Clase - Lunes 8 de noviembre

Monday, November 8, 2021 7:09 AM

Temario

Teoría de probabilidades.

1. Teoría elemental de probabilidades
2. Fundamentos (Axiomatización de Kolmogorov)
3. Convergencia
4. Ley de los grandes números
5. Función característica - Teorema del límite central
6. Condicionamiento.

Bibliografía.

- Chung K.L. Course in Probability (2001)
- Shirayev, A.N. Probability I (2016)

> En esta sección estudiamos las limitaciones de un modelo finito de probabilidades

Teoría elemental de probabilidades

Caso finito:

- Número finito de resultados

$$\Omega = \{\omega_1, \dots, \omega_N\}$$

Espacio muestral (Ω); eventos elementales (ω_i); muestras puntuales

Ejemplo:

1. Lanzamiento de una moneda:

$$\Omega = \{c, s\}$$

2. n lanzamientos de una moneda.

$$\Omega = \{\omega : \omega = (a_1, \dots, a_n) \ , \ a_i \in \{c, s\}\}$$

$$|\Omega| = 2^n$$

3. Muestreo con reemplazo:

bolas numeradas $1, \dots, M$

se eligen al azar n bolas con reemplazo

(3.1) Con orden:

sí importa el orden

$$\Omega = \{\omega : \omega = (a_1, \dots, a_n) \ : \ a_i = 1, \dots, M\} \qquad |\Omega| = M^n$$

(3.2) Sin orden

no importa el orden

$$\Omega = \{\omega : \omega = [a_1, \dots, a_n] \ : \ a_i = 1, \dots, M\} \qquad |\Omega| = \binom{M+n-1}{n}$$

(4) Muestreo sin reemplazo

(4.1) Con orden:

$$\Omega = \{\omega : \omega = (a_1, \dots, a_n), \ a_k \neq a_j \ \text{si} \ k \neq j \ , \ a_i = 1, \dots, M\}$$

$$|\Omega| = M(M-1)\cdots(M-n+1) = \frac{M!}{(M-n)!}$$

(4.2) Sin orden

$\Omega = \{ \omega : \omega = [a_1, \dots, a_n] : a_k \neq a_j \text{ si } k \neq j \quad a_i = 1, \dots, M \}$

$$|\Omega| \cdot n! = \frac{M!}{(M-n)!} \Rightarrow |\Omega| = \frac{M!}{(M-n)!\,n!} = \binom{M}{n}$$

• Si $M = 3$, $n = 2$

(1,1), (1,2), (1,3) (2,1), (2,2), (2,3), (3,1) (3,2), (3,3)	[1,1], [1,2], [1,3] [2,2], [2,3], [3,3]	Con reemplazo
(1,2), (1,3), (2,1), (2,3) (3,1), (3,2)	[1,2], [1,3] [2,3]	sin reemplazo
Con orden	Sin orden	

M^n	$\binom{M+n-1}{n}$	Con reemplazo
$\frac{M!}{(M-n)!}$	$\binom{M}{n}$	Sin reemplazo
Ordenado	Sin orden	Muestra / Tipo

Modelo probabilístico

• Se lanza 3 veces una moneda

$\Omega = \{ccc, ccs, \dots, sss\}$

Sea $A = \{ccc, ccs, csc, scc\}$

¿Qué tan posible es que $\omega \in A$?

Definición 0.1 - Evento- A: Evento ($A \subseteq \Omega$)

↳ operaciones de conjuntos

• $A \cup B = \{\omega : \omega \in A \text{ ó } \omega \in B\}$ • $A \cap B = \{\omega : \omega \in A \text{ y } \omega \in B\}$

• $A \setminus B = \{\omega : \omega \in A \text{ y } \omega \notin B\}$ • $A^c = \{\omega : \omega \notin A\} = \Omega \setminus A$.

• $\emptyset$: Evento imposible • Ω Evento seguro

A y B son incompatibles si $A \cap B = \emptyset$ disjuntos o contradictorios

Definición 0.2. - Álgebra- (Mínima estructura que vamos a exigir para un conjunto de eventos)

Sea $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ tal que

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

2) Si $A \in \mathcal{A}$, $B \in \mathcal{A}$, los conjuntos

$A \cup B$, $A \cap B$, $A \setminus B \in \mathcal{A}$

2*) Si $A \in \mathcal{A}$ y $B \in \mathcal{A}$, entonces

$A^c \in \mathcal{A}$

$A \cup B \in \mathcal{A}$

$A \cap B \in \mathcal{A}$

$\mathcal{A}$ es un álgebra

Ejemplos de álgebras

a) $\{\emptyset, \Omega\}$ álgebra trivial

b) $A \subset \Omega$, $\{A, A^c, \emptyset, \Omega\}$ álgebra generada por A

c) $\mathcal{A} = \{A : A \subset \Omega\}$ álgebra que contiene todos los subconjuntos

Otro ejemplo:

$\mathcal{D} = \{D_1, \dots, D_n\}$ una partición de Ω

$$\boxed{D_i \cap D_j = \emptyset \quad \text{si } i \neq j \qquad \bigcup_{j=1}^{n} D_j = \Omega}$$

$\alpha(D)$: la más pequeña álgebra que contiene a $\mathcal{D}$

$$\alpha(\mathcal{D}) = \{A = \bigcup_{j \in I} D_j : I \subseteq \{1, \dots, n\}\}$$

$(\Omega, \mathcal{A})$ $\quad \Omega = \{w_1, \dots, w_N\}$

1) $0 \leq P(w_i) \leq 1$ (no negatividad)

2) $\sum_{i=1}^{N} P(w_i) = 1$ (normalización)

$(*)$

Sea

$$P(A) = \sum_{i: w_i \in A}' p(w_i)$$

Modelo probabilístico (finito)

$\Omega = \{w_1, \dots, w_N\}$

Sea $\underset{\sim}{p}: \Omega \to \mathbb{R}$ tal que $\sum_{i=1}^{N} p(w_i) = 1$

Para cada $A \subseteq \Omega$

$$P(A) = \sum_{w_i \in A} P(w_i)$$

$(\Omega, \mathcal{A} = \mathcal{P}(\Omega), P) \leftarrow$ modelo probabilístico.

Ejemplos:

Modelo binomial, binomial negativo, etc.

Definición 1.1. $(\Omega, \mathcal{A}, P)$: espacio de probabilidad (FINITO)

* Ω : espacio muestral
* $\mathcal{A}$: álgebra de Ω
* p : como en $(*)$

$P(\{w_i\}) = p(w_i) \leftarrow$ probabilidad discreta.

Propiedades

* $p(\emptyset) = 0$
* $p(\Omega) = 1$
* $p(A \cup B) = p(A) + p(B) - p(A \cap$
* Si $A \cap B = \emptyset$ $\quad p(A \cup B) = p(A) + p(B)$
* $p(A^c) = 1 - p(A)$

Ajustando el modelo probabilístico finito.

Equiprobabilidad ↷ igual probabilidad

$$p(\omega_i) = k, \qquad i = 1, \dots, N$$

$$P(\Omega) = \sum_{i=1}^{N} p(\omega_i) = N \cdot k \Rightarrow p(\omega_i) = \frac{1}{N} \Rightarrow k = \frac{1}{N}$$

En este caso

$A \in \mathcal{A}$:

$$p(A) = \sum_{\omega_i \in A} \frac{1}{N} = \frac{|A|}{N}$$

Propiedades.

- $P(\emptyset) = 0$
- $P(\Omega) = 1$
- Si $A \cap B = \emptyset$, entonces $P(A \cup B) = P(A) + P(B)$
- $P(A^c) = 1 - P(A)$

Ejemplo: (Problema de coincidencia)

Una urna, contiene M bolas numeradas $1, \dots, M$ se extraen n con reemplazo

$$\Omega = \{ \omega : \omega = (a_1, \dots, a_n) : a_i = 1, \dots, M \}$$

$$|\Omega| = M^n$$

Sea el evento A

$$A = \{ \omega : \omega = (a_1, \dots, a_n) : a_i \neq a_j \text{ si } i \neq j \text{ con } a_i = 1, \dots, M \}$$

$$|A| = M(M-1) \cdots (M-n+1)$$

$$P(A) = \frac{M(M-1) \cdots (M-n+1)}{M^n}$$

$$= \left(1 - \frac{1}{M}\right)\left(1 - \frac{2}{M}\right) \cdots \left(1 - \frac{n-1}{M}\right)$$

En particular, cumpleaños:

n estudiantes

365 días son equiprobables.

B^c: todos cumplen años en días diferentes

B: al menos dos personas cumplan años en el mismo día.

$$P(B) = 1 - P(B^c)$$

$$= 1 - \frac{365(365-1) \cdots (365-n+1)}{365^n}$$

Miércoles, 11- de noviembre de 2021

Ejemplos:

i) Binomial

$$\Omega = \{ \omega : \omega = (a_1, \dots, a_n) , \quad a_i \in \{0,1\} \}$$

Para $\omega = (a_1, \dots, a_n)$, $\quad p(\omega) = p^{\sum a_i} q^{n - \sum a_i} \qquad p \geq 0 \qquad p + q = 1.$

Consideremos

$$A_k = \{ \omega : \omega = (a_1, \dots, a_n) : \sum_{i=1}^{n} a_i = k \} \qquad |A_k| = \binom{n}{k}$$

$$\sum_{\omega \in \Omega} p(\omega) = \sum_{k=0}^{n} \sum_{\omega \in A_k} p(\omega)$$

$$= \sum_{k=0}^{n} \sum_{\omega \in A_k} p^k q^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \quad = (p+q)^n = 1$$

$$p(k) = p(A_k) = \binom{n}{k} p^k q^{n-k}$$
$$k = 0, \dots, n$$

2) Marcha aleatoria

Paso adelante o atrás con una probabilidad determinada

Camino $\quad \omega = (a_1, a_2, \dots, a_n)$, $\quad a_i \in \{-1, 1\} \qquad p + q = 1 \qquad p, q \geq 0$

$$p(\omega) = p^{r(\omega)} q^{n - r(\omega)} \qquad \text{donde } r(\omega) : \# \text{ 1's en } \omega$$

$$r(\omega) = \frac{\left(\sum_{i=1}^{n} a_i + n \right)}{2}$$

$A_k :=$ después de n pasos, la partícula se encuentra en la posición k

$$r(\omega) - (n - r(\omega)) = k$$

Esto implica que $r(\omega) = \frac{n+k}{2}$, $\quad r(\omega) \in \mathbb{N} \cup \{0\}$

$$P(A_k) = \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}} q^{\frac{n-k}{2}}$$

→ Por la fórmula de Stirling

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e} \right)^n$$

Si $p = q = 1/2 \qquad P(A_k) = \binom{n}{\frac{n+k}{2}} 2^{-n}$

¿Cuál es la probabilidad de regresar a 0 en n pasos?

$$P(A_0) = \binom{n}{\frac{n}{2}} 2^{-n}$$

$$\binom{n}{\frac{n}{2}} = \frac{n!}{\left(\frac{n}{2}\right)!^2} = \frac{\sqrt{2\pi} \sqrt{n}\, e^{-n} n^n}{\sqrt{2\pi} \left(\sqrt{\frac{n}{2}}\, e^{-\frac{n}{2}} \cdot \left(\frac{n}{2}\right)^{n/2} \right)^2} = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{\frac{n}{2}}} \cdot \frac{n^n}{n^{n/2} \left(\frac{1}{2}\right)^{n/2}}$$

$$= \frac{1}{\sqrt{\pi}} n^{-n/2} 2^{n/2} = 2^{n/2}$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{\sqrt{n}\; n^n}{n/2 \; n^n (1/2)^n} = \frac{2}{\sqrt{2\pi n}} 2^n$$

$$P(A_0) = \binom{n}{n/2} 2^{-n} \approx \frac{2}{\sqrt{2\pi n}} \xrightarrow[n\to\infty]{} 0$$

Probabilidad Condicional - Independencia

Definición $A, B \in \mathcal{A}$ $\quad P(A) > 0 \quad P(B|A) = \frac{P(A\cap B)}{P(A)} \quad P(\{X = x\}) = 0$

Propiedades

- $P(A|A) = 1$
- $P(\emptyset | A) = 0$
- Si $A \subseteq B$ $\quad P(B/A) = 1$
- Si $B_1 \cap B_2 = \emptyset$ $\quad P(B_1 \cup B_2 | A) = P(B_1|A) \cup P(B_2|A)$

Si $\mathcal{D} = \{B_1, \dots, B_n\}$ una partición de Ω tal que $P(B_i) > 0$: $i = 1, \dots, n$

$$P(A) = P\left(\bigcup_{i=1}^{n} A \cap B_i\right) = \sum_{i=1}^{n} P(A\cap B_i)$$
$$= \sum_{i=1}^{n} P(B_i) P(A|B_i)$$

Si $P(A) > 0$, $\quad P(B_j|A) = \frac{P(B_j\cap A)}{P(A)}$, $\quad j = 1, \dots, n$

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}$$ Fórmula de Bayes

Independencia

$P(B) = P(B|A)$ $\; P(A)>0$ y $\qquad P(A) = P(A|B) \qquad P(B) > 0$

Definición A y B son independientes $\quad P(A\cap B) = P(A)\cdot P(B)$

Definición $\mathcal{A}_1$ y $\mathcal{A}_2$ son familias de eventos.

$\mathcal{A}_1$ y $\mathcal{A}_2$ son independientes si $\quad \forall A_i \in \mathcal{A}_i \quad i = 1, 2, \dots$

$$P(A_1 \cap A_2) = P(A_1)\cdot P(A_2)$$

Observación

A_1 y A_2 independientes si y sólo si $\alpha(A_1)$ y $\alpha(A_2)$ son independientes

$\downarrow$

$\{\emptyset, A_1, A_1^c, \Omega\}$ y $\{\emptyset, A_2, A_2^c, \Omega\}$ son independientes

Una familia $\{A_1, \dots, A_n\}$ son independientes si para cada $k = 2, \dots, n$ y $1 \le i_1 < \dots \le i_k \le n$

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k})$$

- $\{\mathcal{A}_1, \dots, \mathcal{A}_n\}$ son independientes si:

$\forall \{A_1, \dots, A_n\}$ tal que $A_i \in \mathcal{A}_i$ son independientes

Observación

Independencia 2 a 2 no implica ~~ant~~ independencia

Ejemplo

$\Omega = \{(i,j) : i,j = 1,\dots,6\}$

$A = \{(i,j) : i,j = 1,2,5\}$

$B = \{(i,j) : i,j = 4,5,6\}$

$C = \{(i,j) : i+j = 9\}$

$A \cap C = \{(i,j) : j = 1,2, \ , \ i+j = 9\}$
$= \{ \cancel{\dots}, (5,4)\}$

$A \cap B = \{(i,j) : j = 1,2,3 \ y \ j = 4,5,6\}$
$= \{(i,j) \ j = 5\}$

$B \cap C = \{(i,j) : j = 4,5,6 \ ; \ i+j = 9\}$
$= \{(4,5), (5,4), (3,6)\}$

$A \cap B \cap C = \{(4,5)\}$

$P(A) = 1/2,$
$P(B) = 1/2$
$P(C) = 1/9$

$P(A\cap B) = \frac{6}{36} = \frac{1}{6} \neq \frac{1}{4} = P(A)P(B)$

$P(B\cap C) = \frac{3}{36} = \frac{1}{12} \neq \frac{1}{18} = P(B)P(C)$

$P(A\cap C) = \frac{1}{36} \neq \frac{1}{18} = P(A)P(C)$

$P(A\cap B\cap C) = \frac{1}{36} = \frac{1}{36} = \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{9} = P(A)P(B)P(C) \Rightarrow \{A,B,C\}$ no son independientes

Variable aleatoria (Espacio muestral finito)

$$X : \Omega \longrightarrow \mathbb{R} \qquad |\Omega| < +\infty$$

Sea $A \subseteq \Omega$

$$\mathbb{1}_A : \Omega \to \mathbb{R}$$
$$\omega \mapsto \mathbb{1}_A(\omega) = \begin{cases} 1 & si \ \omega \in A \\ 0 & si \ \omega \notin A \end{cases} \quad \text{función indicatriz}$$

$\mathbb{1}_{A\cup B}(\omega) = \mathbb{1}_A(\omega) + \mathbb{1}_B(\omega) - \mathbb{1}_{A\cap B}(\omega)$ $\qquad$ $\mathbb{1}_{A\cap B}(\omega) = \mathbb{1}_A(\omega)\cdot \mathbb{1}_B(\omega)$

Observación Como $|\Omega| < +\infty$

$R_X \ \{x = X(\omega) : \omega \in \Omega\}$ $\qquad$ $|R_X| < +\infty$ $\qquad$ $R_X = \{x_1, \dots, x_m\} \subseteq \mathbb{R}$

Sea $B \subseteq \cancel{\Omega} R_X$, se define

$$P_X(B) = P(\{\omega : X(\omega) \in B\}) = P \circ X^{-1}$$

$\forall x_i \in R_X$: $\quad P_X(\{x_i\}) = P(\{\omega : X(\omega) = x_i\})$

$P_X(x_i) \qquad P(\{\omega : X(\omega) = x_i\}) = P(X = x_i)$ ← Abuso de notación

$P_X(B) = P(\{\omega : X(\omega) \in B\}) = P(X \in B)$

$R_X = \{ x = X(\omega) : \omega \in \Omega \}$
$= \{x_1, \ldots, x_m\}$

$P_X(x_i) \quad i = 1, \ldots, m$

1) $P_X(x_i) \geq 0 \quad i=1,\ldots,n$

2) $\sum_{i=1}^{n} P_X(x_i) = 1$

Se define en los números reales una probabilidad (imagen por X)

En el ejemplo ~~binomial~~ Bernoulli $X(\omega) = \begin{cases} 1 & \text{si } \omega \text{ es éxito} \\ 0 & \text{si } \omega \text{ no es éxito (fracaso)} \end{cases}$ $\quad R_X = \{0, 1\}$

$P_X(x) = p^x (1-p)^{1-x} \quad : x = 0, 1$

Función de distribución

Sea $x \in \mathbb{R}$

- $F_X(x) = P(\{\omega : X(\omega) \leq x\})$
$= P(X \leq x)$

- $P_X(x) = \sum_{\{i : x_i \leq x\}} P_X(x_i)$

- $P_X(x_i) = F_X(x_i) - F_X(x_i-)$

Propiedades

1) $F_X(-\infty) = 0, \quad F_X(+\infty) = 1$

2) $F_X(x)$ continua derecha

$F_X(x+) = F_X(x)$

3) F_X es no decreciente

Vectores aleatorios

$X = (X_1, \ldots, X_n) \quad : X_i$ son variables aleatorias

$P_X(x_1, \ldots, x_r) = P(\{\omega : X_1(\omega) = x_1, \ldots, X_r(\omega) = x_r\})$

$F_X(x_1, \ldots, x_r) = P(\{\omega : X_1(\omega) \leq x_1, \ldots, X_r(\omega) \leq x_r\})$

Independencia de variables aleatorias

Definición $X_1, \ldots, X_r$

$P(X_1 = x_1, \ldots, X_r = x_r) = P(X_1 = x_1) \cdots P(X_r = x_r) \quad \forall (x_1, \ldots, x_r) \in R_X$

Esperanza $(\Omega, \mathcal{A}, P) \quad |\Omega| < +\infty$

$X : \Omega \to \mathbb{R} \qquad R_X = \{x_1, \ldots, x_k\}$

Sea $A_i = \{\omega : X(\omega) = x_i\}$. Así
$i = 1, \ldots, k$

$X(\omega) = \sum_{i=1}^{k} x_i \mathbb{1}_{A_i}(\omega)$

$\{A_1, \ldots, A_k\}$ es una partición de Ω. Sea $p_i = P(A_i) = P(X = x_i)$

Definición

$E[X] = \sum_{i=1}^{k} x_i P_X(A_i) = \sum_{i=1}^{k} x_i P_X(x_i)$

Viernes 12 de Noviembre de 2021

1) Modelo Binomial

$\Omega = \{ \omega = (a_1, \dots, a_n) ; a \in \{0,1\} \} \qquad \mathcal{A} = \{ A : A \subseteq \Omega \}$

$p(\omega) = p^{\sum a_i} q^{n - \sum a_i} \quad ; \quad p, q \geq 0, \quad p + q = 1$

Sean $A_k = \{ \omega : a_k = 1 \}$ y $A_k^c = \{ \omega : a_k = 0 \}$

$\mathcal{A}_k = \{ \emptyset, \Omega, A_k, A_k^c \}$ es un álgebra.

$$P(A_k) = \sum_{\{\omega : \omega \in A_k\}} p^{\sum a_j} q^{n - \sum a_j}$$

Si $\omega \in A_k$, $\displaystyle\sum_{i=1}^{N} a_i = 1 + \sum_{i \neq k} a_i$

$P(A_k^c) = 1 - p = q$

$A_k \cap A_\ell = \{ \omega : a_k = 1 \wedge a_\ell = 1 \}$

$P(A_k \cap A_\ell) = p^2$
$P(A_k \cap A_\ell^c) = p \cdot q$
$P(A_k^c \cap A_\ell^c) = q^2$

Se verifica entonces que A_k y A_ℓ son independientes si $k \neq \ell$.

2) Variables aleatorias (finito)

Si X_1, X_2 son variables aleatorias y $f : \mathbb{R}^2 \to \mathbb{R}$, entonces $y = f(x_1, x_2) : \Omega \to \mathbb{R}$ (v.a.)

$Z = X + Y, \quad R_X = \{ x_1, \dots, x_k \} \quad R_Y = \{ y_1, \dots, y_\ell \}$

$R_Z = \{ z = x_i + y_j : i = 1, \dots, k, \ j = 1, \dots, \ell \}$

Sea $z \in R_Z$,

$$p_Z(z) = P(Z = z) = P(X + Y = z)$$
$$= \sum_{\{(i,j) : x_i + y_j = z\}} P(X = x_i, Y = y_j).$$

Si X e Y son independientes, entonces

$$\sum_{\{(i,j) : x_i + y_j = z\}} P(X = x_i) P(Y = y_j)$$
$$= \sum_{\{(i,j) : x_i + y_j = z\}} P_X(x_i) P_Y(y_i)$$
$$= \sum_{i=1}^{k} P_X(x_i) P_Y(z - x_i)$$

$P_Y(z - x_i) = 0 \quad$ si $\quad z - x_i \in R_Y$

Ejemplo. X, Y variables aleatorias Bernoulli independientes e idénticamente distribuidas

$R_X = R_Y = \{0, 1\} \qquad P(X=0) = P(Y=0) = q \qquad P(X=1) = P(Y=1) = p$

$Z = X + Y \quad R_Z = \{0, 1, 2\}$

$P_Z(0) = P_X(0) \cdot P_Y(0-0) + P_Z(1) \cdot P_Z(0-1) = q^2$

$P_Z(1) = P_X(0) \cdot P_Y(1-0) + P_X(1) P_Y(1-1) = 2pq$

$P_Z(2) = P_X(0) \cdot P_Y(2-0) + P_X(1) P_Y(2-1) = p^2$

Por inducción $\quad Z = \sum_{j=1}^{n} x_j \quad x_i$ Bernoulli

$R_Z = \{0, 1, \dots, n\} \qquad P_Z(z) = \binom{n}{z} p^z q^{n-z}, \quad z \in R_Z$

Esperanza.

$X : \Omega \to \mathbb{R} \qquad R_X = \{x_1, \dots, x_k\} \qquad A_i = \{\omega : X(\omega) = x_i\}$

$X(\omega) = \sum_{i=1}^{m} x_i \mathbb{1}_{A_i}(\omega) \qquad p_i = p(A_i) = p(X = x_i)$

$E[X] = \sum_{i=1}^{n} x_i P(A_i)$

Supongamos que existen otros $\{B_j : 1 \leq j \leq \ell\}$, $B_j \subseteq \Omega$ $\quad B_j$ disjuntos

$$X(\omega) = \sum_{j=1}^{\ell} x_j' \mathbb{1}_{B_j}(\omega)$$

$$\sum_{\{j : x_j' = x_i\}} x_j' P(B_j) = x_i \sum_{\{j : x_j' = x_i\}} P(B_j)$$

$$= x_i P(A_i)$$

$$E[X] = \sum_{i=1}^{k} x_i P(A_i)$$

$$= \sum_{i=1}^{k} \sum_{\{j : x_j' = x_i\}} x_j' P(B_j) = \sum_{j=1}^{\ell} x_j' P(B_j)$$

$$\sum_{j=1}^{\ell} x_j' P(B_j) = EX = \sum_{j=1}^{k} x_i P(A_i)$$

Lunes 15 de noviembre de 2021.

Sea $(\Omega, \mathcal{A}, P) : |\Omega| < +\infty \qquad X : \Omega \to \mathbb{R}$

$R_X = \{x_1, \dots, x_m\} \subseteq \mathbb{R}. \qquad P(A_i) = P(X = x_i)$

$A_i = \{\omega : X(\omega) = x_i\}$

$$E[X] = \sum_{i=1}^{m} x_i P(A_i) = \sum_{i=1}^{m} x_i P(X = x_i)$$

$$\Delta F_X(x) = F(x) - F(x_i)$$

$$E[X] = \sum_{i=1}^{m} x_i \Delta F_X(x_i)$$

- **Propiedades**

1) Si $X \geq 0$, entonces $E[X] \geq 0$
2) $E(aX + bY) = a E[X] + b E[Y]$

3. $X \geq Y$, entonces $E[X] \geq E[Y]$
4. $|E[X]| \leq E[|X|]$
5. Si X y Y son independientes, entonces $E[XY] = E[X] \cdot E[Y]$
6. $(E[|XY|])^2 \leq (E[X])^2 (E[Y])^2$ (Cauchy - Bunyakovsky - Schwartz)
7. Si $X = \mathbb{1}_A$, $E[X] = P(A)$

Observación
Todas las propiedades se demuestran fácilmente y se usa la suma finita.

Transformación de variables aleatorias

$$\left.\begin{array}{l} X: \Omega \to \mathbb{R} \\ Y: \mathbb{R} \to \mathbb{R} \end{array}\right\} \Rightarrow Y = \varphi(X): \Omega \to \mathbb{R}$$

Se puede escribir $R_X = \{x_1, \dots, x_k\}$

$$X = \sum_i x_i \mathbb{1}_{A_i} \qquad \text{con} \quad A_i = \{\omega : X(\omega) = x_i\}$$

Sea $B_j = \{\omega : Y(\omega) = \varphi(X(\omega)) = y_j\}$

$\varphi(X(\omega))$ completar.

$$E[Y] = \sum_j y_j P(B_j) = \sum_j P_Y(y_j)$$

Por otro lado $Y(\omega) = \varphi(X(\omega)) = \sum_i \varphi(x_i) \mathbb{1}_{A_i}(\omega)$

y por lo tanto

$$E[Y] = E[\varphi(X)] = \sum_i \varphi(x_i) P(A_i)$$
$$= \sum_i \varphi(x_i) P_X(x_i)$$

Varianza.
$V(X) = E((X - EX)^2)$

Propiedades.

