:

Teorema 31. Sean ξ y η variables aleatorias con esperanza finita y a y b números reales.

1) $\underline{E}\{\underline{E}(\xi|\mathcal{G})\} = \underline{E}(\xi)$

2) $\underline{E}(a\xi+b\eta|\mathcal{G}) = a\underline{E}(\xi|\mathcal{G})+b\underline{E}(\eta|\mathcal{G})$ c.s.

3) Si ξ es $\mathcal{G}$-medible, $\underline{E}(\xi|\mathcal{G}) = \xi$ c.s.

En particular si $\xi = a$ c.s. entonces $\underline{E}(\xi|\mathcal{G}) = a$ c.s.

4) Si $\xi\geq 0$ c.s. entonces $\underline{E}(\xi|\mathcal{G}) \geq 0$ c.s.

5) Si $\xi \leq \eta$ c.s. entonces $\underline{E}(\xi|\mathcal{G}) \leq \underline{E}(\eta|\mathcal{G})$ c.s.

6) $|\underline{E}(\xi|G)| \leq \underline{E}(|\xi||\mathcal{G})$.

Demostración

1) Puesto que $\Omega \in \mathcal{G}$

$$E(\xi) = \int_\Omega \xi dP = \int_\Omega \underline{E}(\xi|\mathcal{G})dP = \underline{E}(\underline{E}(\xi|\mathcal{G}))$$

2) $\forall\ E \in \mathcal{G}$

$$\begin{aligned}\int_E (a\xi+b\eta)dP &= a\int_E \xi dP+b\int_E \eta dP \\ &= a\int_E \underline{E}(\xi|\mathcal{G})dP+b\int_E \underline{E}(\eta|\mathcal{G})dP \\ &= \int_E \{a\underline{E}(\xi|\mathcal{G})+b\underline{E}(\eta|\mathcal{G})\}dP\end{aligned}$$

3) La primera parte es obvia.

Definamos $\eta(\omega) = a \quad \forall\ \omega \in \Omega$, entonces $\forall\ B \in \mathcal{B}(\Omega)$

$$\eta^{-1}(B) = \begin{cases}\phi & \text{si } a \notin B \\ \Omega & \text{si } a \in B\end{cases}$$

o sea $\eta^{-1}(B) \in \mathcal{G}$ y por lo tanto η es $\mathcal{G}$-medible.

4) Si $\xi \geq 0$ el Teorema de Radon-Nikodyn asegura que $\underline{E}(\xi|\mathcal{G}) \geq 0$ c.s.

5) y 6) siguen en forma obvia.

Teorema 32. Sea ξ una variable aleatoria tal que $E|\xi|<\infty$ y $\mathcal{G}_1$ y $\mathcal{G}_2$ dos sub σ-álgebras de $\mathcal{F}$ tal que $\mathcal{G}_2 \subset \mathcal{G}_1 \subset \mathcal{F}$. Entonces

$$\underline{E}\{\underline{E}(\xi|\mathcal{G}_2)|\mathcal{G}_1\} = \underline{E}(\xi|\mathcal{G}_2) = \underline{E}\{\underline{E}(\xi|\mathcal{G}_1)|\mathcal{G}_2\} \quad \text{c.s.}$$

Demostración. Sea $E \in \mathcal{G}_2 \subset \mathcal{G}_1$. Entonces

$$\int_E \underline{E}(\xi|\mathcal{G}_1)dP = \int_E \xi dP = \int_E \underline{E}(\xi|\mathcal{G}_2)dP$$

lo cual implica que $\underline{E}\{\underline{E}(\xi|\mathcal{G}_1)|\mathcal{G}_2\} = \underline{E}(\xi|\mathcal{G}_2)$ c.s.

$\underline{E}(\xi|\mathcal{G}_2)$ es $\mathcal{G}_2$-medible y por lo tanto $\mathcal{G}_1$-medible lo cual da la primera igualdad.

Teorema 33. Sea ξ una variable aleatoria tal que $E|\xi|<\infty$ y supongamos que $\sigma(\xi)$ y $\mathcal{G}$ son independientes. Entonces

$$\underline{E}(\xi|\mathcal{G}) = E\xi \quad \text{c.s.}$$

En particular si ξ y η son variables aleatorias integrables tal que son independientes entonces $\underline{E}(\xi|\eta) = E\xi$ c.s. en donde $\underline{E}(\xi|\eta) = \underline{E}(\xi|\sigma(\eta))$.