1) $V(X) = EX^2 - (\ \ )^2$
2) $V(aX+b) = a^2 V(X)$

Covarianza

$$Cov(X, Y) = E[(X - EX)(Y - EY)]$$
$$= E[XY] - EX \cdot EY.$$

Esquema de Bernoulli

$(\Omega_n, \mathcal{A}_n, P_n) = \Omega_n = \{\omega : \omega = (a_1, \dots, a_n),\ a_i \in \{0,1\}\}$

$\mathcal{A}_n = \{A : A \subseteq \Omega_n\}$ $\qquad P_n(\{\omega\}) = p^{\sum a_i} q^{1 - \sum a_i}$ $\qquad (q = 1 - p)$

$X_{n_i}(\omega) = a_i \qquad i = 1, \dots, n$

$P(X_{n_i} = 1) = p \qquad P(X_{n_i} = 0) = q \qquad i = 1, \dots, n$

$S_{n_0}(\omega) \equiv 0$

$$S_{nk}(\omega) = S_{nk-1}(\omega) + X_{nk}(\omega)$$
$$= \sum_{j=1}^{k}{}' X_{nj}(\omega) \qquad k=1,\dots,n$$

Sean $S_n = S_{nn}$

$E S_n = np$ y en consecuencia $E\left(\frac{S_n}{n}\right) = p$

¿Qué tan grande es el conjunto?

$$\left\{\omega : \left|\frac{S_n(\omega)}{n} - p\right| \leq \varepsilon\right\}$$

Para n suficientemente grande.

Observemos que $0 < p < 1$

$$P\left(\frac{S_n}{n} = 1\right) = P(X_{n_1} = 1, \dots, X_{nn} = 1) = p^n$$

$$P\left(\frac{S_n}{n} = 0\right) = P(X_{n_1} = 0, \dots, X_{nn} = 0) = q^n$$

Podemos observar que sucede con

$$\left\{\omega : \left|\frac{S_n(\omega)}{n} - p\right| \geq \varepsilon\right\}$$

$$P_n\left(\left\{\omega \,\middle|\, \left|\frac{S_n}{n} - p\right| > \varepsilon\right\}\right) = \sum_{\left\{k : \left|\frac{k}{n} - p\right| > \varepsilon\right\}}{}' P_n(k)$$

Lunes 22 de noviembre de 2021.

Probabilidad condicional e independencia.

Definición 1 - Probabilidad condicional-
Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad discreto y sean $A, B \in \mathcal{A}$ tales que $P(A) > 0$, se define la probabilidad condicional del evento B dado el evento A por

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

donde $P(A) = |A|/|\Omega|$ y $P(A \cap B) = |A \cap B|/|\Omega|$.

De la definición 1, se siguen las siguientes propiedades

- $P(A|A) = 1$
- $P(B|A) = 1$ si $A \subseteq B$
- $P(\emptyset|A) = 0$
- Si $B_1 \cap B_2 = \emptyset$, entonces $P(B_1 \cup B_2 | A) = P(B_1|A) + P(B_2|A)$.

Además, tenemos que.

$$P(B|A) + P(B^c|A) = 1$$

Aunque en general, NO se tiene que

$$P(B|A) + P(B|A^c) = 1 \qquad \text{o} \qquad P(B|A) + P(B^c|A^c) = 1.$$

Por otro lado, si consideramos $\mathcal{D} = \{B_1, \ldots, B_n\}$ una partición de Ω tal que $P(B_i) > 0$, para todo $i = 1, \ldots, n$ (dado que es una partición, los conjuntos son disjuntos dos a dos) y A un evento, entonces

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n).$$

Así, se sigue que

$$P(A) = P\left(\bigcup_{i=1}^{n} (A \cap B_i) \right) = \sum_{i=1}^{n} P(A \cap B_i)$$

pero, como

$$P(A \cap B_i) = P(A|B_i) \cdot P(B_i),$$

entonces

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i). \qquad (1)$$

La fórmula (1) se conoce como fórmula de la probabilidad total.
En particular, si $0 < P(B) < 1$, se tiene que

$$P(B) = P(A|B)\,P(B) + P(A|\bar{B})\,P(\bar{A}).$$

Fórmula para la multiplicación de probabilidades.
Es posible generalizar la fórmula

$$P(A \cap B) = P(B|A) \cdot P(A) \qquad (2)$$

por inducción, de la siguiente manera: si $A_1, \ldots, A_{n-1}$ son eventos con $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$, entonces

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)\,P(A_2|A_1)\,P(A_3|(A_1 \cap A_2)) \cdots P(A_n|(A_1 \cap \cdots A_{n-1}))$$

- Supongamos que A y B son dos eventos con $P(A) > 0$ y $P(B) > 0$, entonces usando la fórmula (2), se tiene que

$$P(A \cap B) = P(A|B)\,P(B). \qquad (3)$$

Por lo tanto, de (2) y (3), se deduce que

$$P(B|A)\,P(A) = P(A|B)\,P(B) \implies P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \qquad (4)$$

la fórmula (4) se conoce como la fórmula de Bayes.

Finalmente, si los eventos $A_1, \ldots, A_n$ forman una partición de Ω, entonces de (1) y (4), se tiene que

$$P(A_i|B) = \frac{P(B|A_i)\,P(A_i)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)} \qquad (5)$$

Independencia

En cierto sentido, el concepto de independencia juega un papel fundamental en la teoría de probabilidades pues nos permite distinguirla de la teoría general de la medida.

Así, si A y B son dos eventos, es natural hablar de independencia de B si A si sabemos que la ocurrencia de A no afecta la probabilidad de B

es decir, si

$$P(B|A)=P(B) \tag{6}$$

(note que se supone $P(A)>0$). Así, por la fórmula (2), se concluye que

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

Definición 2. - Independencia -
Sean $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad discreto (finito), y A y B dos eventos. Se dice que A y B son independientes si

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

- Dadas dos álgebras $\mathcal{A}_1$ y $\mathcal{A}_2$, se dicen independientes si para cada par de conjuntos $A_1\in\mathcal{A}_1$ y $A_2\in\mathcal{A}_2$, estos son independientes, es decir

$$P(A_1\cap A_2)=P(A_1)P(A_2)$$

- Una familia de eventos $A_1,\dots,A_n$ se dice independiente si para cada $k=1,\dots,n$ y $1\leq i_1\leq i_2\leq\dots\leq i_k\leq n$

$$P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$$

- Una familia de álgebras $\{\mathcal{A}_1,\mathcal{A}_2,\dots\mathcal{A}_n\}$ es independiente si para toda familia de eventos $\{A_1,\dots,A_n\}$ con $A_i\in\mathcal{A}_i$ con $i=1,\dots,n$, esta es independiente.

Miércoles 24 de Noviembre de 2021

Sección 2: Variables Aleatorias

Definición 3. Cualquier función numérica X definida sobre un espacio muestral finito se denomina variable aleatoria.

$$X:\Omega\longrightarrow\mathbb{R}$$
$$\omega\longmapsto X(\omega)$$

Más adelante se exigirá una estructura de medibilidad, pero para el caso actual esto es trivial

Ejemplo: Un ejemplo importante es, dado $A\subseteq\Omega$.

$$\mathbb{1}_A:\Omega\longrightarrow\mathbb{R}$$
$$\omega\longmapsto\mathbb{1}_A(\omega)=\begin{cases}1 & \text{si } \omega\in A\\ 0 & \text{si } \omega\notin A.\end{cases}$$

Definición 4 - Función de Distribución - Sean $x\in\mathbb{R}$, ~~la función~~ $(X,\mathcal{A},P)$ un espacio de probabilidad finito, se define la función de distribución por

$$F_X(x) = P(\{\omega: X(\omega)\leq x\}) = P(X\leq x)$$

donde X es una variable aleatoria.

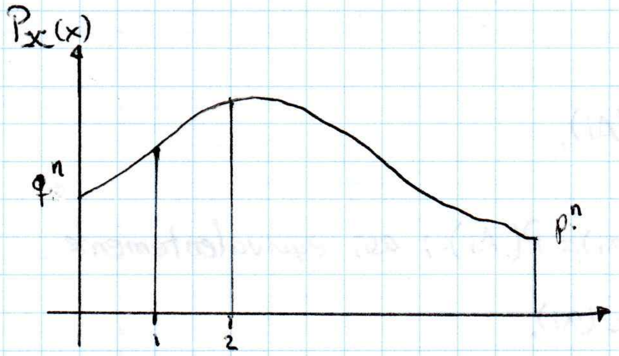
Se observa lo siguiente

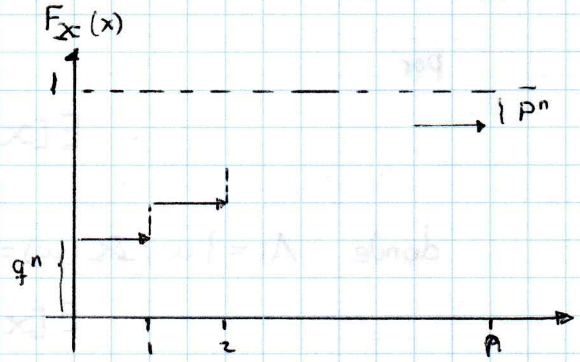
1) $F_X(x)=\sum_{\{i:\, x_i\leq x\}} P_X(x_i)$, donde $P_X(x_i) = P(\{\omega: X(\omega)=x_i\})$

2) $P_X(x_i) = F_X(x_i)-F_X(x_i-)$ donde $F_X(x_i-) = \lim_{y\uparrow x_i} F_X(y)$ con $y\uparrow x_i \Leftrightarrow y\leq x_i$ $t\to x_i$

3) $F_X(-\infty) = 0$ y $F_X(+\infty) = 1$
4) $F_X(x)$ es constante a trozos y continua a la derecha ($F_X(x^+) = F_X(x)$)

Para una variable aleatoria binomial, se tiene los siguientes diagramas





Vectores aleatorios.

En algunos casos, es necesario considerar también vectores aleatorios, es decir, podemos considerar un vector de variables aleatorias

$$X = (X_1, \dots, X_r) \qquad \text{con } X_i \ \ i = 1, \dots, r \text{ variables aleatorias.}$$

En este sentido se define al conjunto de probabilidades

$$P_X(x_1, \dots, x_r) = P(\{\omega : X_1(\omega) = x_1, \dots, X_r(\omega) = x_r\}),$$

como la distribución de probabilidades del vector aleatorio X y a la función

$$F_X(x_1, \dots, x_r) = P(\{\omega : X_1(\omega) \le x_1, \dots, X_r(\omega) \le x_r\})$$

con $x_i \in \mathbb{R}$ para $i = 1, \dots, r$, se la denomina función de distribución aleatoria del vector X del vector aleatorio

Definición 4. - Independencia de vectores - Dadas las variables aleatorias $X_1, \dots, X_r$, estas se dicen independientes si

$$P(\{X_1 = x_1, \dots, X_r = x_r\}) = P(\{X_1 = x_1\}) \cdots P(\{X_r = x_r\})$$

para $x_i \in \mathbb{R}$ con $i \in 1, \dots, r$ o equivalentemente

$$P(\{X_1 \in B_1, \dots, X_r \in B_r\}) = P(\{X_1 \in B_1\}) \cdots P(\{X_r \in B_r\})$$

para todo $B_1, \dots, B_r \in \mathcal{B}$ co $\mathcal{B}$ el conjunto de los Borelianos de $\mathbb{R}$.

Esperanza

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, X una variable aleatoria con valores $X = \{x_1, \dots x_k\}$, si definimos los conjuntos

$$A_i = \{\omega : X(\omega) = x_i\} \qquad i = 1, \dots, k,$$

entonces

$$X(\omega) = \sum_{i=1}^{k} x_i \, \mathbb{1}_{A_i}(\omega)$$

donde los conjuntos $\{A_1, \dots, A_k\}$ forman una partición de Ω.

Definición 5. Sea $(X, \mathcal{A}, P)$ un espacio de probabilidad finito, X una variable aleatoria,
- Esperanza - se define la esperanza o el valor medio de la variable

$$X = \sum_{k=1}^{k} x_i \qquad x_i \in \{x_1, \dots, x_k\} = X$$

por

$$E[X] = \sum_{i=1}^{k} x_i P(A_i),$$

donde $A_i = \{\omega : X(\omega) = x_i\}$ y $P_X(x_i) = P(A_i)$; así, equivalentemente

$$E[X] = \sum_{k=1}^{k} x_i P_X(x_i).$$

Si recordamos que $F_X = F_X(x_i)$ y como

$$\Delta F_X(x) = F_X(x) - F_X(x-),$$

entonces como $P_X(x) = \Delta F_X(x)$, entonces se sigue que

$$E[X] = \sum_{k=1}^{k} x_i \Delta F_X(x_i)$$

Propiedades de la esperanza.

1) Si $X \geq 0$, entonces $E[X] \geq 0$.
2) $E[aX + bY] = aE[X] + bE[Y]$ con a, b constantes
3) Si $X \geq Y$, entonces $E[X] \geq E[Y]$
4) $E[|X|] \geq |E[X]|$.
5) Si X y Y son independientes, $E[X \cdot Y] = E[X] \cdot E[Y]$
6) $E[X^2] \cdot E[Y^2] \geq (E[XY])^2$ Desigualdad de Cauchy-Schwarz.
7) Si $X = \mathbb{1}_A$, $E[X] = P(A)$

La demostración de estas propiedades está en Probability I - Shiryaev, A. (3th edito) (pág 36)

Transformación de variables

Sea $X = \sum_i x_i \mathbb{1}_{A_i}(x_i)$ donde $A_i = \{\omega : X(\omega) = x_i\}$ y sea $\varphi = \varphi(X(\omega))$ una función de $X(\omega)$, si $B_j = \{\omega : \varphi(X(\omega)) = y_i\}$, entonces

$$\varphi(X(\omega)) = \sum_j y_i \mathbb{1}_{B_j}(\omega)$$

y, consecuentemente

$$E[\varphi] = \sum_j y_i P(B_j) = \sum_j y_i P_\varphi(y_i)$$

pero además, es claro que

$$\varphi(X(\omega)) = \sum_i \varphi(x_i) \mathbb{1}_{A_i}(\omega)),$$

Así, se concluye que, con $Y = \varphi$

$$E[Y] = E[\varphi(X)] = \sum_i \varphi(x_i) P_X(x_i)$$

Definición 6 **-Varianza-** Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y X una variable aleatoria, se define la varianza de X, denotada $Var(X)$ por

$$Var(X) = E(X - E[X])^2 \tag{7}$$

Una noción importante sobre la varianza es que esta indica la cantidad de dispersión de los valores de X alrededor de $E[X]$.

El número $\sigma = +\sqrt{Var(X)}$ se conoce como la desviación estándar (de X del valor medio $E[X]$)

Propiedades.

1) $Var[X] = E[X]^2 + (E[X])^2$
2) $Var[a+bX] = b^2 Var[X]$ con a y b constantes.

Covarianza Dadas dos variables aleatorias, se define la covarianza por

$$Cov[X, Y) = E[(X - E[X])\,(Y - E[Y])]. \tag{8}$$

Si $Var[X] > 0$ y $Var[Y] > 0$, entonces se define el coeficiente de correlación como

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var[X] \cdot Var[Y]}} \tag{9}$$

Si las variable X y Y son independientes entonces

$$Var(X + Y) = Var(X) + Var(Y)$$

Teorema 1.
Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad discreto y X una variable aleatoria no negativa, entonces

$$P(X \geq \varepsilon) \leq \frac{E[X]}{\varepsilon} \quad \text{para todo } \varepsilon > 0 \tag{10}$$

Demostración:
Notemos que

$$X = X\,\mathbb{1}_{[\varepsilon, +\infty[}(X) + X\,\mathbb{1}_{]0,\varepsilon[}(X) \geq X\,\mathbb{1}_{[\varepsilon, +\infty[}(X) \geq \varepsilon\,\mathbb{1}_{[\varepsilon, +\infty[}(X)$$

así, por las propiedades de la esperanza (7)

$$E[X] \geq \varepsilon P(X \geq \varepsilon).$$

Corolario Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad finito, para todo $\varepsilon > 0$

1) $P(|X| \geq \varepsilon) \leq E[|X|]/\varepsilon$
2) $P(|X| \geq \varepsilon) = P(X^2 \geq \varepsilon^2) \leq E[X^2]/\varepsilon^2$
3) $P\{|X - E[X]| \geq \varepsilon\} \leq (Var\ X)/\varepsilon^2$
4) $P(|X - E[X]| / \sqrt{Var[X]} \geq \varepsilon) < 1/\varepsilon^2$

Esquema: Ley de los grandes números

Consideremos el siguiente esquema denominado esquema de Bernoulli, sea $(\Omega_n, \mathcal{A}_n, P_n)$ con

$$\Omega_n = \{\omega : \omega = (a_1, \ldots, a_n) \quad a_i = 0, 1,\ i = 1, \ldots n\} \quad \mathcal{A}_n = \{A : A \subseteq \Omega_n\}$$

$$P_n(\{\omega\}) = p^{\sum a_i} q^{n - \sum a_i} \qquad q = 1 - p.$$

Definamos las variables aleatorias $X_{n1},\dots,X_{nn}$ donde $X_{ni}(\omega)=a_i$ $i=1,\dots,n$ donde $\omega=(a_1,\dots,a_n)$. Se sabe que las variables X_{ni} son independientes e identicamente distribuidas

$$P_n(X_{ni}=1)=p \qquad P_n(X_{ni}=0)=q \qquad i=1,\dots,n.$$

La variable X_{ni} describe el resultado del experimento en la i-ésima etapa.

Definamos $S_{n0}(\omega)=0$ y

$$S_{nk}=X_{n1}+\dots+X_{nk} \qquad k=1,\dots,n.$$

Así, se sigue que, con $S_n=S_{nn}$

$$E_n S_n=np \Rightarrow E_n \frac{S_n}{n}=p$$

En otras palabras, la frecuencia de éxitos S_n/n, coincide con la probabilidad de éxito. Así, podemos preguntarnos, cuanto difiere la frecuencia S_n/p de la probabilidad p

- Una primera aproximación se da cuando intentamos un $\varepsilon>0$ suficientemente pequeño y para n suficientemente grande, en ese caso no podemos esperar que la desviación de S_n/n de p sea menor que ε.

- Así, consideremos $0<p<1$; así, tenemos que

$$P\left(\frac{S_n}{n}=1\right)=P(X_1=1,\dots,X_n=1)=p^n$$

y

$$P\left(\frac{S_n}{n}=0\right)=P(X_1=0,\dots,X_n=0)=q^n.$$

Por otro lado, uno puede demostrar que

$$P_n\left(\left|\frac{S_n}{n}-p\right|\geq\varepsilon\right)\to 0 \quad \text{cuando } n\to+\infty.$$

Para ello, usando el teorema 1, se tiene que: Tomando $X=S_n/n$, y como $Var\, S_n=npq$, entonces

$$P\left(\left|\frac{S_n}{n}-p\right|\geq\varepsilon\right)\leq\frac{Var(S_n/n)}{\varepsilon^2}=\frac{Var\,S_n}{n^2\varepsilon^2}=\frac{npq}{n^2\varepsilon^2}=\frac{pq}{n\varepsilon^2}\leq\frac{1}{4n\varepsilon^2}\xrightarrow[n\to+\infty]{}0$$

Capítulo 2: Fundamentos de la teoría de probabilidades.

Definición 7.

Sea Ω un conjunto de resultados posibles de un experimento aleatorio y $\mathcal{P}(\Omega)$ el conjunto potencia, definimos los siguientes términos

1) Un álgebra es una clase $\mathcal{A}\subseteq\mathcal{P}(\Omega)$ tal que

- $\Omega\in\mathcal{A}$
- Si $A,B\in\mathcal{A}$, entonces $A\cup B\in\mathcal{A}$ y $A\cap B\in\mathcal{A}$
- Si $A\in\mathcal{A}$, entonces $A^c\in\mathcal{A}$

2) Una medida finitamente aditiva es una aplicación $\mu:\mathcal{A}\to[0,+\infty]$ tal que

- Si $A\cap B=\emptyset$, entonces $\mu(A\cup B)=\mu(A)+\mu(B)$.
- $\mu(\Omega)\leq+\infty$.

3) Una σ-álgebra $\mathcal{F}$ es un álgebra que satisface:
Si $A_n \in \mathcal{F}$, $n=1,\dots$, entonces

$$\bigcup_n A_n \in \mathcal{F} \quad y \quad \bigcap_n A_n \in \mathcal{F}$$

a la dupla $(\Omega, \mathcal{F})$ se le denomina espacio medible.

4) Una medida finitamente aditiva es contablemente aditiva o simplemente medida si para toda sucesión de elementos disjuntos $\{A_n\}$ con $\{A_i \cap A_j = \emptyset \quad i \neq j\}$, entonces

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n)$$

5) Una medida se dice σ-finita si existe una sucesión de eventos $\Omega_n \in \mathcal{F}$, $n \geq 1$ tales que $\Omega_i \cap \Omega_j = \emptyset$ si $i \neq j$ y $\bigcup_n \Omega_n = \Omega$ tales que

$$\mu(\Omega_n) < +\infty \qquad \forall n$$

6) Una medida contablemente aditiva P sobre una σ-álgebra que satisface $P(\Omega)=1$ se denomina medida de probabilidad o probabilidad.

A la tripleta $(\Omega, \mathcal{A}, P)$ se la denomina espacio de probabilidad o probabilizado.

Propiedades de las medidas de probabilidad

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad
1) $P(\emptyset) = 0$.
2) Si $A, B \in \mathcal{A}$, entonces $P(A\cup B) = P(A) + P(B) - P(A\cap B)$.
3) Si $A, B \in \mathcal{A}$ y $B \subseteq A$, entonces $P(B) \leq P(A)$.
4) Si $A \in \mathcal{A}$, entonces $P(A^c) = 1 - P(A)$
5) Si $A_n \in \mathcal{A}$ con $n\in\mathbb{N}$ y $\bigcup A_n \in \mathcal{A}$, entonces

$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} P(A_n)$$

Consideremos la familia $\{B_n\}_n \in \mathcal{A}$, con

$$B_0 = A_0 \quad y \quad B_n = A_n \setminus \bigcup_{i=0}^{n-1} A_i$$

Así, se tiene que $B_i \subseteq A_i$ para todo i y por lo tanto

$$\mu(B_i) \leq \mu(A_i)$$

de donde como la familia $\{B_n\}_n$ es disjunta, entonces

$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = P\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \sum_{n\in\mathbb{N}} \mu(B_n) \leq \sum_{n\in\mathbb{N}} \mu(A_n).$$

Teorema. Sean $(\Omega, \mathcal{A})$ un espacio medible (probabilizable), P una medida finitamente aditiva sobre $\mathcal{A}$ con $P(\Omega)=1$. Son equivalentes las siguientes afirmaciones:

1) P es σ-aditiva (i.e. P es una probabilidad) (σ-aditiva = contablemente aditiva)
2) P es continua inferiormente, es decir que si $A_n \in \mathcal{A}$, $n \geq 1$ tal que $A_n \subseteq A_{n+1}$ (sucesión creciente), entonces

$$\lim_{n\to+\infty} P(A_n) = P\left(\bigcup_{n=1}^{+\infty} A_n\right)$$

3) P es continua superiormente, es decir para $A_n \in \mathcal{A}$ con $n \in \mathbb{N}$ con $A_n \supseteq A_{n+1}$ (sucesión decreciente), entonces

Espacios Lebesgue
Lema 2.2.3
Vol 1.
Diego Chamorro.

$$\lim_{n\to+\infty} P(A_n) = P\left(\bigcap_{i=1}^{+\infty} A_n\right)$$

4) P es continua en $\emptyset$, es decir, para $A_n \in \mathcal{A}$ con $n \in \mathbb{N}$ tal que $A_n \supseteq A_{n+1}$ con $\bigcap_{n=1}^{\infty} A_n = \emptyset$, entonces

$$\lim_{n\to+\infty} P(A_n) = 0.$$

Demostración

$(1) \Rightarrow (2)$. Notemos que como la familia (A_n) es creciente, entonces

$$\bigcup_{n\in\mathbb{N}} A_n = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots$$

Así, obtenemos una unión disjunta de conjuntos y por lo tanto

$$\begin{aligned} P\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= P\left(A_1 \cup (A_2\setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots\right) \\ &= P(A_1) + P(A_2\setminus A_1) + P(A_3 \setminus A_2) + \cdots \\ &= P(\cancel{A_1}) + P(\cancel{A_2}) - P(\cancel{A_1}) + P(\cancel{A_3}) - P(\cancel{A_2}) + \cdots \\ &= \lim_{n\to+\infty} P(A_n) \end{aligned}$$

$(2) \Rightarrow (3)$ Sea $n \in \mathbb{N}$, entonces

$$P(A_n) = P(A_1 \setminus (A_1 \setminus A_n)) = P(A_1) - P(A_1 \setminus A_n).$$

además, la sucesión $\{A_1 \setminus A_n\}_{n\in\mathbb{N}}$ es creciente y como

$$\bigcup_{n\in\mathbb{N}} (A_1 \setminus A_n) = \bigcup_{n\in\mathbb{N}} A_1 \cap A_n^c = A_1 \cap \left(\bigcup_{n\in\mathbb{N}} A_n^c\right) = A_1 \setminus \left(\bigcup_{n\in\mathbb{N}} A_n^c\right)^c = A_1 \setminus \bigcap_{n\in\mathbb{N}} A_n.$$

Así, por la propiedad anterior, se sigue que

$$\lim_{n\to+\infty} P(A_1 \setminus A_n) = P\left(\bigcup_{n\in\mathbb{N}} A_1 \setminus A_n\right).$$

Finalmente, como

$$\begin{aligned} \lim_{n\to+\infty} P(A_n) = P(A_1) - \lim_{n\to+\infty} P(A_1 \setminus A_n) &= P(A_1) - P\left(\bigcup_{n\in\mathbb{N}} A_1 \setminus A_n\right) \\ &= P(A_1) - P\left(A_1 \setminus \bigcap_{n\in\mathbb{N}} A_n\right) \\ &= P(A_1) - P(A_1) + P\left(\bigcap_{n\in\mathbb{N}} A_n\right) \\ &= P\left(\bigcap_{n\in\mathbb{N}} A_n\right) \end{aligned}$$

se concluye lo requerido.

(3) $\Rightarrow$ (4) Inmediato

(4) $\Rightarrow$ (1) Sea $(A_n) \in \mathcal{A}$ una familia de conjuntos dos a dos. Notemos que,

$$P\left(\bigcup_{i=1}^{+\infty} A_i\right) = P\left(\bigcup_{i=1}^{n} A_i\right) + P\left(\bigcup_{i=n+1}^{+\infty} A_i\right),$$

de donde, tomando $(B_n) = \left(\bigcup_{i=n+1}^{+\infty} A_n\right)$, sabemos que (B_n) es decreciente y ademas $B_n \to \emptyset$ cuando $n \to +\infty$, entonces

$$\sum_{i=1}^{+\infty} P(A_i) = \lim_{n\to+\infty} \sum_{i=1}^{n} P(A_i) = \lim_{n\to+\infty} P\left(\bigcup_{i=1}^{n} A_i\right)$$

$$= \lim_{n\to+\infty}\left[P\left(\bigcup_{i=1}^{+\infty} A_i\right) - P\left(\bigcup_{i=n+1}^{+\infty} A_i\right)\right]$$

$$= P\left(\bigcup_{i=1}^{+\infty} A_i\right) - \lim_{n\to+\infty} P\left(\bigcup_{i=n+1}^{+\infty} A_i\right) = P\left(\bigcup_{i=1}^{+\infty} A_i\right)$$

lo que concluye lo requerido.

Proposición 3. Sean μ una medida σ-finita sobre un espacio $(\Omega, \mathcal{A})$, $\{A_n\}_n \in \mathcal{A}$ y $A = \lim_{n\to+\infty} A_n$, es decir

$$A = \lim_{n\to+\infty} \inf A_n = \lim_{n\to+\infty} \sup A_n,$$

entonces

$$\mu(A) = \lim_{n\to+\infty} \mu(A_n)$$

Sección 2.1: Funciones de distribución

Definición 9. Una función $F: \mathbb{R} \to \mathbb{R}$ se dice función de distribución si

- es no decreciente
- es continua a la derecha
- $F(-\infty) = 0$, $F(+\infty) = 1$

Notemos que una función de distribución tiene a lo sumo una cantidad numerable de saltos, En efecto, sea a_j el conjunto de saltos

$$F(a_j) - F(a_j-) = b_j$$

si el conjunto no fuera numerable, no se puede tener que $\sum_j b_j \leq 1$.

Sea $F_d(x) = \sum_j b_j S_{a_j}(x)$, donde

$$S_a(x) = \begin{cases} 0 & \text{si } x < a \\ 1 & \text{si } x \geq a. \end{cases}$$

entonces se tiene que $F_d(-\infty) = 0$ y $F_d(\infty) = \sum_j b_j \leq 1$. Observe que $F = F_c + F_d$ es una descomposición única con F_c continua.

Definición 10 Una función F es absolutamente continua en $\mathbb{R}$ con respecto a la medida de Lesbegue si existe una función $f \in L^1$ tal que para todo $x < x'$ se tiene que

$$F(x') - F(x) = \int_x^{x'} f(x)\,dx$$

Proposición 9: Si F es una función de distribución absolutamente continua, entonces existe $f \in L^1$, tal que

- $f \geq 0$ en casi todas partes
- $\int_{-\infty}^{+\infty} f(x)\,dx = 1$
- Para todo x, $F(x) = \int_{-\infty}^{x} f(y)\,dy = 1$.

A esta función se la denomina densidad de probabilidad.

Definición 10: Una función se denomina singular si no es idénticamente cero y F' existe y es igual a 0 en casi todo punto.

Los funciones de distribución discretas son singulares, pero también existen funciones singulares continuas, por ejemplo la escalera de Cantor.

Teorema 5 Sea F una función acotada no decreciente con $F(-\infty)=0$, y denotemos F' su derivada donde exista, entonces

1) $S = \{x : F'(x) \text{ existe y } 0 \leq F'(x) < +\infty\}$ entonces $m(S^c)=0$ donde (m es la medida de Lesbegue).

Ver Lema 2.2
• Probability
A Graduate course
Allan Gunt

2) $F' \in L^1$ y si $x < x'$

$$\int_x^{x'} F'(y)\,dy \leq F'(x) - F(x)$$

3) Se define $F_{ac}(x) = \int_{-\infty}^{x} F'(t)\,dt$ y $F_s(x) = F(x) - F_{ac}(x)$, entonces F_s es singular y la descomposición es única en casi todo punto.

Teo. (2.4) [2]
pág (37)

4) Una función de distribución puede escribirse como una combinación convexa (única para casi todo punto) de una discreta, una singular continua y una continua.

Existen dos resultados que determinan una relación biunívoca entre funciones de distribución y medidas de probabilidad.

Lema. Cada medida de probabilidad μ en $(\mathbb{R}, \mathcal{B})$ determina una función de distribución

Lema
pág 35
[3].

$$\forall x \in \mathbb{R} \qquad F(x) = \mu(]-\infty, x])$$

Demostración:

Denotemos por $I_x =]-\infty, x]$, para todo $x \in \mathbb{R}$. Así, es claro que $I_x \in Bor(\mathbb{R})$, así, la medida

$$\mu(I_x) = \mu(]-\infty, x])$$

está definida. Así, definamos

$$F(x) = \mu(I_x).$$

Así, debemos probar que F satisface las condiciones de la (definición 9).

• Puesto que si $x \leq y$, entonces $I_x \subseteq I_y$, como μ es una medida de probabilidad, entonces

$$\mu(I_x) \leq \mu(I_y)$$

 dado que x, y fueron arbitrarios, se concluye que F es no decreciente.

Sea $(x_n)_n$ una sucesión decreciente, cualquiera tal que $x_n \to x$. Así, se sigue que

$$I_{x_n} \to I_x \qquad y \qquad I_{x_{n+1}} \subseteq I_{x_n}$$

así, por monotonía de la medida, se sigue que

$$\mu(I_{x_n}) \to \mu(I_x)$$

y, más aún $\lim_{I_{x_n} \downarrow I_x} \mu(I_{x_n}) = \mu(I_x) = F(x)$. Por lo tanto, F es continua por derecha.

Finalmente, tomando $(x_n)_n$ tal que $x_n \to -\infty$ con $(x_n)_n$ decreciente, entonces

$$I_{x_n} \downarrow I_x$$

y, más aún, tomando $(x_n)_n \downarrow x$ y $(x_n)_n \uparrow x$ respectivamente, se sigue que

- $F(-\infty) = \lim_{x_n \downarrow -\infty} F(x) = \lim_{x_n \to -\infty} \mu(I_{x_n}) = \mu\left(\bigcap_{n\in\mathbb{N}} I_{x_n}\right) = \mu(\emptyset) = 0.$
- $F(+\infty) = \lim_{x_n \uparrow +\infty} F(x) = \lim_{x_n \uparrow +\infty} \mu(I_{x_n}) = \mu\left(\bigcup_{n\in\mathbb{N}} I_{x_n}\right) = \mu(\Omega) = 1.$

Así, se concluye que μ define una función de distribución.

Más información en teorema 1, pág 185 [3]

Teorema 6.- Cada función de distribución F determina una medida de probabilidad μ en $(\mathbb{R}, \mathcal{B})$ a través de la ecuación

[3] pág 185

$$\forall x \in \mathbb{R}, \qquad F(x) = \mu(]-\infty, x]) \qquad (11)$$

Teorema 7 - Teorema de Carathédory- Sea Ω un espacio, $\mathcal{A}$ un álgebra y $\mathcal{B} = \sigma(\mathcal{A})$ ~~ala~~ la σ-álgebra más pequeña que contiene a $\mathcal{A}$. Sea μ_0 a σ-finita
[3] pág 186
y σ-aditiva medida en $(\Omega, \mathcal{A})$, entonces existe una única medida μ en $(\Omega, \sigma(\mathcal{A}))$ la cual es una extensión de μ_0 y satisface que

$$\mu(A) = \mu_0(A) \qquad \forall A \in \mathcal{A}$$

Sección 2: Variables aleatorias. Esperanza

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad

Definición 11 - variable aleatoria- Una variable aleatoria X es una función medible sobre los números reales con los borelianos $(Bor(\mathbb{R}), \mathcal{B})$

$$X: (\Omega, \mathcal{F}) \to (\mathbb{R}, Bor(\mathbb{R}))$$

es decir $\{\omega : X(\omega) \in B\} \in \mathcal{F}$ para todo $B \in \mathcal{B}$, equivalentemente

$$X^{-1}(\mathcal{B}) \subseteq \mathcal{F}.$$

Teorema 8. Cada variable aleatoria sobre $(\Omega, \mathcal{A}, P)$ induce un espacio de probabilidad $(\mathbb{R}, \mathcal{B}, P_X)$ por medio de la correspondencia

teo 3.1.3 [3]

$$\forall B \in Bor(\mathbb{R}), \quad P_X(B) = P(X^{-1}(B)) = P(\{\omega : X(\omega) \in B\}) \qquad (12)$$

Se tiene entonces que puede escribirse $P_X = P \circ X^{-1}$!

Demostración:
Por definición de P_X, se tiene que

$$P_X(b) = P(X^{-1}(B)) \geq 0.$$

Ahora, sea $\{B_n\}_n \in \mathcal{B}$ una familia disjunta, de donde

$$P_X\left(\bigcup_n B_n\right) = P\left(X^{-1}\left(\bigcup_n B_n\right)\right) = P\left(\bigcup_n X^{-1}(B_n)\right) = \sum_n P(X^{-1}(B_n)) = \sum_n P_X(B_n)$$

Finalmente, como

$$P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1 \quad y \quad P_X(\emptyset) = P(X^{-1}(\emptyset)) = P(\emptyset) = 0,$$

entonces, P_X define una medida de probabilidades.

Notemos la familia $\{X^{-1}(S) : S \subseteq \mathbb{R}\}$ forman una clase monótona para toda función P. En partular, si X es una variable aleatoria, es la clase monótona más pequeña de $\mathcal{A}$ que contiene los conjuntos de la forma

$$\{\omega : X(\omega) \leq x\},$$

donde $x \in \mathbb{R}$. Así, podemos escribir

$$P_X = P \circ X^{-1},$$

y a P_X la llamamos la medida de distribución de probabilidad y es una medida de probabilidad asociada la la función de distribución F por el teorema 6; así, tenemos que

$$F(x) = \mu(]-\infty, x]) = P(X \leq x)$$

Teorema 8. teo 3.1.2 [3] Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, entonces X es una variable aleatoria si y solo si para cada número real $x \in \mathbb{R}$ (o en un conjunto denso de $\mathbb{R}$) tenemos que

$$\{\omega : X(\omega) \leq x\} \in \mathcal{F}.$$

Demostración: $(\Leftarrow)$
Notemos que el problema anterior lo podemos escribir equivalentemente como

$$\forall x \in \mathbb{R} \qquad X^{-1}(]-\infty, x]) \in \mathcal{F}.$$

Ahora, tomemos $\mathcal{A} = \{S : S \subseteq \mathbb{R}$ y $X^{-1}(S) \in \mathcal{F}\}$. Así, se sigue que

$$X^{-1}(S^c) = (X^{-1}(S))^c \in \mathcal{F},$$

y si para todo $S_j \in \mathcal{A}$ con $j \in J$, entonces

$$X^{-1}\left(\bigcup_j S_j\right) = \bigcup_j X^{-1}(S_j) \in \mathcal{F}.$$

Así, $\mathcal{A}$ es estable por complementos y uniones arbitrarias; es decir es una clase monótona la cual contiene conjuntos de la forma

$$]-\infty, x]$$

los cuales generan $Bor(\mathbb{R})$ incluso cuando se restringen a un conjunto denso.
Por lo tanto, $Bor(\mathbb{R}) \subseteq \mathcal{A}$ y por lo tanto $X^{-1}(B) \in \mathcal{F}$ para cada $B \in Bor(\mathbb{R})$

Así, por definición X es una variable aleatoria.
Para ($\Rightarrow$) esta es trivial.

Teorema 9. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Si X es una variable aleatoria y f es una función Borel medible, entonces $f(X)$ es también una variable aleatoria.

Demostración.
P.D. $f(X)$ es una variable aleatoria. P.D. $f(x)$ es medible
Notemos que

$$f(X) = f \circ X : \Omega \longrightarrow \mathbb{R}$$
$$\omega \longmapsto f(X(\omega)).$$

Así, se sigue que

$$f(x) \Rightarrow (f \circ X)(\omega) \Rightarrow (f(x))^{-1} = (f \circ X)^{-1}(\omega) = (X^{-1} \circ f^{-1})(\omega).$$

de donde, como

$$X^{-1}(f^{-1}(Bor(\mathbb{R}))) \subseteq X^{-1}(Bor(\mathbb{R})) \subseteq \mathcal{F}.$$

Vector aleatorio.

Ahora, vamos a discutir la noción de vector aleatorio, este es solo un vector en el que cada componente es una variable aleatoria. Para su estudio es suficiente considerar el caso de dimensión 2, los casos de dimensiones más altas no representan una diferencia considerable

Así, consideremos el caso bidimensional. Sea

$$\mathcal{C} = \{(x,y) : a < x \leq b,\ c < y \leq d\} =]a,b] \times]c \times d],$$

notemos que $Bor(\mathbb{R}^2) = \sigma(\mathcal{C})$. Más aún también podemos considerar

$$\mathcal{D} = \{B_1 \times B_2 : B_1, B_2 \in Bor(\mathbb{R})\},$$

y aún se tiene que $\sigma(\mathcal{D}) = Bor(\mathbb{R}^2)$.

Sean X, Y dos variables aleatorias en $(\Omega, \mathcal{F}, P)$, se define el vector aleatorio como una función medible

$$(X,Y) : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}^2, Bor(\mathbb{R})^2),$$

que además induce una medida de probabilidad en el plano real de la siguiente manera

$$\forall \Delta \in Bor(\mathbb{R}^2),\ P_{(X,Y)}(\Delta) = P(\{\omega : (X(\omega), Y(\omega)) \in \Delta\}) = P(\{(X,Y) \in \Delta\})$$

A continuación se presenta un resultado análogo al (teorema 9)

Teorema 10 Sea X, Y dos variables aleatorias y f una función Borel medible de dos variables, entonces $f(X,Y)$ es una variable aleatoria.
teo 3.1.5 [3]

Demostración.
Análogamente la demostración del teorema 9, tenemos que

$$f(X,Y) = (f \circ (X,Y)) : \Omega \longrightarrow \mathbb{R}$$

de donde

$$[f\circ(X,Y)]^{-1}(Bor(\mathbb{R})) = (X,Y)^{-1}\circ f^{-1}(Bor(\mathbb{R})) \subseteq (X,Y)^{-1}(Bor(\mathbb{R}^2)) \in \mathcal{F}$$

La última contenencia nos dice que el mapa $(X,Y)^{-1}$ proyecta cada uno, un boreliano de dos dimensiones en el conjunto $\mathcal{F}$. Esto se prueba de la siguiente manera.
Si $A = B_1\times B_2$ con $B_1\in Bor(\mathbb{R})$, $B_2\in Bor(\mathbb{R})$, entonces

$$(X,Y)^{-1}(A) = X^{-1}(B_1)\cap Y^{-1}(B_2)\in\mathcal{F}$$

Corolario 11. Si X, Y son variables aleatorias, entonces también lo son $\min(X,Y)$, $\max(X,Y)$, $X+Y$, $X-Y$, $X\cdot Y$, X/Y. la última considera $Y\neq 0$.

Teorema 12 Sea $\{X_n\}_n$ una sucesión de variables aleatorias, entonces
teo 3.1.6 [3]

$$\inf_n X_n,\quad \sup_n X_n,\quad \liminf_n X_n,\quad \limsup_n X_n.$$

son variables aleatorias no necesariamente finitas

Demostración: Consideremos el caso $\sup_n X_n$. Para el mismo, notemos la relación

$$\forall x\in\mathbb{R}: \{\sup_n X_n \leq x\} = \bigcap_n \{X_n\leq x\}$$

Así, por el (teorema 8), se sigue que $\sup_n X_n$ es una variable aleatoria.

Para los casos restantes considere $\limsup_n X_n = \inf_j \sup_{n\geq j} X_n$

Definición 12 - Una variable aleatoria X se dice discreta o contable si existe un conjunto numerable $B\in Bor(\mathbb{R})$ tal que $P(X\in B)=1$.

Ejemplo: Sea $A\subseteq\Omega$, se define la función indicatriz por

$$\mathbb{1}_A:\Omega\longrightarrow\mathbb{R}$$
$$\omega\longmapsto \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega\in A\\ 0 & \text{si } \omega\notin A.\end{cases}$$

Claramente $\mathbb{1}_A$ es una variable aleatoria si $A\in\mathcal{F}$.

Si $\{A_i\}_i$ es una partición contable de Ω, es decir que

$$A_i\cap A_j=\emptyset \quad \text{si } j\neq i \qquad \text{y} \qquad \bigcup_i A_i = \Omega,$$

entonces se tiene que

$$1 = 1_\Omega = \sum_j \mathbb{1}_{A_j}$$

Una clase de variable aleatoria importante es la siguiente.

Definición 13. Sea $\{(A_i,b_i): i=1,\dots,k\}$ donde los A_i definen una partición medible de Ω y los $b_i\in\mathbb{R}$ definen los «pesos» de cada A_j, entonces se define la variable aleatoria

$$\varphi(\omega) = \sum_{i=1}^{k} b_i\,\mathbb{1}_{A_i}(\omega)$$

a una función que puede escribirse de esta forma se denomina variable aleatoria simple.

Esperanza.

- Si X es una variable aleatoria simple, es decir puede escribirse como

$$X(\omega) = \sum_{j\in J}' b_j \mathbb{1}_{A_j}(\omega), \qquad \text{con } |J| < +\infty,$$

se define la esperanza como

$$E[X] = \sum_{j\in J}' b_j P(A_j).$$

- Consideremos una variable aleatoria arbitraria positiva $(X \geq 0)$. Para dos enteros positivo m y n, el conjunto

$$A_{mn} = \left\{ \omega : \frac{n}{2^m} \leq X(\omega) \leq \frac{n+1}{2^m} \right\} \in \mathcal{F}.$$

Por otro lado, para cada m, definimos la variable aleatoria X_m simple con el conjunto $\{A_{mn}, n/2^m\}$ por lo tanto, se define de la siguiente manera

$$X_m(\omega) = \sum_{n=1}^{m2^m}{}' \frac{n}{2^m} \mathbb{1}_{A_{mn}}(\omega).$$

Más aún, se tiene que $X_m = n/2^m$ si y solo si $n/2^m \leq X(\omega) \leq (n+1)/2^m$.
Por otro lado, es claro que para cada m

$$\forall \omega : \ X_m(\omega) \leq X_{m+1}(\omega) \quad \text{y} \quad 0 \leq X(\omega) - X_m(\omega) < 1/2^m \tag{13}$$

Así, por (13), si $m \to +\infty$, entonces $X_m \to X$; así,

$$\lim_{m\to+\infty} X_m(\omega) = X(\omega).$$

Por otro lado, la esperanza se define como

$$E[X] = \lim_{m\to+\infty} E[X_n] = \lim_{m\to+\infty} \sum_{n=1}^{m2^m}{}' \frac{n}{2^m} P(A_{mn}) = \lim_{m\to+\infty} \sum_{n=1}^{m2^m}{}' \frac{n}{2^m} P\left(\left\{\frac{n}{2^m} \leq X \leq \frac{n+1}{2^m}\right\}\right)$$

Además, si para algum m, $E[X_m] = +\infty$, entonces $E[X] = +\infty$.

Finalmente, para X una variable aleatoria cualquiera, definimos

$$X^+ = \max\{X, 0\} \quad \text{y} \quad X^- = \max\{-X, 0\} \tag{14}$$

con lo que se tiene que

$$X = X^+ - X^- \tag{15}$$

donde X^+ y X^- son variables aleatorias positivas.

Ahora, considerando (15) y por las propiedades de la esperanza se sigue que

$$E[X] = E[X^+] - E[X^-]. \tag{16}$$

Más aún, cuando la esperanza existe, también se la denota por

$$E[X] = \int_\Omega X(\omega) P(d\omega). \tag{17}$$

Se observa que $|X| = X^+ + X^-$ (18)

Definición 14. Se dice que la esperanza de una variable aleatoria X es finita si $X^+ < +\infty$ y $X^- < +\infty$, por (18) se tiene equivalentemente que es finita si

def 2. (C2, S6) [1]

$$E[|X|] < +\infty$$

En este caso, se dice que la integral de Lesbegue es absolutamente convergente.

Propiedades de la esperanza.

1. Linealidad: $\int_\Omega (aX + bY)dP = a\int_\Omega X dP + b\int_\Omega Y dP.$

2. Aditividad sobre conjuntos, sea $\{A_i\}_{i\in I} \subset \mathcal{F}$ disjuntos
$$\int_{\cup_i A_i} X dP = \sum_{i\in I} \int_{A_i} X dP.$$

3. Positividad: Si $X \geq 0$ casi seguramente en A, entonces
$$\int_A X dP \geq 0.$$

4. Monotonía. Si $X_1 \leq X \leq X_2$ casi seguramente en A, entonces
$$\int_A X_1 dP \leq \int_A X dP \leq \int_A X_2 dP.$$

5. Valor medio: Si $a \leq X \leq b$ casi seguramente en A. $aP(A) \leq \int_A X dP \leq bP(A).$

6. Desigualdad de módulo: $\left|\int_A X dP\right| \leq \int_A |X| dP$

7. Convergencia dominada. Si $\lim_{n\to+\infty} X_n = X$ casi seguramente en A y para todo $n \in \mathbb{N}$ $|X_n| \leq Y$ casi seguramente en A con $\int_A Y dP < +\infty$, entonces
$$\lim_{n\to+\infty} \int_A X_n dP = \int_A X dP = \int_A \lim_{n\to+\infty} X_n dP.$$

8. Convergencia monótona. Si $X_n \geq 0$ casi seguramente en A y $X_n \uparrow X$ casi seguramente en A, entonces
$$\lim_{n\to+\infty} \int_A X_n dP = \int_A X dP = \int_A \lim_{n\to+\infty} X_n dP$$

Note que el resultado obtenido es similar a la convergencia dominada.

9. Integración por términos: Si $\sum_n \int_A |X_n| dP < +\infty$, entonces $\sum_n |X_n| < +\infty$ casi seguramente en A, por lo tanto $\sum_n X_n$ converge casi seguramente en A y
$$\int_A \left(\sum_n X_n\right) dP = \sum_n \int_A X_n dP$$

10. Lema de Fatou: Si $X \geq 0$ casi seguramente en A, entonces

$$\int_A \liminf_{n\to+\infty} X_n \, dP \leq \liminf_{n\to+\infty} \int_A X_n \, dP$$

11. Convergencia acotada. Si $\lim_{n\to+\infty} X_n = X$ casi seguramente y existe M tal que para todo n $|X_n| \leq M$ casi seguramente en A, entonces

$$\lim_{n\to+\infty} \int_A X_n \, dP = \int_A X \, dP = \int_A \lim_{n\to+\infty} X_n \, dP$$

A partir de aquí, si no se especifica nada, se supondrá que las variables aleatorias están definidas en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.

Teorema: 13. Dada una variable aleatoria, se verifica que

teo 3.2.1 [3]

$$\sum_{n=1}^{+\infty} P(|X| \geq n) \leq E[X] \leq 1 + \sum_{n=1}^{+\infty} P(|X| \geq n). \qquad (19)$$

donde $E[X] < +\infty$ si y solo si la serie de la ecuación (19) converge.

Demostración:
Consideremos la familia de conjuntos $A_n = \{n \leq |X| < n+1\}$ así, $\{A_n\}_n$ es una familia disjunta y por la propiedad (2)

$$E[X] = \int_{\bigcup_n A_n} |X| \, dP = \sum_{n=0}^{+\infty} \int_{A_n} |X| \, dP$$

Ahora, si aplicamos la (propiedad 5) valor medio, a cada A_n, se tiene que

$$\sum_{n=0}^{+\infty} nP(A_n) \leq E[|X|] \leq \sum_{n=0}^{+\infty} (n+1) P(A_n) = 1 + \sum_{n=0}^{+\infty} n P(A_n).$$

Así, resta probar que

$$P(A_n) = P(|X| \geq n). \qquad (20)$$

Considerando el método de Abel para sumatorios parciales para $N \geq 1$, se sigue que

$$\sum_{n=0}^{N} n \{P(|X| \geq n) - P(|X| \geq n+1)\} = \sum_{n=1}^{N} \{n - (n-1)\} P(|X| \geq n) - NP(|X| \geq N+1)$$

$$= \sum_{n=1}^{N} P(|X| \geq n) - NP(|X| \geq N+1).$$

Por lo tanto, se sigue que

$$\sum_{n=1}^{N} nP(A_n) \leq \sum_{n=1}^{N} P(|X| \geq n) \leq \sum_{n=1}^{N} nP(A_n) + N P(|X| \geq N+1) \qquad (21)$$

Finalmente si volvemos a aplicar la propiedad del valor medio, se sigue que

$$N P(|X| \geq N+1) \leq \int_{(|X| \geq N+1)} |X| \, dP.$$

Por lo tanto, si $E[|X|] < +\infty$, entonces el último término de (21) converge a cero cuando $N \to +\infty$ y así se obtiene (20). Por otro lado, si $E[|X|] = +\infty$, entonces el segundo término en (10) diverge cuando $N \to +\infty$, así (20) es cierta con ambos lados iguales a infinito.

Teorema 14. Sea X una variable aleatoria y f una función borel medible, entonces se tiene que
teo 3.2.2 [3]

$$\int_{\Omega} f(X(\omega))\, dP(\omega) = \int_{\mathbb{R}} f(x)\, P_X\, dx \qquad (22)$$

siempre que ambos lados de la ecuación existan

Demostración
Sea $B \in Bor(\mathbb{R})$ y $f = \mathbb{1}_B$, entonces se tiene que, por el teorema 8 se tiene la igualdad

$$\int_{\Omega} f(x(\omega))\, dP_{\omega} = P(X \in B) = \mu(B) = \int_{\mathbb{R}} f(x)\, P_x(dx)$$

Ahora, notemos que por linealidad de la integral, el resultado se mantiene si tomamos

$$f = \sum_j b_j \mathbb{1}_{B_j} \qquad (23)$$

una variable aleatoria simple basada en la partición con peso $\{B_j, b_j\}_j$. Y, una función arbitraria Borel medible podemos definir una sucesión $\{f_m\}_m$ de funciones de forma (23), tal que $f_m \uparrow f$ en todas partes, además, como para cada m

$$\int_{\Omega} f_m(x)\, dP = \int_{\mathbb{R}} f_m\, dP_x$$

entonces, cuando $m \to +\infty$ usando la propiedad de convergencia monótona, entonces

$$\int_{\Omega} f(X)\, dP = \int_{\mathbb{R}} f(x)\, dP_x(x).$$

Teorema 15 Sea f una función borel medible de $\mathbb{R}^2$ en $\mathbb{R}$, se tiene que
teo 3.2.3

$$\int_{\Omega} f(X(\omega), Y(\omega))\, dP = \int_{\mathbb{R}^2} f(x,y)\, dp_{x,y}(x,y)$$

siempre que las integrales existan

Como resultado del teorema 14, se tiene entonces que

$$E[f(x)] = \int_{\mathbb{R}} f(x)\, dP_X(x) = \int_{-\infty}^{+\infty} f(x)\, dF_X(x)$$

donde P_X es la medida de probabilidad inducida por X y F_X la función de distribución inducida por X. Por otro lado, también podemos concluir la linealidad de la esperanza. En efecto,

$$E[X+Y] = \int_{\mathbb{R}^2} (x+y)\, dp_{x,y}(x,y) = \int_{\mathbb{R}^2} x\, dp_{x,y}(x,y) + \int_{\mathbb{R}^2} y\, dp_{x,y}(x,y) = E[X] + E[Y]$$

Definición 15: Sean $a \in \mathbb{R}$ y $r > 0$, entonces $E[|X-a|^r]$ se llama el momento absoluto de X de orden r sobre a.

Usando la (definición 15) y el (teorema 14), se tiene que

$$E[|X-a|^r] = \int_{\mathbb{R}} |x-a|^r\, dP_X(x) = \int_{-\infty}^{+\infty} |x-a|^r\, dF_X(x).$$

Análogamente, se tiene que

$$E[(X-a)^r] = \int_{\mathbb{R}} (x-a)^r \, dP_X(x) = \int_{-\infty}^{+\infty} (x-a)^r \, dF_X(x).$$

Para $r=1$ y $a=0$, se reduce a $E[X]$ la cual se conoce como la media de X. Por otro lado, los momentos alrededor de la media, es decir, con $a=E[X]$, se conocen como momentos centrales. Para $r=2$, es conocido como la varianza, $var(X)$.

$$var(X) = \sigma^2(X) = E[(X-E(X))^2] = E[X^2] - E[X]^2$$

De la ecuación anterior, notemos la siguiente desigualdad

$$\sigma^2(X) \leq E[X^2].$$

Para cualquier número positivo p, se dice que X pertenece a $L^p = L^p(\Omega, \mathcal{F}, P)$ si y solo si $E[|X^p|] < +\infty$.

Desigualdades

Existen algunas desigualdades de interés.