Demostración. $\forall\ E \in \mathcal{G}$ ξ y 1_E son variables aleatorias independientes y por lo tanto

$$\int_E \xi dP = \int_\Omega \xi 1_E dP = \underline{E}(\xi 1_E) = E(\xi)E(1_E) = E(\xi)\int_E dP$$
$$= \int_E E(\xi)dP$$

Teorema 34. (Teorema de Convergencia Monótona Condicional). Sea $\{\xi_n\}_{n\geq 1}$ una sucesión creciente de variables aleatorias no-negativas tal que $\lim\limits_{n\to\infty} \xi_n = \xi$ c.s. en donde $E(\xi) < \infty$. Entonces

$$\underline{E}(\xi|\mathcal{G}) = \lim_{n\to\infty} \underline{E}(\xi_n|\mathcal{G})$$

Demostración. Por el teorema 32(5) la sucesión $\{\underline{E}(\xi_n|\mathcal{G})\}$ es no decreciente y no-negativa c.s., $\eta = \lim\limits_{n\to\infty} \underline{E}(\xi_n|\mathcal{G})$ es $\mathcal{G}$-medible y usando el Teorema de Convergencia monótona (Teorema 12) dos veces obtenemos que para $E \in \mathcal{G}$

$$\int_E \lim_{n\to\infty} \underline{E}(\xi_n|\mathcal{G})dP \overset{\text{TCM}}{=} \lim_{n\to\infty} \int_E \underline{E}(\xi_n|\mathcal{G})dP = \lim_{n\to\infty} \int_E \xi_n dP$$
$$\overset{\text{TCM}}{=} \lim_{n\to\infty} \int_E \xi_n dP = \int_E \xi_n dP$$
$$\therefore \underline{E}(\lim_{n\to\infty} \xi|\mathcal{G}) = \lim_{n\to\infty} \underline{E}(\xi_n|\mathcal{G}) \quad \text{c.s.}$$

En forma similar se prueban los siguientes dos resultados.

Teorema 35 (Lema de Fatou Condicional).

Sea $\{\xi_n\}_{n\geq 1}$ una sucesión de variables aleatorias no-negativas tal que $E(\xi_n)<\infty$ y $E(\liminf \xi_n)<\infty$. Entonces

$$\underline{E}(\liminf \xi_n|\mathcal{G}) \leq \liminf_{n\to\infty} \underline{E}(\xi_n|\mathcal{G}) \text{ c.s.}$$

Teorema 36. (Teorema de Convergencia Dominada Condicional)

Sea $\{\xi_n\}$ una sucesión de variables aleatorias tal que $\xi_n \to \xi$ c.s. y $|\xi_n| \leq \eta$ c.s. $\forall\ n \geq 1$ en donde $E|\eta|<\infty$. Entonces

$$\underline{E}(\xi|\mathcal{G}) = \lim_{n\to\infty} \underline{E}(\xi_n|\mathcal{G}) \quad \text{c.s.}$$

Como aplicación del Teorema de Convergencia Dominada Condicional se tiene el siguiente resultado, el cual es bastante útil.

Teorema 37. Sean ξ y η variables aleatorias tales que $E|\eta|<\infty$ y $E|\xi\eta|<\infty$ y tal que ξ es $\mathcal{G}$-medible. Entonces

$$\underline{E}(\xi\eta|\mathcal{G}) = \xi\underline{E}(\eta|\mathcal{G}) \text{ c.s.}$$

Demostración.

La demostración se hace por casos. Primero para $\xi = 1_A$ $A \in \mathcal{G}$. Entonces $\xi E(\eta|\mathcal{G})$ es $\mathcal{G}$-medible y $\forall$ $E \in \mathcal{G}$

$$\int_E \xi\underline{E}(\eta|\mathcal{G})dP = \int_{E\cap A} \underline{E}(\eta|\mathcal{G})dP = \int_{E\cap A} \eta dP = \int_E \xi\eta dP .$$

Trivialmente el resultado es verdad para funciones simples.

Si ξ es $\mathcal{G}$-medible, sea $\{\xi_n\}$ una sucesión de funciones simples $\mathcal{G}$-medibles (ver el Teorema 11) tal que $\forall$ $\omega \in \Omega$ $\lim\limits_{n\to\infty} \xi_n(\omega) = \xi(\omega)$ y $|\xi_n(\omega)| \le |\xi(\omega)|$ $\forall$ $n \ge 1$. Entonces por el Teorema 36

$$\underline{E}(\xi\eta|\mathcal{G}) = \lim_{n\to\infty} \underline{E}(\xi_n\eta|\mathcal{G}) = \lim_{n\to\infty} \xi_n\underline{E}(\eta|\mathcal{G}) = \xi\underline{E}(\eta|\mathcal{G}) \text{ c.s.} \quad \blacksquare$$

BIBLIOGRAFIA

1. R. Ash. *Real Analysis and Probability.* Accademic Press, New York, 1972.

2. K. L. Chung. *A Course in Probability Theory.* Academic Press, New York, 1974

3. A. N. Shiryayev. *Probability.* Springer Verlag, New York, 1984.