Sean X, Y variables aleatorias, $1 < p < +\infty$ con p y q exponentes conjugados; así,

Hölder: $$|E[XY]| \leq E[|XY|] \leq E[|X|^p]^{1/p} E[|Y|^q]^{1/q}. \quad (23)$$

Minkowski $$E[|X+Y|^p]^{1/p} \leq E[|X|^p]^{1/p} + E[|Y|^p]^{1/p} \quad (24)$$

Un caso particular se obtiene con p y q iguales a 2 en (23), esta es la desigualdad de Cauchy-Schwarz-Bunyakovsky

$$E[|XY|] \leq E[|X|^2]^{1/2} E[|Y|^2]^{1/2} \quad (25)$$

Por otro lado, si tomamos $Y \equiv 1$, en (23), se obtiene

$$E|X| \leq E[|X|^p]^{1/p}. \quad (26)$$

En (26), reemplazando $|X|$ por $|X|^r$ y escribiendo $r' = s = pr$, se tiene que

Liapounov: $$E[|X|^r]^{1/r} \leq E[|X|^s]^{1/s} \quad \text{con} \quad 0 < r < s < +\infty$$

Proposición 16: Sea g una función convexa borel medible y $E[|X|] < +\infty$ ($X \in L^1$), entonces
pág 50 [3]

$$g(E[X]) \leq E[g(X)].$$

Demostración:
Por convexidad de g, sabemos que

$$g\left(\sum_{j=1}^{n} \lambda_j y_j\right) \leq \sum_{j=1}^{n} \lambda_j g(y_j) \quad (27)$$

con $\sum_{j=1}^{n} \lambda_j = 1$. Consideremos el caso donde X es una función simple. Así, se tiene que

$$E[X] = \sum_{j=1}^{n} \lambda_j y_j \quad \text{y} \quad E\,g(X) = \sum_{j=1}^{n} \lambda_j\, g(y_j)$$

Así, por (27), se sigue el resultado deseado.

Proposición 17: Desigualdad de Chebyshev. Sea φ una función estrictamente positiva y creciente en $]0,+\infty[$ con $\varphi(u)=\varphi(-u)$ y X una variable aleatoria tal que $E[\varphi(X)]<+\infty$, ($\varphi(X)\in L^1$). Así, para cada $u>0$, se tiene que

$$P(|X|\geq u) \leq \frac{E[\varphi(X)]}{\varphi(u)}.$$

Demostración.
Por el teorema del valor medio, se tiene que

$$E[\varphi(X)] = \int_{\Omega} \varphi(X)\, dP \geq \int_{\{|X|\geq u\}} \varphi(X)\, dP \geq \varphi(u)\, P\{|X|\geq u\}.$$

De donde, se concluye que

$$\frac{E[\varphi(X)]}{\varphi(u)} \geq P(|X|\geq u).$$

Independencia

Definición 16: Independencia. Una familia de variables aleatorias $\{X_j : 1\leq j\leq n\}$ se dice independiente si para cualquier colección de conjuntos borelianos $\{B_j : 1\leq j\leq n\}$ se tiene que

$$P\left(\bigcap_{j=1}^{n}\{X_j\in B_j\}\right) = \prod_{j=1}^{n} P(X_j\in B_j)$$

- En el caso de una familia arbitraria de variables aleatorias, estas son independientes si y solo si la definición se verifica para cualquier subfamilia finita.
- Se dicen independientes dos a dos si para cada par son independientes
- La definición implica que para cualquier subconjunto de variables aleatorias de $\{X_j : 1\leq j\leq n\}$ es también independiente.

Teorema 18. Una familia de variables aleatorias $\{X_j : 1\leq j\leq n\}$ es independiente si
pág 53 [3]

$$P\left(\bigcap_{j=1}^{n}\{X_j\leq x_j\}\right) = \prod_{j=1}^{n} P(X_j\leq x_j), \quad \text{para todo } (x_1,\ldots,x_n)\in\mathbb{R}^n.$$

Demostración
$\Leftarrow$ Sean $(x_1,\ldots,x_n)\in\mathbb{R}^n$ y $\{X_j : 1\leq j\leq n\}$ una familia de variables aleatorias cualesquiera.
Tenemos que

$$P\left(\bigcap_{j=1}^{n}\{X_j\leq x_j\}\right) = \prod_{j=1}^{n} P(X_j\leq x_j)$$

tomando $B_j =]-\infty, x_j]$, se tiene que

$$P\left(\bigcap_{j=1}^{n}\{X_j\in B_j\}\right) = \prod_{j=1}^{n} P(X_j\in B_j).$$

Recíprocamente, tomando $B_j =]-\infty, x_j]$ el cual es borel medible, se sigue el resultado deseado.

Teorema 19. Sea $\{X_j : 1 \le i \le n\}$ una familia de variables aleatorias independientes. y
teo 3.3.1 [3] $\{f_j : 1 \le j \le n\}$ funciones Borel medibles, entonces las variables aleatorias $\{f_j(X_j) : 1 \le j \le n\}$ son independientes.

Demostración:
Sea $A_j \in Bor(\mathbb{R})$, entonces por hipótesis $f_j^{-1}(A_j) \in Bor(\mathbb{R})$.

$$\bigcap_{j=1}^{n} \{f_j(X_j) \in A_j\} = \bigcap_{j=1}^{n} \{X_j \in f^{-1}(A_j)\}.$$

De donde,

$$P\left(\bigcap_{j=1}^{n} \{f_j(X_j) \in A_j\}\right) = P\left(\bigcap_{j=1}^{n} \{X_j \in f^{-1}(A_j)\}\right) = \prod_{j=1}^{n} P(X_j \in f^{-1}(A_j))$$

$$= \prod_{j=1}^{n} P(f_j(X_j) \in A_j).$$

Así, como A_j fue arbitrario, $f_j(X_j)$ es una variable independiente para cada j.

Teorema 20 Sean $1 \le n_1 < n_2 < \cdots < n_k = n$ y las funciones Borel medibles
teo 3.3.2 [3]

$$f_1 : \mathbb{R}^{n_1} \to \mathbb{R}$$
$$f_2 : \mathbb{R}^{n_2 - n_1} \to \mathbb{R}$$
$$f_3 : \mathbb{R}^{n_3 - n_2} \to \mathbb{R}$$
$$\vdots$$
$$f_k : \mathbb{R}^{n_k - n_{k-1}} \to \mathbb{R}.$$

Si la familia de variables aleatorias $\{X_i : 1 \le i \le n\}$ es independiente, entonces las variables aleatorias

$$f_1(X_1, \ldots, X_{n_1}),\ f_2(X_{n_1+1}, \ldots, X_{n_2}), \ldots,\ f_k(X_{n_{k-1}+1}, \ldots, X_{n_k})$$

son independientes.

Teorema 21 Sean X, Y dos variables aleatorias independientes. con esperanza finito,
teo. 3.3.3 [3]. entonces $E[XY] = E[X]E[Y]$.

Antes de revisar la demostración del (teorema 21), notemos lo siguiente. Sean, μ^n la medida de probabilidad inducida por el vector $(X_1, \ldots, X_n)$ en $(\mathbb{R}^n, Bor(\mathbb{R}^n))$ y la familia de medidas de probabilidad $\{\mu_j : 1 \le i \le n\}$ inducidas por cada X_j en $(\mathbb{R}, Bor(\mathbb{R}))$, se tiene la siguiente relacion

$$\mu^n\left(\bigtimes_{j=1}^{n} B_j\right) = \prod_{j=1}^{n} \mu_j(B_j)$$

donde notamos $\bigtimes_{j=1}^{n} B_j$ al producto $B_1 \times B_2 \times \cdots \times B_n$.

Demostración (teorema 21)
Consideremos el vector aleatorio (X, Y), entonces, por el (teorema 15)

$$E[XY] = \int_{\Omega} XY\, dP = \int_{\mathbb{R}^2} xy\, dp_{XY}(x,y) = \int_{\mathbb{R}^2} xy\, dp_X(x)\, dp_Y(y)$$

$$= \int x\, dp_X(x) \int y\, dp_Y(y) = E[X]\,E[Y]$$

Corolario: Sea $\{X_j : 1\leq j\leq n\}$ una familia de variables aleatorias independientes con esperanza finita, entonces

$$E[\Pi_{j=1}^n X_j] = \Pi_{j=1}^n E[X_j]$$

Ejemplo:
Para $n\geq 2$, sean $\{(\Omega_j, \mathcal{F}_j, P_j) : 1\leq j\leq n\}$ espacios de probabilidad discretos. Definimos el espacio producto

$$\Omega^n = \Omega_1 \times \cdots \times \Omega_n$$

el espacio de las n-uplas $\omega^n = (\omega^1, \dots, \omega^n)$ donde $\omega_j \in \Omega_j$ y además, tenemos que

$$P^n(\{\omega^n\}) = \prod_{j=1}^n P_j(\{\omega_j\})$$

Esta medida de probabilidad se denomina medida producto y la notamos $P^n = \times_{j=1}^n P_j$. Más aún se tiene que, si $S_j \in \mathcal{F}_j$ para cada $1\leq j\leq n$, entonces

pág 58 [3].
$$P^n\left(\bigtimes_{j=1}^n S_j\right) = \prod_{j=1}^n P_j(S_j) \tag{28}$$

Ahora, sea X_j una variable aleatoria en Ω_j y $B_j \in Bor(\mathbb{R})$ cualquiera con $S_j = X^{-1}(B_j)$, entonces

$$S_j = \{\omega_j \in \Omega_j : X_j(\omega_j) \in B_j\}.$$

Así, por (28)

$$P^n\left\{\bigtimes_{j=1}^n \{X_j \in B_j\}\right\} = P^n\left\{\bigtimes_{j=1}^n S_j\right\} = \prod_{j=1}^n P_j(S_j) = \prod_{j=1}^n P_j\{X_j \in B_j\}.$$

Ahora, para cada X_j, se define $\tilde{X}_j$ de la siguiente forma

$$\tilde{X}_j : (\Omega^n, \mathcal{F}^n) \longrightarrow (\mathbb{R}, Bor(\mathbb{R}))$$

tal que para todo $\omega \in \Omega^n$, $\tilde{X}_j(\omega) = X_j(\omega_j)$, entonces

$$\bigcap_{j=1}^n \{\omega : \tilde{X}_j(\omega) \in B_j\} = \bigtimes_{j=1}^n \{\omega_j : X_j(\omega_j) \in B_j\},$$

pues

$$\{\omega : \tilde{X}_j(\omega) \in B_j\} = \Omega_1 \times \cdots \times \Omega_{j-1} \times \{\omega_j : X_j(\omega_j) \in B_j\} \times \Omega_{j+1} \times \cdots \times \Omega_n.$$

De donde, por (28), se sigue que

$$P^n\left(\bigcap_{j=1}^n \{\tilde{X}_j \in B_j\}\right) = \prod_{j=1}^n P^n\{\tilde{X}_j \in B_j\}.$$

Por lo tanto, la familia $\{\tilde{X}_j : 1\leq j\leq n\}$ es independiente

Ejemplo:
Sea $\mathcal{U}^n$ el cubo n-dimensional

$$\mathcal{U}^n = \{(x_1, \dots, x_n),\ 0\leq x_j \leq 1;\ 1\leq j\leq n\}$$

Con el espacio $(\mathbb{R}^n, Bor(\mathbb{R}^n), m^n)$ con $\mathbb{R}^n$ el espacio euclideo, $Bor(\mathbb{R}^n)$ y m^2 el conjunto de los borelianos y la medida usual, respectivamente, entonces es un espacio de probabilidad. Entonces, si se restringe a $(\mathcal{U}^n, Bor(\mathbb{R}^n) \cap \mathcal{U}^n, m^n)$, este es un espacio de probabilidad.

Por otro lado, sean $\{f_i, 1 \leq i \leq n\}$ funciones Borel medibles de una variable, tales que

$$X_j((x_1, \ldots, x_n)) = f_j(x_j).$$

entonces la familia de variables aleatorias $\{X_j : 1 \leq j \leq n\}$ es independiente. En particular, si $f_j(x_j) \equiv x_j$, se obtienen las n coordenadas en el cubo.

Ejemplo.
En $(\mathbb{R}^n, Bor(\mathbb{R}^n))$, es posible construir una medida basada en las medidas obtenidas para cada $(\mathbb{R}, Bor(\mathbb{R}))$. Sean $\{\mu_j : 1 \leq j \leq n\}$, se define μ^n de la siguiente manera.

$$\mu^n\left(\hat{X}_{j=1}^{n} B_j\right) = \prod_{j=1}^{n} \mu_j(P_j)$$

para todo $B_j \in Bor(\mathbb{R})$ con $1 \leq j \leq n$.

Teorema 17. teo. 3.3.4 [3]. Sea $\{P_j\}$ una sucesión finita o infinita de medidas de probabilidad en $(\mathbb{R}, Bor(\mathbb{R}))$, o equivalentemente de funciones de distribución, entonces existe un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y una sucesión de variables independientes $\{X_j : 1 \leq j \leq n\}$ tales que para cada j, P_j es una medida de probabilidad inducida por X_j.

Capitulo 3: Convergencia de Variables aleatorias.

Definición 3.1 **Convergencia casi segura (c.s.)** Sea $\{X_n\}$ una sucesión de variables aleatorias, se dice que converge casi seguramente a X una variable aleatoria si y solo si existe un conjunto de probabilidad (o medida) nula, N tal que

$$\forall \omega \in \Omega \setminus N : \lim_{n \to +\infty} X_n(\omega) = X(\omega) < +\infty.$$

Notemos que, por lo tanto, se permite que la sucesión diverja en un conjunto de medida nula.

Teorema 3.2. teo 4.1.1 [3] Una sucesión de variables aleatorias $\{X_n\}$ converge casi seguramente a X si y solo si para todo $\varepsilon > 0$, se tiene que

$$\lim_{m \to +\infty} P(|X_n - X| \leq \varepsilon, \text{ para todo } n \geq m) = 1;$$

equivalentemente

$$\lim_{m \to +\infty} P(|X_n - X| > \varepsilon, \text{ para algún } n \geq m) = 0.$$

Demostración
$\Rightarrow$. Supongamos que $X_n \xrightarrow{c.s.} X$ así, existe un conjunto de probabilidad nula N tal que $\Omega_0 = \Omega \setminus N$. Para $m \geq 1$, definamos

$$A_m(\varepsilon) = \bigcap_{n=m}^{+\infty} \{|X_n - X| \leq \varepsilon\}.$$

Así, $A_m(\varepsilon)$ forma una sucesión creciente con la inclusión de conjuntos. Por otro lado para $\omega_0 \in \Omega_0$, por hipótesis, existe $m(\omega_0, \varepsilon)$ tal que

$$|X_n(\omega_0) - X(\omega_0)| < \varepsilon \qquad \forall n \geq m(\omega_0,\varepsilon);$$

por lo tanto $\omega_0 \in A_m(\varepsilon)$ para algún m y ε apropiados. Así, se sigue que

$$\Omega_0 \subseteq \bigcup_{m=1}^{+\infty} A_m(\varepsilon)$$

de donde,

$$1 = P\left(\bigcup_{m=1}^{+\infty} A_m(\varepsilon)\right) = P\left(\lim_{m\to+\infty} A_m(\varepsilon)\right) = \lim_{m\to+\infty} P(A_m(\varepsilon)),$$

como se quería.

$\Leftarrow$) Recíprocamente, supongamos que

$$\lim_{m\to+\infty} P(|X_n - X| \leq \varepsilon,\ \forall n \geq m) = 1.$$

Así, notemos $A(\varepsilon) = \bigcup_{m=1}^{+\infty} A_m(\varepsilon)$, de donde, se sabe que $P(A(\varepsilon)) = 1$. Por otro lado, se tiene que si $\omega_0 \in A(\omega)$, entonces, existe $m(\omega_0, \varepsilon)$ tal que

$$|X_n(\omega_0) - X(\omega_0)| \leq \varepsilon \qquad \forall n \geq m.$$

Consideremos (ε_n) una sucesión decreciente a cero (por ejemplo $1/n$) y tomando

$$A = \bigcap_{n=1}^{+\infty} A\left(\frac{1}{n}\right),$$

entonces

$$P(A) = \lim_{n\to+\infty} P\left(A\left(\frac{1}{n}\right)\right) = 1.$$

Por otro lado, sea $\omega \in A$, entonces; por definición $|X_n(\omega) - X(\omega)| \leq \varepsilon$ para un n suficientemente grande y por lo tanto, para todo $\omega \in A$,

$$X_n(\omega) \to X(\omega)$$

donde $P(A) = 1$. Es decir, diverge en un conjunto de probabilidad nula.

Definición 3.3. Convergencia en probabilidad Una sucesión de variables aleatorias $\{X_n\}$, se dice que converge en probabilidad (a la variable X) si para todo $\varepsilon > 0$

$$\lim_{n\to+\infty} P(|X_n - X| > \varepsilon) = 0.$$

Teorema 3.4. Sea $\{X_n\}$ una sucesión de variables aleatorias, si $\{X_n\}$ converge a X casi seguramente, entonces también converge en probabilidad.

Demostración:
Para todo $\varepsilon > 0$, se tiene que

$$P(|X_n - X| > \varepsilon) \leq P\left(\bigcup_{m\geq n} \{|X_n - X| > \varepsilon\}\right) = \lim_{n\to+\infty} P(|X_m - X| > \varepsilon, \text{ para algún } m \geq n) = 0$$

Por lo tanto, si $\{X_n\}$ converge casi seguramente, entonces también converge en probabilidad

Teorema 3.5. teo 4.1.3 [3] (Caracterizaciones de Cauchy) Sea $\{X_n\}$ una sucesión de variables aleatorias.

- $\{X_n\}$ converge casi-seguramente si y sólo si

$$\lim_{k\to+\infty} P(|X_n - X_m| > \varepsilon \text{ para algunos } m>n>k) = 0.$$

- $\{X_n\}$ converge en probabilidad si y sólo si

$$\lim_{\substack{n\to+\infty\\ m\to+\infty}} P(|X_n - X_m| > \varepsilon) = 0.$$

Teorema 3.6. Sea $\{X_n\}$ una sucesión de variables aleatorias. Si $\{X_n\}$ converge en probabilidad a X, entonces, existe una subsucesión $\{X_{n_k}\}$ que converge casi-seguramente a X

Demostración.
Supongamos que $\{X_n\}$ converge en probabilidad. Así, existe $n_1 \in \mathbb{N}$ tal que

$$P(\{\omega\in\Omega : |X_{n_1}(\omega) - X(\omega)| \geq 2^{-1}\}) < 2^{-1}.$$

Para $k>1$, hallamos $n_k > n_{k-1}$ tal que

$$P(\{\omega\in\Omega : |X_{n_k}(\omega) - X(\omega)| \geq 2^{-k}\}) < 2^{-k}.$$

Así, definamos

$$S_k = \bigcup_{i\geq k} \{\omega\in\Omega : |X_{n_i}(\omega) - X(\omega)| \geq 2^{-i}\}.$$

Notemos entonces que la sucesión (S_k) es decreciente. Por otro lado, definamos

$$S = \bigcap_{k\geq 1} S_k$$

Así, se sigue que

$$P(S_k) \leq 2^{-k} + 2^{-k-1} + 2^{-k-2} + \cdots = 2^{1-k}$$

y por lo tanto

$$P(S) = \lim_{k\to+\infty} P(S_k) = 0.$$

Corolario: 3.7. Sea $\{X_n\}$ una sucesión de variables aleatorias. Así $\{X_n\}$ converge en probabilidad a X si y sólo si cada subsucesión $\{X_{n_k}\}$ contiene una subsucesión $\{X_{n_{k_i}}\}$ que converge a X casi seguramente

Corolario 3.8. Sean g una función continua y $\{X_n\}$ una sucesión de variables aleatorias

a) Si $X_n \xrightarrow{c.s.} X$, entonces $g(X_n) \xrightarrow{c.s.} g(X)$
b) Si $X_n \xrightarrow{P} X$, entonces $g(X_n) \xrightarrow{P} g(X)$

Definición 3.9. - Convergencia en L_p - Sea $\{X_n\}$ una sucesión de variables aleatorias, entonces $X_n \xrightarrow{L^p} X$ (converge en L^p) si

$$\lim_n E|X_n - X|^p = 0,$$

con X y $X_n \in L_p$ para cada $n \in \mathbb{N}$ y para todo $1 < p < +\infty$

Observación: Equivalentemente, si $X_n \xrightarrow{L_p} X$, entonces $(X_n - X) \xrightarrow{L_p} 0$

Teorema 3.10. Si $X_n \xrightarrow{L_p} X$, entonces $X_n \xrightarrow{P} X$, el recíproco es cierto si (X_n) está dominada por algún $Y \in L^p$, es decir, $(|X_n| \leq Y)$

Demostración
- Para la primera parte, notemos que

$$\begin{aligned}
P(|X_n - X| \geq \varepsilon) &= \int \mathbb{1}_{\{|X_n - X| > \varepsilon\}} \, dP \\
&= \int \mathbb{1}_{\{|X_n - X|^p > \varepsilon^p\}} \, dP \\
&\leq \int \frac{|X_n - X|^p}{\varepsilon^p} \mathbb{1}_{\{|X_n - X|^p > \varepsilon^p\}} \, dP \quad \text{Por Chebyshev} \\
&\leq \frac{1}{\varepsilon^p} \int |X_n - X|^p \, dP \to 0 \quad \text{cuando } n \to +\infty.
\end{aligned}$$

Para la segunda parte, si suponemos $|X_n| \leq Y$ (en casi todo punto con $E(Y^p) < +\infty$, y sin pérdida de generalidad, suponiendo que $X \equiv 0$, se sigue que

$$E[|X_n|^p] = \int_{\{|X| < \varepsilon\}} |X_n|^p \, dP + \int_{\{|X| \geq \varepsilon\}} |X_n|^p \, dP \leq \varepsilon^p + \int_{\{|X_n| \geq \varepsilon\}} Y^p \, dP. \tag{28}$$

Así, como $X_n \xrightarrow{P} 0$, entonces $P(\{|X_n| \geq \varepsilon\}) \to 0$.

* **Proposición 3.11** Si $E[|X|] < +\infty$ y $\lim P(A_n) = 0$, entonces

Ejer 2, 3.2 [3]

$$\lim_{n \to +\infty} \int_{A_n} X \, dP = 0.$$

En particular

$$\lim_{n \to +\infty} \int_{\{|X| > n\}} X \, dP = 0.$$

Usando la proposición (3.11), como $E(Y^p) < +\infty$ y dado que $P(\{X_n \geq \varepsilon\}) \to 0$, entonces

$$\int_{\{|X_n| \geq \varepsilon\}} Y^p \, dP = 0$$

de donde, si $\varepsilon \to 0$, se sigue que en (28)

$$\lim_{n \to +\infty} E[|X_n|^p] = 0$$

Por lo tanto, $X_n \xrightarrow{L^p} 0$.

Teorema 3.12. Sea X una variable aleatoria, $X_n \xrightarrow{P} 0$ si y solo si

$$E\left[\frac{|X_n|}{1+|X_n|}\right] \to 0.$$

Más aún,

$$\rho(X,Y) = E\left[\frac{|X-Y|}{1+|X-Y|}\right]$$

definen una métrica en el espacio de las variables aleatorias.

Ejercicio 3.13. Si $X \geq 0$ casi en todo punto, en A y $\int_A X\,dP = 0$, entonces $X=0$ en casi todo punto.

Demostración (teorema 3.12)

Para probar que ρ define una métrica, notemos que si $\rho(X,Y)=0$, entonces $E|X-Y|=0$, de donde, por el (ejercicio 3.13), $X-Y=0$; así, $X=Y$. Finalmente, basta probar que se satisface la desigualdad triangular; así, para todo real x, y

P.D. $$\frac{|x+y|}{1+|x+y|} \leq \frac{|x|}{1+|x|} + \frac{|y|}{1+|y|}$$

Por simetría, supongamos que $|y| \leq |x|$, así

$$\frac{|x+y|}{1+|x+y|} - \frac{|x|}{1+|x|} = \frac{|x+y|-|x|}{(1+|x+y|)(1+|x|)} \leq \frac{||x+y|-|x||}{1+|x|} \leq \frac{|y|}{1+|y|}$$

Así, $\rho(X,Y)$ define una métrica.

Para la primera parte del teorema, notemos que $|X|/(1+|X|) \leq 1$; así, por el recíproco del (teorema 3.10) la primera afirmación del (teorema 3.12) se tiene mostramos que

$$|X_n| \xrightarrow{P} 0 \iff \frac{|X_n|}{1+|X_n|} \xrightarrow{P} 0$$

Pero, para $\varepsilon > 0$, cualquiera, se tiene que

$$|x| \leq \varepsilon \iff \frac{|x|}{1+|x|} \leq \frac{\varepsilon}{1+\varepsilon} \qquad \forall x \in \mathbb{R}.$$

De donde, se sigue el resultado deseado.

Observación: En general:

- Convergencia en probabilidad no implica convergencia en L^p.
- Convergencia casi segura no implica convergencia en L^p

Más información pág. 73 [3].

Sección 3.2 Lemas de Borel - Cantelli

Un concepto importante en teoría de conjuntos está relacionado con «lim sup» y «lim inf», estos se pueden definir en cualquier espacio arbitrario Ω.

Definición 3.14. Sea $\{E_n\}_n$ una sucesión de subconjuntos de Ω, entonces

$$\limsup_n E_n = \bigcap_{n=1}^{+\infty} \bigcup_{m=n}^{+\infty} E_n \quad \text{y} \quad \liminf_n E_n = \bigcup_{n=1}^{+\infty} \bigcap_{m=n}^{+\infty} E_n. \tag{29}$$

Además, obsevemos que

$$\liminf_n E_n = \left(\limsup_n E_n^c\right)^c. \tag{30}$$

Proposición 3.15 a) Un elemento pertenece a $\limsup_n E_n$ si y solo si pertenece a infinitamente muchos E_n.
b) Un elemento pertenece a $\liminf_n E_n$ si y solo si pertenece a todos los E_n a partir de cierto n.

Demostración:
Para a), un elemento pertenece a infinitamente muchos E_n si y solo si no pertenece a todos los E_n a partir de cierto n; así, por (30) se sigue el resultado.

Ahora, si ω pertenece infinitamente a muchos E_n's, entonces pertenece a

$$F_m = \bigcup_{n \geq m} E_n \quad \text{para cada } m,$$

luego, por construcción

$$\omega \in \bigcap_{m=1}^{+\infty} F_m = \limsup E_n.$$

Recíprocamente, si $\omega \in \bigcap_{m=1}^{+\infty} F_m$, entonces $\omega \in F_m$ para cada m. Por contradicción, supongamos que ω no pertenece a un infinitamente muchos F_n, es decir pertenece para un número finito. De esta manera, existe un $p \in \mathbb{N}$ tal que

$$\omega \notin F_m = \bigcup_{n \geq m} E_n \quad \text{para } m \geq p.$$

Pero esto contradice nuestra hipótesis pues $\omega \in \bigcap_{m=1}^{+\infty} F_m$. Así ω pertenece a infinitamente muchos E_n.

En un lenguaje más intuitivo, se dice que el evento $\limsup_n E_n$ ocurre si y solo si los eventos E_n ocurren (muchas veces infinitamente) infinitamente muchas veces y se denota

$$\limsup_n E_n = \{E_n;\ i.o.\}.$$

Por otro lado, también se tiene que

$$P\left(\limsup_n E_n\right) = \lim_n P\left(\bigcup_{m \geq n} E_n\right) \quad \text{y} \quad P\left(\liminf_n E_n\right) = \lim_n P\left(\bigcap_{m \geq n} E_n\right) \tag{31}$$

la demostración de este resultado es inmediata.

Teorema 3.16 - Primer lema de Borel-Cantelli- Sea $\{E_n : n \geq 1\} \subseteq \mathcal{F}$, se tiene que
teo 4.2.1 [3]

$$\sum_n P(E_n) < +\infty \Rightarrow P(\{E_n : i.o.\}) = 0.$$

Demostración:
Por la desigualdad de medidas de probabilidad, se tiene que

$$P\left(\bigcup_{n=m}^{+\infty} E_n\right) \leq \sum_{n=m} P(E_n)$$

Así, puesto que por hipótesis, la serie es convergente entonces la cola de la misma converge a cero, es decir $\sum_{n=m}^{+\infty} P(E_n) \to 0$ cuando $m \to +\infty$. Y, junto a la desigualdad anterior

$$0 \leq \lim_{m\to+\infty} P\left(\bigcup_{n=m}^{+\infty} E_n\right) \leq 0 \Rightarrow \lim_{m\to+\infty} P\left(\bigcup_{n=m}^{+\infty} E_n\right) = 0$$

Por lo tanto, junto con (31) se sigue el resultado deseado.

Teorema 3.17. Sea $(X_n)_n$ una sucesión de variables aleatorias, entonces $X_n \xrightarrow{c.s.} 0$ si y solo si $P(\{|X_n| > \varepsilon \text{ . i.o.}\}) = 0$, para todo $\varepsilon > 0$.

teo. 4.2.2 [3].

Demostración:

Usando la notación del (teorema 3.2), consideremos, para $\varepsilon > 0$, arbitrario pero fijo

$$A_m = \bigcap_{n=m}^{+\infty} \{|X_n| \leq \varepsilon\}.$$

De donde, se sigue que

$$\{|X_n| > \varepsilon \text{ , i.o.}\} = \bigcap_{m=1}^{+\infty} \bigcup_{n=m}^{+\infty} \{|X_n| > \varepsilon\} = \bigcap_{m=1}^{+\infty} A_m^c.$$

Así, por el (teorema 3.2); sabemos que $X_n \xrightarrow[a.s.]{} 0$ si y solo si para todo $\varepsilon > 0$

$$P(A_m^c) = P\left(\bigcup_{n=m}^{+\infty} \{|X_n| > \varepsilon\}\right) \to 0 \quad \text{cuando } m \to +\infty.$$

Por lo tando, como (A_m^c) es una sucesión decreciente, entonces junto con la hipótesis, se sigue el resultado deseado.

Una prueba alternativa del (teorema 3.6) puede ser realizada usando el teorema anterior, para más información consultar el (teorema 4.2.3) [3].

Teorema 3.18. Sea $(E_n)_n$ una sucesión de eventos independientes, entonces se tiene que

teo 4.2.4 [3]

$$\sum_n P(E_n) = +\infty \Rightarrow P(\{E_n : i.o.\}) = P(\limsup_n E_n) = 1. \quad (32)$$

Demostración:

Por (31), tenemos que

$$P(\liminf_n E_n^c) = \lim_{m\to+\infty} P\left(\bigcap_{n=m}^{+\infty} E_n^c\right). \quad (33)$$

Además, dado que cada E_n es independiente, entonces E_n^c también lo es; de esta manera, para $k \geq m$, se sigue que

$$P\left(\bigcap_{n=m}^{k} E_n^c\right) = \prod_{n=m}^{k} P(E_n^c) = \prod_{n=m}^{k} (1 - P(E_n)).$$

Ahora, para todo $x \geq 0$, se tiene que $e^{-x} \geq 1 - x$, usando esto en la expresión anterior, se tiene que

$$\prod_{n=m}^{k} (1 - P(E_n)) \leq \prod_{n=m}^{k} e^{-P(E_n)} = e^{-\sum_{n=m}^{k} P(E_n)}$$

Y, tomando $k \to +\infty$, entonces por hipótesis $\sum_{n=m}^{k} P(E_n) \to +\infty$; así $e^{-\sum_{n=m}^{k} P(E_n)} \to 0$. Luego, por monotonía se sigue que

$$P\left(\bigcap_{n=m}^{+\infty} E_n^c\right) = \lim_{k\to+\infty} P\left(\bigcap_{n=m}^{k} E_n^c\right) = 0,$$

y se concluye que

$$P(\liminf_n E_n^c) = 0.$$

Finalmente, por (30), se sigue el resultado deseado.

La implicación de (32) se mantiene si los eventos $\{E_n\}_n$ son independientes dos a dos.

Sección 3.3: Convergencia vaga.

Consideremos $(X_n)_n$ una sucesión de variables aleatorias, entonces si $(X_n)_n$ converge a algún límite, entonces la sucesión corespondente de medidas de probabilidad $(P_n)_n$ debería converger a algún límite en algún sentido.

Una primera pregunta sería, ¿Existe el $\lim_{n\to+\infty} P_n(A)$ para todo $A \in Bor(\mathbb{R})$?

Definición 3.19 - Medida de subprobabilidad- Sea μ una medida sobre $(\mathbb{R}, Bor(\mathbb{R}))$, con $\mu(\mathbb{R}) \leq 1$, entonces μ se denomina una medida de subprobabilidad.

Definición 3.20 - Convergencia vaga- Sea $(\mu_n)_n$ una sucesión de medidas de subprobabilidad. Se dice que (μ_n) converge vagamente a μ una medida de subprobabilidad si y solo si existe D un subconjunto denso de $\mathbb{R}$ tal que

$$(\forall a \in D)(\forall b \in D)(a < b) \quad \mu_n(]a,b]) \longrightarrow \mu(]a,b]).$$

Se nota $\mu_n \xrightarrow{v} \mu$.

Definición 3.21. Se dice que un intervalo (a,b) es un intervalo de continuidad de μ si y solo si ni a ni b son átomos de μ, es decir $\mu(]a,b[) = \mu([a,b])$.

Como convención, se tiene que $\mu(]a,b[) = 0$ cuando $a > b$.

Teorema 3.22. Sean $(\mu_n)_n$ y μ medidas de subprobabilidad, entonces las siguientes proposiciones son equivalentes:

a) Para todo intervalo de continuidad (a,b) y $\varepsilon > 0$, existe $n_0(a,b,\varepsilon)$, tal que para todo $n \geq n_0$ $\quad \mu(a+\varepsilon, a-\varepsilon) - \varepsilon \leq \mu_n(a,b) \leq \mu(a-\varepsilon, b+\varepsilon) + \varepsilon$

b) Para todo intervalo de continuidad $\mu_n(a,b] \longrightarrow \mu(a,b]$.

c) $\mu_n \xrightarrow{v} \mu$.

Demostración:

a) $\Rightarrow$ b) Sea (a,b) un intervalo de continuidad de μ, por la monotonía de una medida, se tiene que

$$\lim_{\varepsilon \downarrow 0} \mu(a+\varepsilon, b-\varepsilon) = \mu(a,b) = \mu[a,b] = \lim_{\varepsilon \downarrow 0} \mu_n(a-\varepsilon, b+\varepsilon).$$

De donde si $n \to +\infty$ y $\varepsilon \downarrow 0$ en nuestra hipótesis, entonces

$$\mu(a,b) \leq \liminf_{n\to+\infty} \mu_n(a,b) \leq \limsup_{n\to+\infty} \mu_n[a,b] \leq \mu[a,b] = \mu(a,b)$$

por lo tanto, se concluye que

$$\mu_n(a,b] \longrightarrow \mu(a,b].$$

b) $\Rightarrow$ c)

Notemos que $(a,b]$ puede ser reemplazado por (a,b), $[a,b]$ o $[a,b)$. Así, como el conjunto de átomos de μ es contable, el complemento D es denso en $\mathbb{R}$. Así, si $a,b \in D$, entonces (a,b) es un intervalo de continuidad por definición de D. Por lo tanto,

$$\mu(a,b) = \mu_n(a,b] \longrightarrow \mu(a,b] = \mu(a,b).$$

c) $\Rightarrow$ a)

Supongamos que $\mu_n \xrightarrow{v} \mu$. Sea $\varepsilon > 0$ cualquiera, entonces existen $a_1, a_2, b_1, b_2 \in D$ tales que

$$a - \varepsilon < a_1 < a < a_2 < a + \varepsilon \qquad \text{y} \qquad b - \varepsilon < b_1 < b < b_2 < b + \varepsilon \tag{34}$$

pues D es denso en $\mathbb{R}$, para cualquier intervalo (a,b). De donde por definición de convergencia vaga, existe n_0 tal que para $n \geq n_0$

$$|\mu_n(a_i, b_j) - \mu(a_i, b_j)| < \varepsilon$$

así, notemos que

$$\mu(a_2, b_1) - \varepsilon \leq \mu_n(a_2, b_1) \leq \varepsilon + \mu(a_2, b_1) \qquad \text{y} \qquad \mu(a_1, b_2) - \varepsilon \leq \mu_n(a_1, b_2) \leq \mu(a_1, b_2) + \varepsilon$$

combinando las desigualdades anteriores y por (34), se sigue que

$$\mu(a+\varepsilon, b-\varepsilon) - \varepsilon \leq \mu_n(a,b) \leq \mu(a-\varepsilon), b+\varepsilon) + \varepsilon, \qquad \forall n \geq n_0.$$ ∎

Teorema 3.23. Sean $(\mu_n)_n$ y μ una m.p., entonces (a), (b) y (c) del teorema anterior, son equivalentes a

(35) $\quad (\forall \delta > 0)(\forall \varepsilon > 0)(\exists n_0(\delta,\varepsilon) \in \mathbb{N})(\mu(a+\delta, b-\delta) - \varepsilon \leq \mu_n(a,b) \leq \mu(a-\delta, b+\delta) + \varepsilon \quad \forall n \geq n_0)$.

Demostración:

($\Rightarrow$) Notemos que claramente (35) implica a) del teorema anterior

($\Leftarrow$) Supongamos b), probemos (35), considerando $(\mu_n)_n$ y μ medidas de probabilidad. Denotemos por A al conjunto de átomos de μ, entonces existe un entero ℓ tal que $a_j \in A^c$, $1 \leq j \leq \ell$ y además $a_j \leq a_{j+1} \leq a_j + \delta$ para $1 \leq j \leq \ell - 1$ y además, por hipótesis

$$\mu((a_1, a_\ell)^c) < \varepsilon/4. \tag{36}$$

Así, por (b) existe $n_0(\varepsilon, \ell)$ tal que

$$\sup_{1 \leq j \leq \ell-1} |\mu(a_j, a_{j+1}] - \mu_n(a_j, a_{j+1}]| \leq \varepsilon/4\ell. \tag{37}$$

y, por aditividad

$$|\mu(a_1, a_\ell] - \mu_n(a_1, a_\ell]| < \varepsilon/4$$

 que, junto con (36), nos permite concluir que

$$\mu_n((a_1, a_\ell)^c) < \varepsilon/2 \qquad (38)$$

y, por (36) y (38) ignorando la parte de (a,b) fuera de (a_1, a_ℓ), nosotras cometemos un error menor a $\varepsilon/2$ en cualquiera de las desigualdades de (35). Por lo tanto, es suficiente probar (35) con δ y $\varepsilon/2$ asumiendo que $(a,b) \subseteq (a_1, a_\ell)$. Así, consideremos $a_j \le a \le a_{j+1}$ y $a_k \le b \le a_{k+1}$ donde $0 \le j \le k \le \ell - 1$. Finalmente, la desigualdad buscada se sigue de (37) pues cuando $n \ge n_0$

$$\begin{aligned}\mu_n(a+\delta, b-\delta) - \varepsilon/4 &\le \mu_n(a_{j+1}, a_k) - \varepsilon/4 \le \mu(a_{j+1}, a_k) \le \mu(a,b)\\ &\le \mu(a_j, a_{k+1}) \le \mu_n(a_j, a_{k+1}) + \varepsilon/4\\ &\le \mu_n(a-\delta, b+\delta) + \varepsilon/4.\end{aligned}$$

Teorema 3.24. Sea $(\mu_n)_n$ una sucesión de m.s.p, existe una subsucesión que converge vagamente a μ una m.s.p.

Demostración:

Consideremos la función de subdistribución (f.s.d) F_n, definida por

$$\forall x \quad F_n(x) = \mu_n(-\infty, x]$$

Si μ_n es una medida de probabilidad, entonces F_n es una función de distribución. Además, en general, F_n es creciente, continua a la derecha con $F_n(-\infty) = 0$ y $F_n(+\infty) = \mu_n(\mathbb{R}) \le 1$.

Sea D un conjunto contable y denso en $\mathbb{R}$; sin pérdida de generalidad, considere $D = \{r_1, r_2, \dots\}$. De esta manera, la sucesión $(F_n(r_1))_n$ es acotada y por lo tanto, por el teorema de Bolzano-Weirestrass, existe una subsucesión convergente $(F_{1k})_k$ de la sucesión tal que el límite,

$$\lim_{k\to+\infty} F_{1k}(r_1) = \ell_1,$$

exista; claramente $0 \le \ell_1 \le 1$. Nuevamente, consideremos la sucesión $(F_{1k}(r_2))_k$ la cual sigue siendo acotada y por lo tanto, existe una subsucesión $(F_{2k})_k$ de $(F_{1k})_k$ tal que

$$\lim_{k\to+\infty} F_{2k}(r_2) = \ell_2$$

con $0 \le \ell_2 \le 1$. Notemos que, como $(F_{2k})_k$ es una subsucesión de $(F_{1k})_k$, que converge a ℓ_1 en r_1, entonces F_{2k} también converge a ℓ_1 en r_1. Repitiendo el proceso, se tiene que

$$\begin{array}{llllll} F_{11} & F_{12} & F_{13} & \cdots & F_{1k} \cdots & \text{converge en } r_1 \\ F_{2k} & F_{22} & F_{23} & \cdots & F_{2k} \cdots & \text{converge en } r_1, r_2 \\ \vdots & \vdots & & \ddots & \vdots & \\ F_{j1} & F_{j2} & \cdots & \cdots & F_{jk} \cdots & \text{converge en } r_1, r_2, \dots, r_j \\ \vdots & \vdots & & & \vdots & \end{array}$$

Así, consideremos la sucesión diagonal $(F_{kk})_k$, notemos que esta converge en todo r_j, $j \ge 1$, es decir $F_{kk}(r_j) \to \ell_j$; en efecto, sea r_j con $j \in \mathbb{N}$, notemos que para los primeros $j-1$ términos, la sucesión $(F_{kk})_k$ es una subsucesión de $(F_{jk})_k$ donde $F_{jk}(r_j) \to \ell_j$ como se quería.

Así, hemos probado la existencia de una subsucesión (n_k) y una función G creciente en D tal que

$$\forall r \in D \quad \lim_{k\to+\infty} F_{n_k}(r) = G(r). \qquad (39)$$

A partir de la función G, podemos definir la función F en $\mathbb{R}$ de la siguiente manera

$$\forall x \in \mathbb{R}: F(x) = \inf_{\substack{x<r\\ r\in D}} G(r)$$

la cual es creciente y continua a la derecha (ver (vii) sec 1.1 [3]). Ahora, definamos C como el conjunto de los puntos de continuidad de F (es decir, donde F es continua); notemos que C es denso en $\mathbb{R}$ y mostremos que

$$\forall x \in C \quad \lim_{k\to+\infty} F_{n_k}(x) = F(x).$$

Sean $x \in C$ y $\varepsilon > 0$, cualesquiera, entonces existen r, r' y $r'' \in D$ tales que $r < r' < x < r''$ y tales que $F(r'') - F(r) \le \varepsilon$. De donde

$$F(r) \le G(r') \le F(x) \le G(r'') \le F(r'') \le F(r) + \varepsilon$$

de donde, por (39), se sigue que

$$F(r) \le F_{n_k}(r') \le F_{n_k}(x) \le F_{n_k}(r'') \le F(r'') + \varepsilon \qquad k \to +\infty$$

y así, como $\varepsilon > 0$ fue arbitrario, entonces a F le corresponde una única medida de subprobabilidad tal que $F(x) - F(-\infty) = \mu(-\infty, x]$. Más aún, se tiene que

$\forall a \in C, \ \forall b \in C, \ a<b: \ \lim_{k \to +\infty} \mathcal{U}_{n_k}(a,b] = \mathcal{U}(a,b]$,

(i.e.) $\mathcal{U}_{n_k} \xrightarrow{v} \mathcal{U}$.

Diremos que F_n converge vagamente a F y escribimos $F_n \xrightarrow{v} F$ para $\mathcal{U}_n \xrightarrow{v} \mathcal{U}$ donde $\mathcal{U}_n$ y $\mathcal{U}$ son m.s.p. correspondientes a las f.s.d. F_n y F.

Teorema 3.25. Sea $(\mathcal{U}_n)_n$ una sucesión de m.s.p, si $\mathcal{U}_{n_k} \xrightarrow{v} \mathcal{U}$ para cada subsucesión de $(\mathcal{U}_n)$, entonces $\mathcal{U}_n \xrightarrow{v} \mathcal{U}$

Demostración:

Para la demostración de este teorema, vamos a proceder por la contrapositiva.

Supongamos que $\mathcal{U}_n \not\xrightarrow{v} \mathcal{U}$. Así, por el (teorema 3.22) existe un intervalo de continuidad (a,b) tal que $\mathcal{U}_n(a,b) \not\to \mathcal{U}$. Por Bolzano - Weierstrass, existe una subsucesión $(\mathcal{U}_{n_k})$ tal que

$$\mathcal{U}_{n_k}(a,b) = L \neq \mathcal{U}(a,b).$$

Por el (teorema 3.24) la sucesión $(\mathcal{U}_{n_k})_k$ contiene una subsucesión $(\mathcal{U}_{n_k'})_k$ que converge vagamente a $\mathcal{U}$. Nuevamente por el (teorema 3.22)

$$\mathcal{U}_{n_k'}(a,b) \longrightarrow \mathcal{U}(a,b)$$

pero como $\mathcal{U}_{n_k'}(a,b) \longrightarrow L$, entonces $L = \mathcal{U}(a,b)$ lo cual no es posible y por contrapositiva completa nuestra prueba. ■

Continuemos con la discusión de otro tipo de criterios más populares en teoría de la medida y análisis funcional. Estos criterios tienen relación con las clases de funciones continuas en $\mathbb{R}$.

- C_K = el conjunto de las funciones continuas con soporte compacto
- C_0 = el conjunto de las funciones continuas tales que $\lim_{|x| \to +\infty} f(x) = 0$;
- C_B = el conjunto de las funciones continuas acotadas.
- C = el conjunto de las funciones continuas.

Así, es claro que $C_K \subseteq C_0 \subseteq C_B \subseteq C$; además C_0 es la clausura de C_K respecto a la convergencia uniforme.

Definición 3.26. Sea f una función definida en un espacio arbitrario, decimos que f tiene soporte en algún subconjunto S del espacio si y solo si se anula fuera de S.

En el caso de que sea soporte compacto, tiene soporte en algún subconjunto compacto.

Lema de aproximación [5] Sea $f \in C_K$ con soporte en el intervalo compacto $[a,b]$. Dado cualquier subconjunto denso A de $\mathbb{R}$, y $\varepsilon > 0$, entonces existe una función simple con valores en A f_ε en (a,b) tal que

$$\sup_{x \in \mathbb{R}} |f(x) - f_\varepsilon(x)| \leq \varepsilon.$$

Si $f \in C_0$, el resultado es igual si se reemplaza (a,b) por $\mathbb{R}$.

Teorema. 3.27. Sean $(\mathcal{U}_n)_n$ y $\mathcal{U}$ m.s.p., entonces $\mathcal{U}_n \xrightarrow{v} \mathcal{U}$ si y solo si

$$\forall f \in C_K (o \in C_0) \quad \int_{\mathbb{R}} f(x)\,\mathcal{U}_n(dx) \longrightarrow \int_{\mathbb{R}} f(x)\,\mathcal{U}(dx). \qquad (40)$$

Demostración

$\Rightarrow$) Supongamos que $\mathcal{U}_n \xrightarrow{v} \mathcal{U}$. Notemos entonces que (40) es cierta si consideramos f la función indicatriz sobre $(a,b]$ para $a,b \in D$, donde D es el conjunto de la (definición 3.20). Así, por linealidad de la integral, (40) también es cierta cuando f es cualquier función simple definida en D.

Ahora, sean $f \in C_0$ y $\varepsilon > 0$, cualesquiera; por el lema de aproximación, existe una función simple definida en D f_ε tal que

$$\sup_{x \in \mathbb{R}} |f(x) - f_\varepsilon(x)| \leq \varepsilon.$$

$$\left| \int f\,d\mu_n - \int f\,d\mu \right| \leq \left| \int (f-f_\varepsilon)\,d\mu_n \right| + \left| \int f_\varepsilon\,d\mu_n - \int f_\varepsilon\,d\mu \right| + \left| \int (f_\varepsilon - f)\,d\mu \right|$$

Así, notemos que para el primer término

$$\left| \int (f-f_\varepsilon)\,d\mu_n \right| \leq \int |f-f_\varepsilon|\,d\mu_n \leq \varepsilon \int d\mu_n \leq \varepsilon,$$

similarmente, es posible acotar el último término por ε; para el segundo término, notemos que cuando $n \to +\infty$, dicho término tiende a cero; por lo tanto

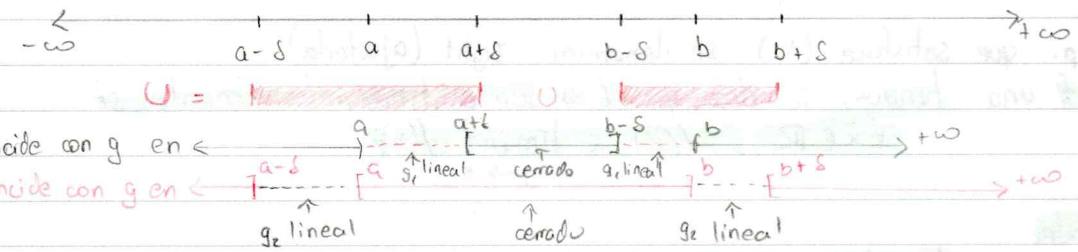
$$\left| \int f\,d\mu_n - \int f\,d\mu \right| \leq 2\varepsilon$$

y dado que $\varepsilon > 0$ fue arbitrario, entonces $\int f\,d\mu_n \to \int f\,d\mu$.

Recíprocamente, si (40) es verdadero para $f \in C_K$, sea A el conjunto de átomos de μ mencionado en la demostración del (teorema 3.23), vamos a mostrar que la convergencia vaga se mantiene si $D = A^c$. Sea $g = \mathbb{1}_{(a,b]}$ la función indicatriz de $(a,b]$ con $a,b \in D$, entonces, dado $\varepsilon > 0$, existe $\delta(\varepsilon) > 0$ tal que $a+\delta < b-\delta$ y tal que $\mu(U) < \varepsilon$ donde

$$U = (a-\delta, a+\delta) \cup (b-\delta, b+\delta).$$

Ahora, definamos g_1 como la función que coincide con g en $(-\infty, a] \cup [a+\delta, b-\delta] \cup [b, +\infty)$ y que es linear en $(a, a+\delta)$ y en $(b-\delta, b)$. Análogamente, definimos g_2 como la función que coincide con g en $(-\infty, a-\delta] \cup [a,b] \cup [b+\delta, +\infty)$ y que es linear en $(a-\delta, a)$ y en $(b, b+\delta)$.



Así, se sigue que $g_1 \leq g \leq g_2 \leq g_1 + 1$ y en consecuencia

$$\int g_1\,d\mu_n \leq \int g\,d\mu_n \leq \int g_2\,d\mu_n, \qquad (41)$$

de donde como $g_1, g_2 \in C_K$ por hipótesis, se tiene que de la desigualdad anterior

$$\int g_1\,d\mu \leq \int g\,d\mu \leq \int g_2\,d\mu,$$

notemos que únicamente los términos extremos convergen. Ahora, como $g_1 \leq g \leq g_2 \leq g+1$, se sigue

$$\int g_2\,d\mu - \int g_1\,d\mu \leq \int_U 1\,d\mu = \mu(U) < \varepsilon$$

y, puesto que $\varepsilon > 0$ fue arbitrario, entonces el término medio en (41) también converge es decir

$$\int g\,d\mu_n \to \int g\,d\mu$$

de donde, por definición de g

$$\mu_n(a,b] \xrightarrow{v} \mu(a,b].$$

Corolario 3.28. Sea $(\mu_n)_n$ una sucesión de m.s.p. tal que para todo $f \in C_K$

$$\lim_{n \to +\infty} \int_{\mathbb{R}} f(x)\,\mu_n(dx)$$

existe, entonces $(\mu_n)_n$ converge vagamente.

Esquema de demostración

Por el (teorema 3.24) $(\mu_n)_n$ admite una subsucesión convergente vagamente digamos a μ, por el (teorema 3.27) el límite superior es igual a $\int_{\mathbb{R}} f(x)\,\mu(x)$. Notemos que esta cota encontrada debería ser la misma para toda subsucesión que converge vagamente y así el corolario se sigue por por el (teorema 3.25)

Teorema. 3.29. Sea $(\mu_n)_n$ una sucesión de medidas de subprobabilidad, entonces $\mu_n \xrightarrow{v} \mu$ si y solo si

$$\forall f \in C_B: \int_{\mathbb{R}} f(x)\,\mu_n(dx) \to \int_{\mathbb{R}} f(x)\,\mu(dx).$$

La demostración de este resultado corresponde al teo 4.4.2 [3].

Teorema . 3.30. Sea $\{\mu_\alpha\}_{\alpha \in A}$ una familia de m.p. Para que cada sucesión de dicha familia tenga una subsucesión que converga vagamente a una medida de probabilidad es suficiente y necesario que para todo $\varepsilon > 0$ exista un intervalo finito tal que

$$\inf_{\alpha \in A} \mu_\alpha(I) > 1 - \varepsilon. \qquad (42)$$

Demostración
$\Rightarrow$) Supongamos (42). Así, dada una sucesión de m.p. $(\mu_n)_n$ de la familia $\{\mu_\alpha\}_{\alpha\in A}$, existe una subsucesión tal que $\mu_{n_k} \xrightarrow{n} \mu$, vamos a mostrar que μ es una m.p. Sea J un intervalo de continuidad de μ tal que $J \subseteq I$, entonces

$$\mu(\mathbb{R}) \geq \mu(J) = \lim_{k\to+\infty} \mu_{n_k}(J) \geq \limsup_{k\to+\infty} \mu_{n_k}(I) \geq 1-\varepsilon,$$

puesto que $\varepsilon>0$ es arbitrario, entonces, si $\varepsilon \to 0$, se concluye que $\mu(\mathbb{R})=1$, como se quería.

$\Leftarrow$) Por contradicción, supongamos que (42) no se cumple; así existe $\varepsilon>0$ y sea I_n una sucesión de intervalos finitos I_n creciente en $\mathbb{R}$ y una sucesión $(\mu_n)_n$ de la familia $\{\mu_\alpha\}_{\alpha\in A}$ tal que

$$\forall n: \quad \mu_n(I_n) \leq 1-\varepsilon,$$

así, por hipótesis, existen $(\mu_{n_k})_k$ y μ, tales que $\mu_{n_k} \longrightarrow \mu$. Así, tomando nuevamente un intervalo de continuidad de μ, entonces $J \subseteq I_n$ para algún n suficientemente grande y por lo tanto

$$\mu(I) = \lim_{k\to+\infty} \mu_{n_k}(J) \leq \liminf_{k\to+\infty} \mu_{n_k}(I_n) \leq 1-\varepsilon$$

Es decir, $\mu(\mathbb{R}) \leq 1-\varepsilon$, pero por hipótesis, μ es una medida de probabilidad. Así, por contradicción se sigue el resultado deseado.

A una familia de m.p. que satisface (42) se denomina tight (ajustada)

Definición 3.31. Sea f una función, se dice que f es semicontinua inferiormente si

$$\forall x \in \mathbb{R}, \quad f(x) \leq \liminf_{y\to x} f(y) \tag{43}$$

donde

$$\liminf_{y\to x} f(y) = \sup_{r>0} \inf \{ f(y) : |x-y|<r, \ x\neq y \}.$$

Análogamente, una función es semicontinua superiormente si

$$\forall x \in \mathbb{R}: \quad f(x) \geq \limsup_{y\to x} f(y) \tag{44}$$

donde

$$\limsup_{y\to x} f(y) = \inf_{r>0} \sup \{ f(y) : |x-y|<r, \ x\neq y \}.$$

Proposición 3.32. a) f es acotada y semicontinua inferiormente si y solo si existe una sucesión $f_k \in C_B$ tal que $f_k \uparrow f$.
b) f es semicontinua superiormente si y solo si $-f$ es semicontinua inferiormente.

Notación: Vamos a notar por L el conjunto de las funciones semicontinuas inferiormente y, por U a las funciones semicontinuas superiormente.

Teorema 3.34. Sean $(\mu_n)_n$ y μ m.p., entonces $\mu_n \xrightarrow{v} \mu$ si y solo si se satisface una de las siguientes condiciones

a) $$\forall f \in L: \ \liminf_n \int f(x)\,\mu_n(dx) \geq \int f(x)\,\mu(dx)$$

b) $$\forall g \in U: \ \limsup_n \int g(x)\,\mu_n(dx) \leq \int g(x)\,\mu(dx)$$

Demostración:
Notemos que ambas condiciones son equivalentes, en efecto, tomando $f=-g$, se sigue a) a partir de b) y viceversa.

$\Rightarrow$) Supongamos que $\mu_n \xrightarrow{v} \mu$ y sea $f_k \in C_B$ tal que $f_k \uparrow f$, entonces, tenemos que

$$\liminf_n \int f(x)\,\mu_n(dx) \geq \lim_n \int f_k(x)\,\mu_n(dx) = \int f_k(x)\,\mu(dx),$$

por el (teorema 3.29). Así, si $k \to +\infty$

$$\int f_k(x)\,\mu d(x) \longrightarrow \int f(x)\,\mu d(x)$$

Recíprocamente, supongamos que a) y b) son ciertas y sea $\varphi \in C_B$, entonces $\varphi \in L$ y $\varphi \in U$ y por lo tanto

$$\int \varphi(x)\mu(dx) \leq \liminf_n \int \varphi(x)\mu_n(dx) \leq \limsup_n \int \varphi(x)\mu_n(dx) \leq \int \varphi(x)\mu(dx)$$

de donde, se sigue que

$$\int \varphi(x)\mu_n(dx) \to \int \varphi(x)\mu(dx)$$

y, nuevamente por el (teorema 3.29), $\mu_n \xrightarrow{v} \mu$

Observación; el teorema anterior se mantiene si f es semicontinua con $f \in C$ en el dominio de la función, pero acotada inferiormente.

Corolario 3.35. Las condiciones a) y b) del teorema anterior, se pueden reemplazar por

a') Para todo abierto O: $\liminf_n \mu_n(O) \geq \mu(O)$

b') Para todo cerrado C: $\limsup_n \mu_n(C) \leq \mu(C)$

Convergencia en distribución

Definición 3.36. Sea $(X_n)_n$ una sucesión de v.a., se dice que converge en distribución a F si y solo la sucesión $(F_n)_n$ donde $F_n = P \circ X_n^{-1}$ converge vagamente a la f.d. F.

Si X es una v.a. con f.d. F, se dice, como abuso de lenguaje, que X_n converge en distribución a X.

Teorema 3.37. Sean (F_n), F f.d de las v.a. $(X_n)_n$, X. Si $X_n \xrightarrow{P} X$, entonces $F_n \xrightarrow{v} F$ (i.e. convergencia en probabilidad implica convergencia en distribución).

Demostración:

Si $X_n \xrightarrow{P} X$, entonces para cada $f \in C_K$, $f(X_n) \to f(X)$ (ver Corolario 3.8). Y, como f es acotada, entonces, por el (teorema 3.10) $f(X_n) \xrightarrow{L^1} f(X)$, es decir

$$E f(X_n) \to E f(X) \qquad \forall f \in C_K$$

de donde, se deduce (40) y por lo tanto $F_n \xrightarrow{v} F$

Teorema 3.38. Sea $(X_n)_n$ una sucesión de v.a. y $c \in \mathbb{R}$. Si $X_n \xrightarrow{d} c$, entonces $X_n \xrightarrow{P} c$

Demostración:

Para $c \in \mathbb{R}$, notemos que

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

y, como $X_n \xrightarrow{d} c$, entonces para $\varepsilon > 0$

$$\lim_n F_n(c+\varepsilon/2) = F(c+\varepsilon/2) = 1 \quad \text{y} \quad \lim_n F_n(c-\varepsilon) = F(c-\varepsilon) = 0.$$

Así, sea $\varepsilon > 0$, cualquiera, se tiene que

$$\lim_n P(|X_n - c| > \varepsilon) = \lim_n P(-\varepsilon < X_n - c < \varepsilon) = \lim_n P(-\varepsilon + c \leq X_n \leq \varepsilon + c)$$

$$= \lim_n \left(P(X_n > c-\varepsilon) + P(X_n \leq \varepsilon + c)\right)$$

$$= \lim_n \left[\left(1 - P(X_n \leq c-\varepsilon)\right) + F_X(\varepsilon + c)\right)$$

$$= \lim_n \left[\left(1 - F_n(c-\varepsilon)\right]\right.$$

¡Terminar!

Teorema 3.39. Si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} 0$, entonces

a) $X_n + Y_n \xrightarrow{d} X$

b) $X_n Y_n \xrightarrow{d} 0$

Demostración:

a) Sea $f \in C_k$, con $|f| \leq M$. Como f es uniformemente continua, entonces para $\varepsilon > 0$, existe $\delta > 0$ tal que

$$|x-y| \leq \delta \Rightarrow |f(x) - f(y)| < \varepsilon.$$

Por lo tanto,

$$E|f(X_n+Y_n) - f(X_n)| \leq \varepsilon P\{|f(X_n+Y_n) - f(X_n)| \leq \varepsilon\} + 2M P(\{|f(X_n+Y_n) - f(X_n)| > \varepsilon\}$$
$$\leq \varepsilon + 2M \underbrace{P(|Y_n| > \delta)}_{\text{por contrarecíproco}}$$

de donde cuando $n \to +\infty$, entonces

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

y por el (teorema 3.10), se sigue que $X_n + Y_n \xrightarrow{d} 0$.

b) Sea $\varepsilon > 0$, tomemos A_0 tal que $\pm A_0$ son puntos de continuidad de la f.d. de X

$$\lim_{n\to+\infty} P(|X_n| > A_0) = P(|X| > A_0) < \varepsilon.$$

Es decir, $P(|X_n| > A_0) < \varepsilon$ para todo $n \geq n_0(\varepsilon)$. Más aún, tomemos $A \geq A_0$ tal que la desigualdad se mantenga para $n \leq n_0(\varepsilon)$. Así,

$$P(|X_n Y_n| > \varepsilon) \leq P(|X_n| > A) + P(|Y_n| > \varepsilon/A) \leq \varepsilon + P(|Y_n| > \varepsilon/A)$$

y, cuando $n \to +\infty$, se sigue que

$$P(|X_n Y_n| > \varepsilon) = 0$$

Así, $X_n Y_n \xrightarrow{P} 0$ y por el (teorema 3.37), se sigue el resultado deseado.

Corolario 3.40. Si $X_n \xrightarrow{d} X$, $\alpha_n \xrightarrow{d} \alpha$ y $\beta_n \xrightarrow{d} \beta$, con $\alpha, \beta \in \mathbb{R}$, entonces

$$\alpha_n X_n + \beta_n \xrightarrow{d} \alpha X + \beta$$

Integrabilidad uniforme, convergencia de momentos.

Notemos que la función $|x|^r$, $r > 0$ es continua pero no acotada; así, no es posible aplicar el (teorema 3.27).

Ejemplo: Convergencia casi segura no implica convergencia en L^p.

Consideremos

$$X_n(\omega) = \begin{cases} 2^n & \text{si } \omega \in (0, 1/n) \\ 0 & \text{caso contrario.} \end{cases}$$

Así, notemos que

$$E|X_n|^p = \int_0^{1/n} |2^n|^p \mu(dx) = 2^{np}(1/n - 0) = \frac{2^{np}}{n} \to +\infty$$

para cada $p > 0$, pero $X_n \to 0$ en todas partes.

Notemos que en el ejemplo anterior, la convergencia casi segura no implica la convergencia de cualquier momento de orden $r > 0$.

Una modificación del ejemplo anterior, sería que si $X_n \xrightarrow{c.s.} X$ $E|X_n|^r = 1$ pero $E|X|^r = 0$.

Es útil tener condiciones para garantizar la convergencia de momentos cuando se tiene la convergencia casi segura.

Teorema 3.41 Sea $(X_n)_n$ una sucesión de v.a. tal que $X_n \xrightarrow{c.s.} X$, con X v.a., entonces para todo $r > 0$

$$E|X^r| \leq \liminf_{n\to+\infty} E|X_n|^r.$$

Si $X_n \to X$ en L^r y $X \in L^r$, entonces $E|X_n|^r \to E|X|^r$ pág 99 [3]

Demostración:
Notemos que la demostración de este resultado es solo un caso del lema de Fatou (ver sección 2).
Así, tenemos que

$$E|X|^r = \int_\Omega |X|^r dP = \int_\Omega \lim_n |X_n|^r dP = \int_\Omega \liminf_n |X_n|^r dP \leq \liminf_n \int_\Omega |X_n|^r dP$$

Por lo tanto,

$$E|X|^r \leq \liminf_n E|X_n|^r.$$

Ahora, si $X_n \xrightarrow{L^r} X$, para $r \geq 1$; por la desigualdad de Minkowski y como $X = X_n + (X - X_n) = X_n - (X_n - X)$; entonces

$$E(|X_n|^r)^{1/r} - E(|X_n - X|^r)^{1/r} \leq E(|X|^r)^{1/r} \leq E(|X_n|^r)^{1/r} + E(|X_n - X|^r)^{1/r}.$$

Así, cuando $n \to +\infty$, entonces

$$E(|X_n|^r)^{1/r} \to E(|X|^r)^{1/r}$$

Ahora, para $0 < r < 1$, sabemos que

$$|x+y|^r \leq |x|^r + |y|^r \qquad \forall x, y \in \mathbb{R}$$

De donde, nuevamente

$$E|X_n|^r - E|X - X_n|^r \leq E|X|^r \leq E|X_n|^r + E|X - X_n|^r$$

y se sigue la conclusión deseada

Teorema. 3.42. Sea $(X_n)_n$ una sucesión de v.a. tal que $X_n \xrightarrow{d} X$ y para algún $p > 0$, $\sup_n E|X_n|^p = M < +\infty$, entonces para $r < p$, cualquiera:

teo 4.5.2 [3]

$$\lim_{n \to +\infty} E|X_n|^r = E|X|^r.$$

En particular, si r es un entero positivo, entonces podemos reemplazar $|X_n|^r$ y $|X|^r$ por X_n^r y X^r respectivamente.

Demostración
Para la demostración de este teorema vamos a considerar únicamente el segundo caso, debido a que la demostración del primero es similar. Sean $(F_n)_n$ y F f.d asociadas a $(X_n)_n$ y X respectivamente. Para todo $A > 0$, definimos $f_A : \mathbb{R} \to \mathbb{R}$ como

$$f_A(x) = \begin{cases} x^r & \text{si } |x| \leq A \\ A^r & \text{si } x > A \\ (-A)^r & \text{si } x < -A, \end{cases}$$

Así, es claro que $f_A \in C_B$. Así, por el (teorema 3.27) los «momentos truncados» convergen:

$$\int_{-\infty}^{+\infty} f_A(x)\, dF_n(x) \to \int_{-\infty}^{+\infty} f_A(x)\, dF(x).$$

Luego, se sigue que

$$\int_{-\infty}^{+\infty} |f_A(x) - x^r|\, dF_n(x) \leq \int_{|x|>A} |x|^r dF_n(x) = \int_{|X_n|>A} |X_n|^r dP \leq \frac{1}{A^{p-r}} \int_\Omega |X_n|^p dP \leq \frac{M}{A^{p-r}}.$$

Así, cuando $A \to +\infty$, entonces $M/A^{p-r} \to 0$ y por lo tanto $\int_{-\infty}^{+\infty} f_A\, dF_n$ converge uniformemente a $\int_{-\infty}^{+\infty} x^r\, dF$.

Luego,

$$\int_{-\infty}^{+\infty} x^r\, dF = \lim_{A\to+\infty} \int_{-\infty}^{+\infty} f_A\, dF = \lim_{A\to+\infty} \lim_{n\to+\infty} \int_{-\infty}^{+\infty} f_A\, dF_n$$

$$= \lim_{n\to+\infty} \lim_{A\to+\infty} \int_{-\infty}^{+\infty} f_A\, dF_n = \lim_{n\to+\infty} \int_{-\infty}^{+\infty} x^r\, dF_n.$$

de donde, si $F_n(x) = P(X_n \leq x)$ y $F(x) = P(X \leq x)$, entonces se ha probado que

$$E|X_n|^r \longrightarrow E|X|^r$$

Definición 3.43. Una familia de v.a. $(X_t)_{t\in T}$ con T un conjunto de índices arbitrarios, se dice que es uniformemente integrable si y solo si

$$\lim_{A\to+\infty} \int_{|X_t|>A} |X_t|\, dP = 0 \qquad (45)$$

uniformemente en $t\in T$.

Teorema 3.44. Una familia $\{X_t\}_{t\in T}$ es uniformemente integrable si y solo si las siguientes condiciones se satisfacen.

a) $E|X_t|$ es acotada en $t\in T$

b) Para todo $\varepsilon>0$, existe $\delta=\delta(\varepsilon)>0$ tal que para todo $E\in\mathcal{F}$

$$P(E)<\delta \Rightarrow \int_E |X_t|\, dP < \varepsilon \qquad \forall t\in T$$

Demostración

$\Rightarrow$ Notemos que claramente, (45) implica a)

Luego, sea $E\in\mathcal{F}$, y denotemos $E_t = \{\omega : |X_t(\omega)| > A\}$. Así, por el teorema del valor medio integral

$$\int_E |X_t|\, dP = \left(\int_{E\cap E_t} + \int_{E\setminus E_t}\right)|X_t|\, dP \leq \int_{E_t} |X_t|\, dP + A\,P(E). \qquad (46)$$

Así, sea $\varepsilon>0$, cualquiera, entonces existe $A=A(\varepsilon)$ tal que la última integral es menor que $\varepsilon/2$ para todo $t\in T$, por (45). Así, b) se sigue si tomamos $\delta = \varepsilon/2A$. En efecto, si $P(E)<\delta = \varepsilon/2A$, por (46)

$$\int_E |X_t|\, dP \leq \int_{E_t} |X_t|\, dP + A\left(\frac{\varepsilon}{2A}\right) = \frac{\varepsilon}{2} + \underbrace{\int_{E_t} |X_t|\, dP \leq \frac{\varepsilon}{2}}_{\text{por (45)}} + \frac{\varepsilon}{2} = \varepsilon$$

$\Leftarrow$) Supongamos que se tiene a) y b). Luego, por la desigualdad de Chebyshev, se tiene que, para $t\in T$

$$P(|X_t|>A) \leq \frac{E|X_t|}{A} \leq \frac{H}{A},$$

donde H satisface a). Así, si $A>H/\delta \Rightarrow \delta > H/A$, entonces $P(E_t)<\delta$ entonces por b)

$$\int_{E_t} |X_t|\, dP < \varepsilon$$

y así, (45) es cierto, lo que completa la demostración.

Teorema. 4.5.4. Sea $r>0$, $X_n\in L^r$, y $X_n \xrightarrow{P} X$, entonces las siguientes proposiciones son equivalentes

a) $(|X_n|^r)$ es uniformemente integrable.

b) $X_n \xrightarrow{L^r} X$

c) $E|X_n|^r \longrightarrow E|X|^r$.

a) $\Rightarrow$ b)

Puesto que $X_n \xrightarrow{P} X$, por el (teorema 3.6), existe una subsucesión $(X_{n_k})_k$ tal que X_{n_k} converge c.s. a X. Luego, por el (teorema 3.41) y a) $X \in L^r$. Luego, consideremos la siguiente desigualdad

$$|X_n - X|^r \leq 2^r \{ |X_n|^r + |X|^r \}$$

para todo $r>0$, y junto con el (teorema 3.44), entonces la sucesión $(|X_n - X|^r)_n$ también es uniformemente integrable.

Por otro lado, para $\varepsilon > 0$, tenemos que

$$\int_\Omega |X_n - X|^r dP = \int_{|X_n - X| > \varepsilon} |X_n - X|^r dP + \int_{|X_n - X| \leq \varepsilon} |X_n - X|^r dP$$

$$\leq \int_{|X_n - X| > \varepsilon} |X_n - X|^r dP + \varepsilon^r.$$

luego, com por hipotesis $P(|X_n - X| > \varepsilon) \longrightarrow 0$ cuando $n \to +\infty$. Así, por b) del teorema anterior la última integral tiende a 0 para $\varepsilon > 0$, cualquiera. Así, se ha probado b)

b) $\Rightarrow$ c) Supongamos b), entonces por b) del (teorema 3.41) se sigue el resultado.

c) $\Rightarrow$ a)

Sea $A > 0$, cualquiera. Consideremos la siguiente función $f_A \in C_K$ tal que

$$f_A(x) = \begin{cases} = |x|^r & \text{para } |x|^r \leq A \\ \leq |x|^r & \text{para } A \leq |x|^r \leq A+1 \\ = 0. & \end{cases}$$

Por lo tanto, se tiene que

$$\liminf_{n \to +\infty} \int_{|X_n|^r \leq A+1} |X_n|^r dP \geq \lim_{n \to +\infty} E[f_A(X_n)] = E[f_A(X)] \geq \int_{|X|^r \geq A} |X|^r dP,$$

debido a la definición de f_A y el límite se deduce del (teorema 3.37). Y restando del límite de c), se sigue que

$$\limsup_{n \to +\infty} \int_{|X_n|^r > A+1} |X_n|^r dP \leq \int_{|X|^r > A} |X|^r dP.$$

Así, notemos que la última integral converge a cero. Por lo tanto, para todo $\varepsilon > 0$, existe $A_0 = A_0(\varepsilon)$ y $n_0 = n_0(A(\varepsilon))$ tal que

$$\sup_{n > n_0} \int_{|X_n|^r > A+1} |X_n|^r dP < \varepsilon,$$

cuando $A > A_0$. Luego, como cada $|X_n|^r$ es integrable, existe $A_1 = A_1(\varepsilon)$ tal que el último supremo tomado con $n \geq 1$ nos permite concluir que es $A > A_0 \vee A_1$, lo que completa la demostración.

Capítulo 4: Ley de los grandes números

Varios de los conceptos analizados en el capítulo anterior los vamos a aplicar en lo que se conoce como la Ley de los grandes números. Para ello, definamos las sumas parciales

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

de una sucesión de variables aleatorias $(X_n)_{n\in\mathbb{N}}$. La formulación débil o fuerte de la ley nos dice que

$$\frac{S_n - E[S_n]}{n} \to 0 \qquad (1)$$

en probabilidad o casi-seguramente, claramente, cuando $E[S_n]$ es finita. Una generalización natural de (1) está dada por

$$\frac{S_n - a_n}{b_n} \longrightarrow 0,$$

donde $(a_n)_n$ es una sucesión de números reales y $(b_n)_n$ una sucesión de numeros positivos que tiende a infinito.

Uno de los primeros casos se sigue del (teorema 3.10) y (teorema 3.6) de los cuales, si $(Z_n)_n$ es una sucesión de v.a., entonces $E[Z_n^2] \to 0$ implica que $Z_n \xrightarrow{P} 0$ y así, $Z_{n_k} \xrightarrow{c.s.} 0$ para $\{n_k\}$ una sucesión creciente. Aplicando $Z_n = S_n/n$, entonces aplicando el razonamiento anterior

$$E[S_n^2] = o(n^2) \quad (\text{intuitivamente } n^2 \text{ crece más rápido que } E[S_n^2])$$

entonces

$$S_n/n \xrightarrow{P} 0. \qquad (2)$$

Podemos calcular $E[S_n^2]$ de manera más explícita de la siguiente manera

$$E[S_n^2] = E\left(\left(\sum_{j=1}^n X_j\right)^2\right) = E\left(\sum_{j=1}^n X_j^2 + 2\sum_{1\le j\le k\le n} X_j X_k\right) = \sum_{j=1}^n E[X_j^2] + 2\sum_{1\le j\le k\le n} E[X_j X_k]$$

Notemos que la expresión anterior posee n^2 términos. Así, incluso si todos los términos están acotados, entonces $E[S_n] = O(n^2)$ lo que haría que (2) no se cumpla pues el cociente no convergería necesariamente. Por lo tanto, la idea es introducir hipótesis para poder cancelar algunos de los «términos mezclados».

Definición 4.1. Sean X y Y v.a., se dice que no están relacionadas si tienen segundo momento finito ($E[X^2]$, $E[Y^2]$ son finitos) y

$$E[XY] = E[X]E[Y] \qquad (3)$$

Además, se dicen ortogonales, si

$$E[XY] = 0. \qquad (4)$$

Una familia $\{X_n\}_n$ se dice no relaciona [ortogonal] si es dos a dos no relaciona [ortogonal].

Notemos que (3) es equivalente a

$$E[X - E[X]]\,E[Y - E[Y]] = 0$$

y así, si $E[X] = E[Y] = 0$, entonces se obtiene (4).

Teorema 4.2. Sea $(X_n)_n$ v.a. no relacionadas con segundo momento finito acotado, entonces (1) converge en L^p y por lo tanto en Probabilidad

En efecto, aplicando la desigualdad de Chevychev, para $\varepsilon > 0$, cualquiera.

$$P\left(\left|\frac{S_n}{n}\right| > \varepsilon\right) \le V(S_n/n)/\varepsilon^2 \le \frac{E[S_n^2]}{n^2\varepsilon^2}$$

y, si $E[S_n^2] = o(n^2)$, entonces $S_n/n \xrightarrow{P} 0$. [ver más en pág.108 [3]]

Teorema. 4.3. Bajo las mismas hipótesis del (teorema 4.2), (1) se mantiene y más aún converge casi-seguramente.

Demostración:

① sin pérdida de generalidad, supongamos que $E[X_j] = 0$ para cada j. Así, $\{X_j\}_j$ es ortogonal

Así, se sigue que, como $\sigma^2(S_n) = \sum_{j=1} \sigma^2(X_j)$, entonces

$$E[S_n^2] \leq Hn$$

donde H es la cota de los segundos momentos. Así, por la desigualdad de Chevychev para $\varepsilon > 0$, cualquiera

$$P(|S_n| \geq n\varepsilon) \leq Hn/(n^2\varepsilon^2) = H/n\varepsilon^2.$$

Así, sumando sobre n, se tiene la serie armónica, divergente. No obstante, si consideramos la subsucesión $\{n^2\}$, entonces

$$\sum_{n\in\mathbb{N}} P(|S_{n^2}| > n^2\varepsilon) \leq \sum_{n\in\mathbb{N}} H/n^2\varepsilon^2 < +\infty.$$

Así, por el primer lema de Borel-Cantelli,

$$P(|S_{n^2}| > n^2\varepsilon \ : i.o.) = 0.$$

Es decir, se tiene que $S_{n^2}/n^2 \xrightarrow{c.s.} 0$. Notemos que haci, hemos probado el resultado para una subsucesión. Para probar el caso general, vamos a probar que S_k no es muy diferente de S_{n^2}. Así, para $n \geq 1$, definamos

$$D_n = \max_{n^2 \leq k \leq (n+1)^2} |S_k - S_{n^2}|.$$

Así, se sigue que

$$E[D_n^2] \leq 2n\, E[|S_{(n+1)^2} - S_{n^2}|^2] = 2n \sum_{j=n^2+1}^{(n+1)^2} \sigma^2(X_j) \leq 4n^2 H$$

y, por la desigualdad de Chevychev

$$P(D_n > n^2\varepsilon) \leq 4H/n^2\varepsilon^2$$

y nuevamente $D_n/n^2 \xrightarrow{c.s.} 0$. Así, se tiene que si $S_{n^2}/n^2 \xrightarrow{c.s} 0$ y $D_n/n^2 \xrightarrow{c.s.} 0$, entonces, se tiene (1) pues

$$\frac{|S_k|}{k} \leq \frac{|S_{n^2}| + D_n}{n^2} \qquad \text{para } n^2 \leq k \leq (n+1)^2$$

y así, se ha probado el resultado deseado.

Sección 4.1. Ley débil de los grandes números

La ley de los grandes números de la sección anterior (1) incluye la hipótesis de segundo momento finito. Así, para debilitar estas hipótesis, podemos introducir un nuevo concepto.

Definición 4.4. Sean $(X_n)_n$ y $(Y_n)_n$ se dicen sucesiones equivalentes de v.a. si

$$\sum_{n\in\mathbb{N}} P(X_n \neq Y_n) < +\infty. \tag{5}$$

Teorema 4.5. Sean $(X_n)_n$ y $(Y_n)_n$ sucesiones de v.a., si son equivalentes, entonces

$$\sum_{n\in\mathbb{N}} (X_n - Y_n) \tag{6}$$

converge casi-seguramente.

Más aún, si existe $(a_n)_n$ tal que $a_n \uparrow +\infty$, entonces

$$\frac{1}{a_n}\sum_{j=1}^{n} (X_j - Y_j) \xrightarrow{c.s.} 0 \tag{7}$$

Demostración.

Por el Lema de Borel-Cantelli, de (5), se sigue que

$$P\{X_n \neq Y_n;\ i.o\} = 0$$

Así, existe N, un conjunto de medida nula tal que si $\omega \in \Omega \setminus N$, existe $n_0 = n_0(\omega)$ tal que

$$n \geq n_0(\varepsilon) \implies X_n(\omega) = Y_n(\omega).$$

Así, se concluye que las sucesiones $(X_n)_n$ y $(Y_n)_n$ difieren únicamente en un número finito de términos. Es decir, la serie $\sum_n X_n - Y_n$, consiste en una serie de ceros a partir de algún punto.

Notemos entonces que ambas afirmaciones del teorema, se concluyen de este razonamiento

Corolario 4.6. Con probabilidad uno, la expresión

$$\sum_{n\in\mathbb{N}} x_n \quad \text{o} \quad \frac{1}{a_n}\sum_{j=1}^{n} x_j$$

converge, diverge a $+\infty$ o $-\infty$ o fluctua de la misma forma que

$$\sum_{n\in\mathbb{N}} Y_n \quad \text{o} \quad \frac{1}{a_n}\sum_{j=1}^{n} Y_j.$$

En particular, si $1/a_n \sum_{j=1}^{n} X_j \xrightarrow{P} X$, entonces $1/a_n \sum_{j=1}^{n} Y_j \xrightarrow{P} X$

Para la segunda afirmación, como convergencia c.s. implica P., entonces la relación (7) se mantiene en probabilidad; es decir

$$\frac{1}{a_n}\sum_{j=1}^{n} X_j \xrightarrow{P} X,$$

Por lo tanto,

$$\frac{1}{a_n}\sum_{j=1}^{n} Y_j = \frac{1}{a_n}\sum_{j=1}^{n} X_j + \frac{1}{a_n}\sum_{j=1}^{n}(Y_j - X_j) \xrightarrow{P} X + 0 = X$$

(Recuerde que $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} Y \Rightarrow X_n + Y_n \xrightarrow{P} X+Y$).

A continuación, consideremos la siguiente ley de los grandes números de Khintche bajo la hipótesis de independencia total

Teorema 4.7. Sea $(X_n)_n$ una s.v.a. independientes dos a dos e idénticamente distribuidas con esperanza finita $E[X_n] = m < +\infty$ para $n \in \mathbb{N}$, entonces

$$\frac{S_n}{n} \xrightarrow{P} m \tag{8}$$

Demostración:

Sea F la función de distribución de cada X_n. Así

$$m = E[X_n] = \int_{\mathbb{R}} x\,dF(x), \quad E[|X_n|] = \int_{\mathbb{R}} |x|\,dF(x) < +\infty$$

Y, por el (teorema 13), como $E[|X_1|]$ es finita

$$\sum_n P(|X_1| > n) < +\infty,$$

luego como son idénticamente distribuidas

$$\sum_n P(|X_n| > n) < +\infty \tag{9}$$

Ahora, definamos una sucesión $(Y_n)_n$ de v.a por «truncamiento» de la sucesión $(X_n)_n$

$$Y_n(\omega) = \begin{cases} X_n(\omega) & \text{si } |X_n(\omega)| \leq n \\ 0 & \text{si } |X_n(\omega)| > n, \end{cases}$$

la cual es equivalente a $(X_n)_n$ por (9) pues

$$\sum_n P(|X_n| > n) = \sum_n P(X_n \neq Y_n) < +\infty.$$

Luego, por el (corolario 4.6); entonces se tiene (3) si y solo si, definiendo

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

probamos que $T_n/n \xrightarrow{P} m$. Notemos que $(Y_n)_n$ es independiente dos a dos aplicando el (teorema 19) a cada par. Por lo tanto, no están relacionadas pues cada una tiene segundo momento finito pues es acotada.

Ahora, vamos a calcular $\sigma^2(T_n)$, para ello, como

$$\sigma^2(S_n) = \sum_{j=1}^{n} \sigma^2(X_j) \tag{10}$$

Entonces

$$\sigma^2(T_n) = \sum_{j=1}^{n} \sigma^2(Y_j) \leq \sum_{j=1}^{n} E[Y_j^2] = \sum_{j=1}^{n} \int_{|x|\leq j} x^2\, dF(x)$$

En una primera estimación, se tiene que

$$\sum_{j=1}^{n} \int_{|x|\leq j} x^2 dF(x) \leq \sum_{j=1}^{n} j \int_{|x|\leq j} |x|\, dF(x) \leq \frac{n(n+1)}{2} \int_{\mathbb{R}} |x| dF(x) = O(n^2).$$

Para mejorar la estimación, sea $(a_n)_n$ una sucesión de de enteros tal que $0 < a_n < n$, $a_n \to +\infty$ pero $a_n = o(n)$, es decir $(a_n/n \to 0)$. Así, notemos que

$$\sum_{j=1}^{n} \int_{|x|\leq j} x^2 dF(x) = \left(\sum_{j\leq a_n} + \sum_{a_n<j\leq n}\right) \int_{|x|\leq j} x^2 dF(x)$$

$$\leq \sum_{j\leq a_n} a_n \int_{|x|\leq a_n} |x|\, dF(x) + \sum_{a_n<j\leq n} j \int_{|x|\leq j} |x| dF(x)$$

$$= \sum_{j\leq a_n} a_n \int_{|x|\leq a_n} |x| dF(x) + \sum_{a_n<j\leq n} a_n \int_{|x|\leq a_n} |x| dF(x) + \sum_{a_n<j\leq n} n \int_{a_n<|x|\leq n} |x|\, dF(x)$$

$$= \sum_{j\leq a_n} a_n \int_{|x|\leq a_n} |x|\, dF(x) + \sum_{a_n<j\leq n} a_n \int_{|x|\leq a_n} |x|\, dF(x) + \sum_{a_n<j\leq n} (j-a_n) \int_{|x|\leq a_n} |x|\, dF(x)$$

$$+ \sum_{a_n<j\leq n} j \int_{a_n<|x|\leq j} |x|\, dF(x)$$

$$= \sum_{j=1}^{n} a_n \int_{|x|\leq a_n} |x|\, dF(x) + \sum_{a_n<j\leq n} (2j-a_n) \int_{|x|\leq a_n} |x|\, dF(x)$$

$$\leq n a_n \int_{|x|\leq a_n} |x|\, dF(x) + 2n^2 \int_{|x|>a_n} |x|\, dF(x).$$

Notemos que, para el primer término $O(na_n) = o(n^2)$ y para el segundo $n^2 o(1) = o(n^2)$. Luego, como el conjunto $\{x : |x| > a_n\}$ decrece al vacío, entonces la última integral converge a cero. Así, hemos probado que $\sigma^2(T_n) = o(n^2)$. Así, por (2), se sigue que

$$\frac{T_n - E[T_n]}{n} = \frac{1}{n}\sum_{j=1}^{n} \{Y_j - E[Y_j]\} \xrightarrow{P} 0.$$

Finalmente, por definición, si $n \to +\infty$, entonces $E[Y_n] \to E[X] = m$. Así, se sigue que

$$\frac{1}{n}\sum_{j=1}^{n} E[Y_j] \to m.$$

De donde, se concluye que

$$\frac{T_n}{n} = \frac{1}{n}\sum_{j=1}^{n} Y_j \xrightarrow{P} m$$

Sección 4.2. Convergencia de series

Teorema 4.8. Sea $(X_n)_n$ una sucesión de v.a. tales que

$$\forall n: E[X_n] = 0 \quad y \quad E[X^2] = \sigma^2(X_n) < +\infty.$$

Entonces, para $\varepsilon > 0$, cualquiera

$$P\left(\max_{1\leq j\leq n} |S_j| > \varepsilon\right) \leq \frac{\sigma^2(S_n)}{\varepsilon^2} \tag{11}$$

Note que si reemplazamos $\max_{1\leq j\leq n} |S_j|$ por $|S_n|$, se obtiene la desigualdad de Chebychev usual.

Demostración. Sea $\varepsilon > 0$, cualquiera. Para ω cualquiera en

$$A = \{\omega : \max_{1\leq j\leq n} |S_j(\omega)| > \varepsilon\}$$

definamos

$$v(\omega) = \min\{j : 1 \leq j \leq n, |S_j(\omega)| > \varepsilon\}.$$

Así, claramente v es una variable aleatoria con dominio A. Por otro lado,

$$A_k = \{\omega : v(\omega) = k\} = \{\omega : \max_{1\leq j\leq k-1} |S_j(\omega)| \leq \varepsilon, |S_k(\omega)| > \varepsilon\}$$

de donde, para $k=1$, definimos $\max_{1\leq j\leq 0}|S_j(\omega)|$ como cero. De esta manera, v indica la primera vez que el máximo indicado, excede a ε y A_k es el evento donde esto ocurre «la primera vez en el k-ésimo paso». Los A_k son disjuntos y además

$$A = \bigcup_{k=1}^{n} A_k$$

Así, se sigue que

$$\int_A S_n^2\,dP = \sum_{k=1}^{n} \int_{A_k} S_n^2\,dP = \sum_{k=1}^{n} \int_{A_k} [S_k + (S_n - S_k)]^2\,dP \qquad (12)$$

$$= \sum_{k=1}^{n} \int_{A_k} [S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2]\,dP.$$

Notemos φ_k a la indicatriz de A_k, entonces las siguientes variables aleatorias $\varphi_k S_k$ y $S_n - S_k$ son independientes por el (teorema 20). Luego, se sigue que

$$\int_{A_k} S_k(S_n - S_k)\,dP = \int_{\Omega} (\varphi_k S_k)(S_n - S_k)\,dP$$

$$= \int_{\Omega} \varphi_k S_k\,dP \int_{\Omega} (S_n - S_k)\,dP = 0.$$

Y por definición, sabemos que

$$E[S_n - S_k] = \sum_{j=k+1}^{n} E[X_j] = 0.$$

Usando la última expresión en (12), concluimos que

$$\sigma^2(S_n) = \int_{\Omega} S_n^2\,dP \geq \int_A S_n^2\,dP \geq \sum_{k=1}^{n} \int_{A_k} S_k^2\,dP$$

$$\geq \varepsilon^2 \sum_{k=1}^{n} P(A_k) = \varepsilon^2 P(A).$$

Teorema 4.9. Sea $(X_n)_n$ una s.v.a. con media finita y existe $A > 0$ tal que

$$\forall n : |X_n - E[X_n]| \leq A < +\infty. \qquad (13)$$

Entonces, para todo $\varepsilon > 0$, tenemos que

$$P\left(\max_{1\leq j\leq n} |S_j| \leq \varepsilon\right) \leq \frac{(2A + 4\varepsilon)^2}{\sigma^2(S_n)} \qquad (14).$$

Demostración:

Tomemos $H_0 = \Omega$; así, para $1 \leq k \leq n$

$$H_k = \{\omega : \max_{1\leq j\leq k} |S_j| \leq \varepsilon\}$$

$$\Delta_k = H_{k-1} - H_k$$

Además, supondremos que $P(H_n) > 0$, caso contrario (14) es trivial. Más aún, sea $S_0' = 0$ y para $k \geq 1$

$$X_k' = X_k - E[X_k], \qquad S_k' = \sum_{j=1}^{k} X_j'.$$

Por otro lado, definamos a_k, con $0 \leq k \leq n$ de la siguiente manera

$$a_k = \frac{1}{P(H_k)} \int_{H_k} S_k'\,dP.$$

de modo que

$$\int_{H_k} (S_k' - a_k)\, dP = 0 \tag{15}$$

Así, escribamos

$$\int_{H_{k+1}} (S_{k+1}' - a_{k+1})^2\, dP = \underbrace{\int_{H_k} (S_k' - a_k + a_k - a_{k+1} + X_{k+1}')^2\, dP}_{I_1} - \underbrace{\int_{\Delta_{k+1}} (S_k' - a_k + a_k - a_{k+1} + X_{k+1}')^2\, dP}_{I_2} \tag{16}$$

Usando la definición de H_k y por (13)

$$|S_k' - a_k| = \left| S_k - E[S_k] - 1/P(H_k) \int_{H_k} [S_k - E[S_k]\, dP \right|$$
$$= \left| S_k - \frac{1}{P(H_k)} \int_{H_k} S_k\, dP \right| \leq |S_k| + \varepsilon$$

$$|a_k - a_{k+1}| = \left| \frac{1}{P(H_k)} \int_{H_k} S_k\, dP - \frac{1}{P(H_{k+1})} \int_{H_{k+1}} S_k\, dP - \frac{1}{P(H_{k+1})} \int_{H_{k+1}} X_{k+1}'\, dP \right| \leq 2\varepsilon + A$$

y así, como $|S_k| \leq \varepsilon$ en Δ_{k+1}

$$I_2 \leq \int_{\Delta_{k+1}} (|S_k| + \varepsilon + 2\varepsilon + A + A)^2\, dP \leq (4\varepsilon + 2A)^2\, P(\Delta_{k+1}).$$

Por otro lado

$$I_1 = \int_{H_k} \{ (S_k' - a_k)^2 + (a_k - a_{k+1})^2 + X_{k+1}'^2 + 2(S_k' - a_k)(a_k - a_{k+1}) + 2(S_k' - a_k) X_{k+1}' + 2(a_k - a_{k+1}) X_{k+1}' \}\, dP$$

Así, por independencia y (15)

$$I_1 \geq \int_{H_k} (S_k' - a_k)^2\, dP + \int_{H_k} X_{k+1}'^2\, dP = \int_{H_k} (S_k' - a_k)^2\, dP + P(H_k)\, \sigma^2(X_{k+1}).$$

Sustituyendo en (16) y como $H_k \supseteq H_n$, para $0 \leq k \leq n-1$

$$\int_{H_{k+1}} (S_{k+1}' - a_{k+1})^2\, dP - \int_{H_k} (S_k' - a_k)^2\, dP \geq P(H_n) \sum_{j=1}^{n}{}' \sigma^2(X_j) - (4\varepsilon + 2A)^2\, P(\Omega \setminus H_n).$$

Por lo tanto,

$$(2A + 4\varepsilon)^2 \geq P(H_n) \sum_{j=1}^{n}{}' \sigma^2(X_j)$$

lo que es (14) y completa la prueba.

Teorema 4.10. Sea $(X_n)_n$ una s.v.a. independientes, entonces para $A > 0$ fijo, definimos

$$Y_n(\omega) = \begin{cases} X_n(\omega) & \text{si } |X_n(\omega)| \leq A \\ 0 & \text{si } |X_n(\omega)| > A \end{cases},$$

entonces la serie $\sum_{n \in \mathbb{N}} X_n$ converge c.s. si y sólo si las tres series convergen

a) $\sum_n P(|X_n| > A) = \sum_n P(X_n \neq Y_n)$

b) $\sum_n E[Y_n]$

c) $\sum_n \sigma^2(Y_n)$.

Demostración

$\Leftarrow$) Supongamos que las tres series convergen. Así, aplicando el (teorema 4.8) a la sucesión $(Y_n - E[Y_n])$, se tiene que para todo $m \geq 1$.

$$\underbrace{P\left(\max_{n \leq k \leq n'} \left| \sum_{j=n}^{k}{}' \{Y_j - E[Y_j]\} \right| \leq 1/m \right)}_{P(m,n,n')} \geq 1 - m^2 \sum_{j=n}^{n'}{}' \sigma^2(Y_j)$$

Entonces, como c) converge, para cada m

$$\lim_{n \to +\infty} \lim_{n' \to +\infty} P(m,n,n') = 1.$$

Lo que implica que la cola de $\sum_n |Y_n - E[Y_n]|$ converge a cero c.s.; por lo tanto, la serie converge c.s. Luego, como b) converge, entonces $\sum_n Y_n$ también converge. Además, como a) converge $(X_n)_n$ y $(Y_n)_n$ son equivalentes por definición.

Por lo tanto, $\sum_n X_n$ también converge c.s. por el (teorema 4.5).

$\Rightarrow$ Supongamos que $\sum_n X_n$ converge c.s. Así, por Borel - Cantelli, para todo $A>0$

$$P(|X_n|>A \text{ i.o.})=0.$$

y nuevamente, por el mismo resultado, a) converge. Utilizando un razonamiento análogo al anterior, se concluye que $\sum_n Y_n$ converge c.s. Pero como $|Y_n - E[Y_n]| \leq 2A$, entonces por el (teorema 4.9)

$$P\left\{ \max_{n\leq k\leq n'} \left|\sum_{j=n}^{k} Y_j\right| \leq 1 \right\} \leq \frac{(4A+4)^2}{\sum_{j=n}^{n'} \sigma^2(Y_j)}$$

Donde, si c) no converge (i.e. diverge) la probabilidad anterior tiende a cero cuando $n' \to +\infty$ para cada n. Por lo tanto, la cola de $\sum_n Y_n$ casi seguramente no está acotada por uno y la serie no converge. No obstante esto no es posible pues ya sabemos que esta converge. Así, por contradicción, hemos probado c).

Finalmente, consideremos la serie $\sum_n \{Y_n - E[Y_n]\}$ y aplicando la parte a) (ya demostrada) $P(|Y_n - E[Y_n]|>2A)=0$ y así $E[Y_n - E[Y_n]]=0$. Por lo tanto, las series correspondientes de a) y b) con $2A$ en lugar de A se anulan idénticamente. Así, del criterio suficiente, se sigue que $\sum_n [Y_n - E[Y_n]]$ converge c.s. Más aún $\sum_n (X_n - E[Y_n])$ también converge c.s. Finalmente, como $\sum_n X_n$ converge por hipótesis, entonces, por substracción la serie b) converge.

Teorema 4.11. Sea $(X_n)_n$ una s.v.a. independientes, entonces la convergencia de la serie $\sum_n X_n$ en probabilidad, equivale a la convergencia c.s.

teo 5.3.4 [3]

Demostración:

Por el (teorema 3.4), basta probar que la convergencia en P de $\sum_n X_n$ implica c.s.

Así, supongamos que $\sum_n X_n$ converge en P, entonces para ε con $0<\varepsilon<1$, existe $m_0 \in \mathbb{N}$ tal que si $n>m>n_0$

$$P(|S_{m,n}|>\varepsilon)<\varepsilon \qquad (17)$$

donde

$$S_{m,n} = \sum_{j=m+1}^{n} X_j.$$

Ahora, es claro que para $m<k\leq n$

$$\bigcup_{k=m+1}^{n} \left\{ \max_{m<j\leq k-1} |S_{m,j}| \leq 2\varepsilon \,;\, |S_{m,k}|>2\varepsilon,\, |S_{k,n}|\leq\varepsilon \right\} \subseteq \{|S_{m,n}|>\varepsilon\} \qquad (18)$$

donde los conjunto de la unión son disjuntos. Así, usando probabilidades y por independencia

$$\sum_{k=m+1}^{n} P\left(\max_{m\leq j\leq k-1} |S_{m,j}| \leq 2\varepsilon \,;\, |S_{m,k}|>2\varepsilon\right) P(|S_{k,n}|\leq\varepsilon) \leq P(|S_{m,n}|>\varepsilon)$$

Eliminando los términos $P(|S_{k,n}|\leq\varepsilon)$, la sumatoria es igual a

$$P\left(\max_{m\leq j\leq n} |S_{m,j}|>2\varepsilon\right).$$

Así, se sigue que

$$P\left(\max_{m\leq j\leq n} |S_{m,j}|>2\varepsilon\right) \min_{m\leq k\leq n} P(|S_{k,n}|\leq\varepsilon) \leq P(|S_{m,n}|>\varepsilon)$$

Luego, por (17) el segundo factor de la izquierda excede $1-\varepsilon$; así, si $m>m_0$

$$P\left(\max_{m\leq j\leq n} |S_{m,j}|>2\varepsilon\right) \leq \frac{1}{1-\varepsilon} P(|S_{m,n}|>\varepsilon) < \frac{\varepsilon}{1-\varepsilon}$$

Así, cuando $n\to+\infty$, $m\to+\infty$ y $\varepsilon\to 0$, entonces el triple límite en la primera probabilidad es cero. lo que prueba la convergencia. c.s.

Lema 4.12 [Lema de Kronecker's] Sea $(x_k)_k$ una sucesión de números reales, $(a_k)_k$ una sucesión de números positivos ($a_k > 0$ $\forall k$) tal que $a_k \uparrow +\infty$, entonces

pág 129 [3]

$$\text{si } \sum_n \frac{x_n}{a_n} \text{ converge, entonces } \frac{1}{a_n}\sum_{j=1}^{n} x_j \to 0.$$

Demostración:

Para $1 \le n \le +\infty$, definamos

$$b_n = \sum_{j=1}^{n} x_j/a_j$$

y, si además, $a_0 = 0$, $b_0 = 0$, entonces

$$x_n = a_n(b_n - b_{n-1})$$

y así

$$1/a_n \sum_{j=1}^{n} x_j = 1/a_n \sum_{j=1}^{n} a_j(b_j - b_{j-1}) = b_n - 1/a_n \sum_{j=0}^{n-1} b_j(a_{j+1} - a_j)$$

(Método de sumas parciales de Abel). Luego, como $a_{j+1} - a_j \geq 0$, entonces

$$\frac{1}{a_n}\sum_{j=0}^{n-1}(a_{j+1} - a_j) = 1$$

y así, $b_n \to b_\infty$, de donde

$$\frac{1}{a_n}\sum_{j=1}^{n} x_j \to b_\infty - b_\infty = 0$$

lo que completa la demostración del lema

Teorema 4.13. Sea $(X_n)_n$ una s.v.a. independiente con $E[X_n] = 0$, para todo n, $(a_n)_n$ con $a_n > 0$, para todo n y $a_n \uparrow +\infty$. Si φ es positiva, par. y continua en $\mathbb{R}$ tal que

teo 5.4.1 [3]

$$\frac{\varphi(x)}{|x|} \uparrow, \quad \frac{\varphi(x)}{x^2} \downarrow \tag{19}$$

y

$$\sum_n \frac{E[\varphi(X_n)]}{\varphi(a_n)} < +\infty. \tag{20}$$

Entonces

$$\sum_n \frac{X_n}{a_n} \text{ converge casi-seguramente.}$$

Demostración.

Para cada $n \in \mathbb{N}$, denotemos por F_n a la f.d. de cada X_n, respectivamente. Luego, para cada n, definamos

$$Y_n(\omega) = \begin{cases} X_n(\omega) & \text{si } |X_n(\omega)| \geq a_n \\ 0 & \text{si } |X_n(\omega)| < a_n \end{cases} \tag{21}$$

Entonces,

$$\sum_n E[Y_n^2/a_n^2] = \sum_n \int_{|x| \le a_n} x^2/a_n^2 \, dF_n(x).$$

Luego, por la segunda hipótesis de (19)

$$\frac{x^2}{a_n^2} \le \frac{\varphi(x)}{\varphi(a_n)} \qquad \text{para todo } |x| \le a_n.$$

Así, se sigue que

$$\sum_n \sigma^2(Y_n/a_n) \le \sum_n E(Y_n^2/a_n^2) \le \sum_n \int_{|x| \le a_n} \varphi(x)/\varphi(a_n)\, dF_n(x)$$
$$\le \sum_n E[\varphi(X_n)]/\varphi(a_n) < +\infty \qquad (\text{por } (20))$$

De esta manera, para la s.v.a. $((Y_n - E[Y_n])/a_n)$, la serie asociada c) del (teorema 4.10) converge; mientras que las series a) y b) se anulan para $\Delta = 2$, En efecto, como

$$|Y_n - E[Y_n]| \le 2a_n$$

entonces

$$\sum_n \frac{1}{a_n}\{Y_n - E[Y_n]\} \tag{22}$$

converge c.s.

Luego, se sigue que

$$\sum_n{}' \frac{|E[Y_n]|}{a_n} = \sum_n{}' \frac{1}{a_n}\left| \int_{|x|\leq a_n} x\, dF_n(x) \right| = \sum_n{}' \frac{1}{a_n}\left| \int_{|x|>a_n} x\, dF_n(x) \right|$$

$$\leq \sum_n{}' \int_{|x|>a_n} \frac{|x|}{a_n}\, dF_n(x).$$

donde la segunda ecuación se sigue de que $E[X_n] = \int_{-\infty}^{+\infty} x\, dF_n(x) = 0$. Así, por la primera hipótesis en (19), se sigue que

$$\frac{|x|}{a_n} \leq \frac{\varphi(x)}{\varphi(a_n)} \quad \text{para} \quad |x| \geq a_n.$$

y en consecuencia,

$$\sum_n{}' \frac{|E[Y_n]|}{a_n} \leq \sum_n{}' \int_{|x|>a_n} \frac{\varphi(x)}{\varphi(a_n)}\, dF_n(x) \leq \sum_n{}' \frac{E[\varphi(X_n)]}{\varphi(a_n)} < +\infty$$

De donde, usando la expresión anterior y (22), entonces $\sum_n (Y_n/a_n)$ converge casi seguramente. Finalmente, como φ es creciente, entonces

$$\sum_n{}' P(X_n \neq Y_n) = \sum_n{}' \int_{|x|>a_n} dF_n(x) \leq \sum_n{}' \int_{|x|>a_n} \frac{\varphi(x)}{\varphi(a_n)}\, dF_n(x) \leq \sum_n{}' \frac{E[\varphi(X_n)]}{\varphi(a_n)} < +\infty$$

Así, las sucesiones $(X_n)_n$ y $(Y_n)_n$ son equivalentes y usando el (teorema 4.10) se concluye el resultado deseado.

Aplicando el lema de Kronecker a $\sum_n X_n/a_n$ para cada ω en un conjunto de probabilidad uno se obtiene el siguiente resultado.

Corolario. 4.14. Bajo las mismas hipótesis del (teorema 4.13) tenemos que

$$\frac{1}{a_n} \sum_{j=1}^{n} X_j \to 0 \tag{23}$$

casi seguramente.

Casos particulares.

a) Sea $\varphi(x) = |x|^p$, $1 \leq p \leq 2$; $a_n = n$, entonces

$$\sum_n \frac{1}{n^p} E[|X_n|^p] < +\infty \Rightarrow \frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{c.s.} 0 \tag{24}$$

En particular, para $p=2$ el resultado se debe a Kolmogorov, para $1 \leq p < 2$ el resultado se debe a Marcinkiewicz y Zygmund.

b) Sea δ con $0 < \delta \leq 1$ y $M < +\infty$, entonces

$$\forall n: E[|X_n|^{1+\delta}] \leq M,$$

entonces aplicamos (24) con $p = 1+\delta$

c) Para una elección apropiada de $(a_n)_n$, podemos mejorar la conclusión (23). Supongamos que

$$\forall n: \sigma^2(X_n) = \sigma_n^2 < +\infty \qquad \sigma^2(S_n) = s_n^2 = \sum_{j=1}^{n} \sigma_j^2 \longrightarrow +\infty$$

Tomando $\varphi(x) = x^2$ y $a_n = s_n (\log s_n)^{(1/2)+\varepsilon}$, $\varepsilon > 0$ en el (corolario 4.14), entonces

$$\sum_n{}' \frac{E[X_n^2]}{a_n^2} = \sum{}' \frac{\sigma_n^2}{s_n^2 (\log s_n)^{1+2\varepsilon}} < +\infty$$

por el teorema de Dini y en consecuencia

$$\frac{S_n}{s_n (\log(s_n))^{1/2+\varepsilon}} \xrightarrow{c.s.} 0$$

Teorema 4.15. Sea $(X_n)_n$ una s.v.a. i.i (independientes, idénticamente distribuidas), entonces.
teo 5.4.2 [3]
$$E[|X_1|] < +\infty \Rightarrow S_n/n \longrightarrow E[X_1] \quad \text{casi seguramente} \tag{25}$$
$$E|X_1| = +\infty \Rightarrow \overline{\lim_{n\to+\infty}} \frac{|S_n|}{n} = +\infty \quad \text{casi seguramente} \tag{26}$$

Demostración

• Para demostrar (25), definamos $(Y_n)_n$ como en (21) y tomemos $a_n = n$. Así, se tiene que
$$\sum_n P(X_n \neq Y_n) = \sum_n P(|X_n| > n) = \sum_n P(|X_1| > n) < +\infty$$
por el (teorema 13). Así, se sigue que $(X_n)_n$ y $(Y_n)_n$ son sucesiones equivalentes. Por otro lado, vamos a aplicar (24) a $Y_n - E[Y_n]$ con $\varphi(x) = x^2$. Así
$$\sum_n \sigma^2(Y_n)/n^2 \leq \sum_n E[Y_n^2]/n^2 = \sum_n 1/n^2 \int_{|x|\leq n} x^2\, dF(x) \tag{27}$$
Para continuar, debemos estimar el segundo momento en términos del primero, dado que es la única hipótesis conocida. La técnica estándar es separar el intervalo de integración y luego invertir la sumatoria repetida.
$$\sum_{n=1}^{+\infty} 1/n^2 \sum_{j=1}^{n} \int_{j-1<|x|\leq j} x^2\, dF(x) = \sum_{j=1}^{+\infty} \int_{j-1<|x|\leq j} x^2\, dF(x) \sum_{n=j}^{\infty} 1/n^2$$
$$\leq \sum_{j=1}^{+\infty} j \int_{j-1<|x|\leq j} |x|\, dF(x) \cdot c/j$$
$$\leq c \sum_{j=1}^{+\infty} \int_{j-1<|x|\leq j} |x|\, dF(x) = c\,E[|X_1|] < +\infty$$
En el desarrollo anterior se usó que $\sum_{n=j}^{+\infty} n^{-2} \leq cj^{-1}$ para alguna constante c y para todo $j \geq 1$.
Así, se sigue que (27) converge y usando (21)
$$\frac{1}{n}\sum_{j=1}^{n} \{Y_j - E[Y_j]\} \xrightarrow{c.s.} 0.$$
Claramente, $E[Y_n] \to E[X_1]$ cuando $n \to +\infty$. Por lo tanto,
$$\frac{1}{n}\sum_{j=1}^{n} E[Y_j] \longrightarrow E[X_1]$$
y así
$$\frac{1}{n}\sum_{j=1}^{n} Y_j \to E[X_1] \quad \text{casi seguramente.}$$
Por el (teorema 4.5) el lado izquierdo puede ser reemplazado por
$$\frac{1}{n}\sum_{j=1}^{n} X_j = \frac{1}{n} S_n \longrightarrow E[X_1]$$
lo que prueba (25).

• Para demostrar (26), notemos que si $E|X_1| = +\infty$, entonces $E[|X_1|/A] = +\infty$ para cada $A > 0$ y así, por el (teorema 13)
$$\sum_n P(|X_n| > An) = +\infty.$$
Finalmente, por el teorema de Borel-Cantelli (teorema 3.18)
$$P\{|X_n| > An \ \text{i.o.}\} = 1$$
No obstante, $|S_n - S_{n-1}| = |X_n| > An$ implica que $|S_n| > An/2$ o $|S_{n-1}| > An/2$. Así
$$P\left\{|S_n| > \frac{An}{2} \ \text{i.o.}\right\} = 1$$
lo que significa que, existe un conjunto de medida nula $Z(A)$ tal que si $\omega \in \Omega \setminus Z(A)$, entonces
$$\limsup_{n\to+\infty} \frac{|S_n(\omega)|}{n} \geq \frac{A}{2} \tag{28}$$
Definamos $Z = \bigcup_{m=1}^{+\infty} Z(m)$, así Z es de medida nula y tomando $\omega \in \Omega \setminus Z$, entonces (28) se tiene para cada A y así
$$\limsup_{n\to+\infty} \frac{|S_n|}{n} = +\infty$$

Teorema 4.16 Sean $(X_n)_n$ una s.v.a i.i con $E|X_1| = +\infty$, y $(a_n)_n$ una sucesión de números positivos tales que $a_n/n \uparrow$, entonces
$$\limsup_{n\to+\infty} \frac{|S_n|}{a_n} = 0 \ \text{c.s.} \quad \text{o} \quad \limsup_{n\to+\infty} \frac{|S_n|}{a_n} = +\infty \tag{29}$$
Consecuentemente

$$\sum_n P(|X_n| \geq a_n) = \sum_n \int_{|x|\geq a_n} dF(x) < +\infty \quad o \quad \sum_n \int_{|x|\geq a_n} dF(x) = +\infty$$

Corolario 4.17 Bajo las mismas condiciones del teorema 4.16

$$P(|S_n| \geq a_n \text{ i.o.}) = P(|X_n| \geq a_n \text{ i.o.})$$

Aplicaciones

Una introducción al tema se presenta en las páginas 138-140 [3].

Teorema 4.18. Cuando $n \to +\infty$

$$\sup_{x\in\mathbb{R}} |F_n(x,\omega) - F(x)| \longrightarrow 0 \quad c.s.$$

Demostración.

Sea J el conjunto de saltos de F. Así, para cada $x \in J$, se define

$$\eta_j(x,\omega) = \begin{cases} 1 & \text{si } X_j(\omega) = x \\ 0 & \text{si } X_j(\omega) \neq x. \end{cases}$$

Así, se sigue que, para $x \in J$

$$F_n(x+,\omega) - F_n(x-,\omega) = \frac{1}{n}\sum_{j=1}^{n} \eta_j(x,\omega).$$

Así, existe un conjunto $N(x)$ de medida nula tal que para $x \in \Omega \setminus N(x)$

$$F_n(x+,\omega) - F_n(x-,\omega) \to F(x+) - F(x-). \tag{30}$$

Tomando $N = \bigcup_{x\in\mathbb{Q}\cup J} N(x)$, entonces, para todo $x \in J$, se tiene (30) y así

$$F_n(x,\omega) \longrightarrow F(x).$$

Finalmente, el resultado se sigue de la aplicación del siguiente lema.

Lema 4.19. Sean F_n y F (continuas a la derecha) funciones de distribución y $\mathbb{Q}$ y J definidas en la introducción. Supongamos que

$$\forall x \in \mathbb{Q}: \ F_n(x) \to F(x)$$
$$\forall x \in J: \ F_n(x) - F_n(x-) \to F(x) - F(x-),$$

entonces F_n converge uniformemente en $\mathbb{R}$.

Demostración. Pág 140-141 [3]

Teorema 4.20 Sea f una función continua en $[0,1]$, se define los polinomios de Bernstein $\{p_n\}$

teo 5.5.4 [3].

$$p_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}.$$

entonces p_n converge uniformemente a f en $[0,1]$.

Capítulo 5: Función característica y teorema del límite central

Una herramienta importante para el estudio de las v.a. y las m.p. son las funciones características (f.c).

Definición 5.1. [Función característica]. Sean X una v.a. con P m.p asociada y f.d. F, se define la función característica como

$$\varphi(t) = f(t) = E[e^{itX}] = \int_\Omega e^{itX(\omega)} P(d\omega) = \int_{\mathbb{R}} e^{itx} P_X(dx) = \int_{\mathbb{R}} e^{itx} dF(x) \tag{1}$$

(ver teorema 15 o 3.2.3 [3]) (ver definición 4)

Note que como la función es compleja, entonces

$$Re(\varphi(t)) = \int_{\mathbb{R}} \cos(xt) F(dx) \quad y \quad Im(\varphi(t)) = \int_{\mathbb{R}} sen(xt) F(dx)$$

En análisis, la función característica es conocida como la transformada de Fourier-Stieltjes de P_X o F.

Propiedades de la función característica.

1) $\forall t \in \mathbb{R}$ $|f(t)| \leq 1$, $\varphi(0) = f(0) = 1$, $f(-t) = \overline{f(t)}$
donde $\bar{z}$ denota el conjugado de z.
2) φ es uniformemente continua en $\mathbb{R}$
3) Si denotamos φ_X a la f.c. de X, entonces para cualesquiera $a, b \in \mathbb{R}$
$$f_{aX+b}(t) = f_X(at)e^{itb}$$
4) Sea $(f_n)_n$ una sucesión de f.c., $\lambda_n \geq 0$ con $\sum_{n=1}^{+\infty} \lambda_n = 1$, entonces
$$\sum_{n=1}^{+\infty} \lambda_n f_n$$
es una f.c. (i.e. la combinación convexa de funciones características, es una función característica. Más aún, si $(\mu_n)_n$ son las correspondientes m.p., entonces $\sum_{n=1}^{+\infty} \lambda_n \mu_n$ es m.p. con función característica asociada $\sum_{n=1}^{+\infty} \lambda_n f_n$.
5) Si $(\varphi_n)_{n=1}^{m}$ son f.c., entonces
$$\prod_{n=1}^{m} \varphi_n$$
es una función característica.

Sección 5.1 Convolución

Definición 5.2. [Convolución] Sean F_1 y F_2 dos f.d., se define el producto en convolución como
$$F(x) = \int_{\mathbb{R}} F_1(x-y)\, dF_2(y) \tag{2}$$
y lo notamos $F = F_1 * F_2$

Teorema 5.3. Sean X_1, X_2 dos v.a. con f.d asociadas F_1 y F_2, respectivamente, entonces $X_1 + X_2$ tiene f.d. $F_1 * F_2$.

Demostración
Recordemos que queremos probar que $\forall: P(X_1 + X_2 \leq x) = (F_1 * F_2)(x)$.
Así, definamos, para $x \in \mathbb{R}$, cualquiera
$$f(x_1, x_2) = \begin{cases} 1 & \text{si } x_1 + x_2 \leq x \\ 0 & \text{caso contrario} \end{cases}$$
notemos que f es Borel medible. Por el (teorema 21) y usando notaciones de (teo 3.3.3 [3])
$$\int_{\Omega} f(X_1, X_2)\, dP = \iint_{\mathbb{R}^2} f(x_1, x_2)\, \mu^2(dx_1, dx_2)$$
$$= \int_{\mathbb{R}} \mu_2(dx_2) \int_{\mathbb{R}} f(x_1, x_2)\, \mu_1(dx_1) \qquad \text{(Por independencia)}$$
$$= \int_{\mathbb{R}} \mu_2(dx_2) \int_{(-\infty, x-x_2]} \mu_1(dx_1)$$
$$= \int_{\mathbb{R}} dF_2(x_2)\, F_1(x - x_2)$$
de donde, por definición, se sigue el resultado deseado (2)

Corolario 5.4. La operación convolución (*) es conmutativa y asociativa.

Definición 5.5 Sean p_1, p_2 dos funciones de densidad se define el producto en convolución es la función de densidad p, definida por
$$\forall x \in \mathbb{R}: \quad p(x) = \int_{-\infty}^{+\infty} p_1(x-y) p_2(y)\, dy \tag{3}$$
notada por $p = p_1 * p_2$

Teorema 5.6. El producto en convolución de dos funciones de distribución absolutamente continuas con densidades p_1 y p_2 es absolutamente continuo con densidad $p = p_1 * p_2$

Demostración

Por el teorema de Fubini

$$\int_{-\infty}^{x} p(u)\,du = \int_{-\infty}^{x} du \int_{-\infty}^{+\infty} p_1(u-v)\,p_2(v)\,dv$$

$$= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{x} p_1(u-v)\,du \right] p_2(v)\,dv$$

$$= \int_{-\infty}^{+\infty} F_1(u-v)\,p_2(v)\,dv$$

$$= \int_{-\infty}^{+\infty} F_1(u-v)\,dF_2(v) = (F_1 * F_2)(x)$$

Así, p es la función de densidad de $F_1 * F_2$

Teorema 5.7 Para cada $B \in Bor(\mathbb{R})$, tenemos que

$$(\mu_1 * \mu_2)(B) = \int_{\mathbb{R}} \mu_1(B-y)\,\mu_2(dy). \qquad (4)$$

y, para g una función medible integrable respecto a $\mu_1 * \mu_2$

$$\int_{\mathbb{R}} g(u)\,(\mu_1 * \mu_2)(du) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x+y)\,\mu_1(dx)\,\mu_2(dy) \qquad (5)$$

Demostración

No es difícil ver que (4) es una m.p. Para demostrar que su f.d es $F_1 * F_2$, basta probar para $B =]-\infty, x]$ se tiene que es igual a (2). Así, como

$$\int_{\mathbb{R}} F_1(x-y)\,\mu_2(dy) = \int_{-\infty}^{+\infty} F_1(x-y)\,dF_2(y) = F(x)$$

Por otro lado, sea g la función indicatriz de B, entonces para cada y se define

$$g_y(x) = g(y+x)$$

como la indicatriz del conjunto $B-y$. Por lo tanto

$$\int_{\mathbb{R}} g(x+y)\,\mu_1(dx) = \mu_1(B-y)$$

Substituyendo la expresión anterior en (5) y por el literal anterior

$$\int_{\mathbb{R}} \int_{\mathbb{R}} g(x+y)\,\mu_1(dx)\,\mu_2(dy) = \int_{\mathbb{R}} \mu_1(B-y)\,\mu_2(dy) = (\mu_1 * \mu_2)(B) = \int_{\mathbb{R}} g(u)\,(\mu_1 * \mu_2)(du)$$

El caso general se demuestra considerando g una función simple, luego tomando el límite para funciones integrables.

Ejemplo.

Calculemos la f.c. del producto en convolución $\mu_1 * \mu_2$, por (5)

$$\int e^{itu}\,(\mu_1 * \mu_2)(du) = \iint e^{ity} e^{itx}\,\mu_1(dx)\,\mu_2(dy) = \int e^{itx}\mu_1(dx) \int e^{ity}\mu_2(dy)$$

Teorema 5.8. A la suma finita de variables aleatorias independientes le corresponde como función de distribución la convolución de sus funciones de distribución y como función característica, la multiplicación de sus funciones características.

Demostración: Por inducción

Corolario 5.9 Si φ es una función característica, entonces $|\varphi|^2$ también lo es.

Demostración

Sea X una v.a y φ su respectiva f.c., entonces existe en algún Ω una v.a. Y independiente de X con su misma f.d y f.c. Notemos que la f.c. de $X-Y$ es

$$E[e^{it(X-Y)}] = E[e^{itX}]E[e^{itY}] = f(t)f(-t) = |f(t)|^2$$

La técnica de considerar $X-Y$ y $|e|^2$ en lugar de X y f se conoce como simetrización

Algunas funciones características se pueden encontrar en las páginas 155 - 156 [3]

Teorema 5.10. Para cada $\delta>0$, $f_\delta \in C_B^{+\infty}$. Más aún, si $f \in C_U$, entonces $f_\delta \to f$ uniformemente en $\mathbb{R}$.

Donde:

- C_B^∞: es el conjunto de las funciones con derivadas de todo orden continuas y acotadas.
- C_U: es el conjunto de las funciones acotadas uniformemente continuas.

y

$$f_\delta(x) = (f * n_\delta)(x) = \int_{-\infty}^{+\infty} f(x-y)\,n_\delta(y)\,dy = \int_{-\infty}^{+\infty} n_\delta(x-y)\,f(y)\,dy \qquad (6)$$

$$n_\delta(x) = \frac{1}{\sqrt{2\pi\delta}} \exp\left(-\frac{x^2}{2\delta^2}\right) \quad ; -\infty < x < +\infty$$

Demostración.

No es difícil ver que $n_\delta \in C_B^\infty$. Más aún, su k-ésima derivada está acotada por $C_{k,\delta}\, n_{2\delta}$ con $C_{k,\delta}$ es una constante que solo depende de k y δ; así

$$\left|\int_{-\infty}^{+\infty} n_\delta^{(k)}(x-y)f(y)\,dy\right| \leq C_{k,\delta}\,\|f\| \int_{-\infty}^{+\infty} n_{2\delta}(x-y)\,dy = C_{k\delta}\,\|f\|$$

Así, la primera afirmación se sigue por la diferenciación bajo el signo integral del último término en (6)

La seguna afirmación se prueba de la siguiente manera, para $\eta>0$

$$|f(x) - f_\delta(x)| \leq \int_{-\infty}^{+\infty} |f(x) - f(x-y)|\,n_\delta(y)\,dy$$

$$\leq \sup_{|y|\leq\eta} |f(x) - f(x-y)| + 2\|f\| \int_{|y|>\eta} n_\delta(y)\,dy$$

Teorema 5.11. Sean $(\mu_n)_n$ y μ m.s.p. tales que

$$\forall f \in C_B^\infty \quad \int_{\mathbb{R}} f(x)\,\mu_n(x) \to \int f(x)\,\mu(x),$$

entonces $\mu_n \xrightarrow{v} \mu$.

La demostración de este teorema es consecuencia de los (teoremas 3.25 y 5.10) si notamos que $C_0 \subseteq C_U$

Sección 5.2. Unicidad e inversión.

Para estudiar con mayor profundidad las propiedades de la transformada de Fourier-Stieltjes, requerimos revisar las integrales de Dirichlet

$$\forall y \geq 0 : \int_0^{\pi} \frac{\operatorname{sen}(x)}{x}\,dx \geq \left(\int_0^{J} \frac{\operatorname{sen}(\alpha x)}{x}\,dx\right)\operatorname{sgn}(\alpha) \geq 0 \qquad (7)$$

$$\int_0^{\infty} \frac{\operatorname{sen}(\alpha x)}{x}\,dx = \frac{\pi}{2}\operatorname{sgn}(\alpha) \qquad (8)$$

$$\int_0^{+\infty} \frac{1-\cos(\alpha x)}{x^2} = \frac{\pi}{2}|\alpha| \qquad (9)$$

Un esquema de la demostración de las fórmulas anteriores, se encuentra en pág. 161 [3]

Teorema 5.12. Si $x_1 < x_2$, entonces

$$\mu((x_1,x_2)) + \frac{1}{2}\mu(\{x_1\}) + \frac{1}{2}\mu(\{x_2\}) \qquad (10)$$
$$= \lim_{T\to+\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-itx_1}-e^{-itx_2}}{it} f(t)\,dt,$$

el integrando se define por continuidad en $t=0$.

Demostración:
Note que la integral está acotada por $|x_1-x_2|$ en todo punto y es $O(|t|^{-1})$ cuando $|t|\to+\infty$. Así, para demostrar el teorema, vamos a sustituir la definición de f y realizar los cálculos respectivos. Así, tenemos que

$$\frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-itx_1}-e^{-itx_2}}{it}\left[\int_{-\infty}^{+\infty} e^{itx}\mu(dx)\right]dt \qquad (11)$$
$$= \int_{-\infty}^{+\infty} \underbrace{\left[\int_{-T}^{T} \frac{e^{-it(x-x_1)}-e^{it(x-x_2)}}{2\pi i t}\,dt\right]}_{I(T,x,x_1,x_2)} \mu(dx)$$

De esta manera, con la notación anterior

$$I(T,x,x_1,x_2) = \frac{1}{\pi}\int_0^T \frac{\operatorname{sen} t(x-x_1)}{t}\,dt - \frac{1}{\pi}\int_0^T \frac{\operatorname{sen} t(x-x_2)}{t}\,dt.$$

y así, de (8), se sigue que

$$\lim_{T\to+\infty} I(T,x,x_1,x_2) = \begin{cases} -1/2-(-1/2)=0 & \text{para } x<x_1 \\ 0-(-1/2)=1/2 & \text{para } x=x_1 \\ 1/2-\left(\frac{1}{2}\right)=1 & \text{para } x_1<x<x_2 \\ 1/2-0 = 1/2 & \text{para } x=x_2 \\ 1/2-1/2=0 & \text{para } x>x_2 \end{cases}$$

Más aún, se tiene que I está acotada por (7). Por lo tanto, podemos tomar $T\to+\infty$ bajo el signo integral en (5) por convergencia monótona. Así, como

$$|I(T,x,x_1,x_2)| \le \frac{2}{\pi}\int_0^{\pi} \frac{\operatorname{sen} x}{x}\,dx$$

por (7), entonces

$$\left\{\int_{(-\infty,x_1)} 0 + \int_{\{x_1\}} \frac{1}{2} + \int_{(x_1,x_2)} 1 + \int_{\{x_2\}} \frac{1}{2} + \int_{(x_2,+\infty)} 0\right\} \mu(dx)$$
$$= \frac{1}{2}\mu(\{x_1\}) + \mu\{(x_1,x_2)\} + \frac{1}{2}\mu(\{x_2\}).$$

lo que prueba el resultado.

Ahora, el intercambio de integrales en (5) está justificado por el teorema de Fubini y notando que

$$\left|\frac{e^{it(x-x_1)}-e^{it(x-x_2)}}{it}\right| = \left|\int_{x_1}^{x_2} e^{-itu}\,du\right| \le |x_1-x_2|$$

donde la integral se toma en el eje real y más aún

$$\int_{\mathbb{R}}\int_{-T}^{T}|x_1-x_2|\,dt\,\mu(dx) \leq 2T|x_1-x_2| < +\infty$$

De esta manera, la integral en (9) está dominada por una función integrable finita respecto al producto finito de medidas $dt\cdot\mu(dx)$ en $[-T,T]\times\mathbb{R}$.

Teorema. 5.13. Si dos m.p. o f.d. tienen la misma f.c., entonces son las mismas m.p. o f.d.

Demostración.

Si x_1 y x_2 no son átomos de μ, entonces la fórmula (10) nos muestra que el valor de μ en el intervalo (x_1,x_2) está determinado por su función característica. Así, se sigue que dos medidas de probabilidad tienen la misma f.ch en cada intervalo en el los extremos del mismo no sean átomos de alguna de las dos medidas. Luego, como cada m.p. tiene solo un conjunto contable de átomos, los puntos de $\mathbb{R}$ que no son átomos para cualquiera de las dos medidas forman un conjunto denso. Por lo tanto, las medidas coinciden en un conjunto denso de intervalos y por lo tanto por el (teo 2.2.3 [3]) son idénticas.

Teorema 5.14. Si $f\in L^1(-\infty,+\infty)$, entonces F es continuamente diferenciable y además

$$F'(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{itx} f(t)\,dt \qquad (12)$$

Demostración

Aplicando (10) con $x_2=x$, $x_1=x-h$ con $h>0$ y usando F en lugar de μ, se tiene que

$$\frac{F(x)+F(x-)}{2} - \frac{F(x-h)+F((x-h)-)}{2} = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{e^{ith}-1}{it}e^{-itx}f(t)\,dt$$

Así, por hipótesis, la integral existe y como esta está dominada por $|hf(t)|$. Entonces, cuando $h\to 0$ y por convergencia dominada concluimos que el lado izquierdo es igual a cero. De esta manera, hemos probado que F es continua por izquierda y más aún continua en $\mathbb{R}$. De esta manera, podemos escribir

$$\frac{F(x)-F(x-h)}{h} = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{e^{ith}-1}{ith}e^{-itx}f(t)\,dt.$$

Utilizando un argumento similar se muestra que el límite cuando $h\to 0$ existe. Por lo tanto, F tiene derivada por izquierda en x y es igual al lado derecho de (12).
Similarmente, F tiene derivada por derecha.
Para más detalles revisar ([6] pág 354).

Corolario 5.15. Si $f\in L^1$, entonces $p\in L^1$, donde

$$f(t) = \int_{-\infty}^{+\infty} e^{itx}p(x)\,dx \quad \text{y} \quad p(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}f(t)\,dt \qquad (13)$$

Teorema 5.16. Para cada x_0, tenemos que

$$\lim_{T\to+\infty}\frac{1}{2T}\int_{-T}^{T} e^{-itx_0}f(t)\,dt = \mu(\{x_0\}).$$

Demostración

Siguiendo una idea similar a la realizada en el teorema 5.12, se obtiene para la integral promedio en el lado izquierdo de (7)

$$\int_{\mathbb{R}\setminus\{x_0\}}\frac{\operatorname{sen}T(x-x_0)}{T(x-x_0)}\,\mu(dx) + \int_{\{x_0\}} 1\,\mu(dx) \qquad (14)$$

El primer integrando está acotado por 1 y tiende a cero cuando $T \to +\infty$ en todo punto del dominio de integración. Así, por convergencia acotada la integral converge a cero. El segundo término es simplemente $\mu(\{x_0\})$.

Teorema. 5.17. Tenemos que

$$\lim_{T \to +\infty} \frac{1}{2T} \int_{-T}^{T} |f(t)|^2 \, dt = \sum_{x \in \mathbb{R}}' \mu(\{x\})^2 \qquad (15)$$

Demostración

Puesto que el conjunto de átomos es contable, entonces una cantidad contable de números en la sumatoria se anulan y esta queda acotada por 1.

Por el (corolario 5.9) $|f|^2$ es la función característica de $X - Y$, donde su distribución es $\mu * \mu'$ donde $\mu'(B) = \mu(-B)$ para cada $B \in Bor(\mathbb{R})$ Aplicando el (teorema 5.19) para $x_0 = 0$, se tiene que el lado derecho de (15) es igual a

$$(\mu * \mu')(\{0\})$$

y por (4)

$$\int_{\mathbb{R}} \mu'(\{-y\}) (\mu \, dy) = \sum_{y \in \mathbb{R}}' \mu(\{y\}) \, \mu(\{y\})$$

Finalmente, como la última integral es cero a menos que $-y$ sea un átomo de μ' lo que se tiene si y solo si y es un átomo de μ. Esto nos permite obtener el lado derecho de (15)

Corolario 5.18. μ no tiene átomos (i.e. F es continua) si y solo si el límite del lado izquierdo en (15) es cero.

Definición 5.19. Sea X una v.a., se dice que X es simétrica si X y $-X$ tienen la misma f.d.

Teorema 5.20 Una v.a. X o μ m.p. es simétrica si y solo si su función característica es real valuada para todo t.

Sección 5.3. Teoremas de convergencia.

Teorema 5.21 (teo 6.3.1 [3]) Sea $\{\mu_n : 1 \leq n \leq +\infty\}$ m.p. con funciones características asociadas $\{f_n : 1 \leq n \leq +\infty\}$. Si $\mu_n \xrightarrow{v} \mu_\infty$, entonces f converge uniformemente a f_∞ en cada intervalo finito. Simbólicamente

$$\mu_n \xrightarrow{v} \mu_\infty \implies f_n \xrightarrow{\mu} f_\infty \qquad (16)$$

Más aún, la familia $\{f_n\}$ es continua en $\mathbb{R}$.

Demostración

Notemos que e^{itx} es una función acotada y continua para cada $x \in \mathbb{R}$, en el campo complejo. Así, aplicando el (teorema 3.29) a las partes reales e imaginarias; así se tiene (16) sin la uniformidad.

Ahora, para todo t y h se tiene que

$$|f_n(t+h) - f_n(t)| \leq \int |e^{ihx} - 1| \, \mu_n(dx) \leq \int_{|x| \leq A} |hx| \, \mu_n(dx) + 2 \int_{|x| > A} \mu_n(dx)$$

$$\leq |h| A + 2 \int_{|x| > A} \mu(dx) + \varepsilon$$

para todo $\varepsilon > 0$ y con A y $n \geq n_0(A, \varepsilon)$. Así, se sigue la equicontinuidad de $\{f_n\}$. Por lo tanto, junto a la convergencia puntual $f_n \to f_\infty$ implica que $f_n \xrightarrow{\mu} f_\infty$.

Teorema 5.22. Sea $(\mu_n)_n$ una sucesión de m.p. en $\mathbb{R}$ con funciones características $(f_n)_n$. Supongamos que

a) f_n converge en todo punto en $\mathbb{R}$ y definamos f_∞

b) f_∞ es continua en $t=0$

Entonces, se tiene que

i) $\mu_n \to \mu_\infty$ con μ_∞ una m.p.

ii) f_∞ es la función característica de μ_∞

Lema 5.23 Para cada $A>0$, tenemos que

$$\mu([-2A,2A]) \geq A \left| \int_{-A^{-1}}^{A^{-1}} f(t)dt \right| - 1 \tag{17}$$

Demostración

Por (14), tenemos que

$$\frac{1}{2T}\int_{-T}^{T} f(t)dt = \int_{-\infty}^{+\infty} \frac{\operatorname{sen} Tx}{Tx}\,\mu(dx). \tag{18}$$

Así, como la integral de la derecha está acotada por 1 para todo x (se define como 1 para $x=0$) y como $|Tx|^{-1} \leq (2TA)^{-1}$ para $|x|>2A$, la integral está acotada por

$$\mu([-2A,2A]) + \frac{1}{2TA}[1-\mu([-2A,2A])\}$$

$$= \left(1-\frac{1}{2TA}\right)\mu([-2A,2A]) + \frac{1}{2TA}$$

Tomando $T=A^{-1}$ en (18) se sigue que

$$\left|\frac{A}{2}\int_{-A^{-1}}^{A^{-1}} f(t)\right| \leq \frac{1}{2}\mu([-2A,2A]) + \frac{1}{2}$$

de donde se sigue (17) como se quería.

Demostración (teorema 5.22)

Para cada δ con $0<\delta<\delta_0$, tenemos que

$$\underbrace{\left|\frac{1}{2\delta}\int_{-\delta}^{\delta} f_n(t)\,dt\right|}_{(C)} \geq \underbrace{\left|\frac{1}{2\delta}\int_{-\delta}^{\delta} f(t)\,dt\right|}_{(A)} - \underbrace{\frac{1}{2\delta}\int_{-\delta}^{\delta}|f_n(t)-f(t)|\,dt}_{(B)} \tag{19}$$

Luego $A\to 1$ cuando $\delta\downarrow 0$ pues $f(0)=1$ y f es continua en 0. Para δ fijo $B\to 0$ por convergencia acotada pues $|f_n - f|\leq 2$. Así, se sigue que para $\varepsilon>0$ dado, existe $\delta=\delta(\varepsilon)$ con $\delta<\delta_0$ y $n_0=n_0(\varepsilon)$ tal que para todo $n\geq n_0$, entonces (C) no es menor que $1-\varepsilon$.

Así, por (17)

$$\mu_n([-2\delta^{-1},2\delta^{-1}]) \geq 2(1-\varepsilon)-1 \geq 1-2\varepsilon. \tag{20}$$

Sea $(\mu_{n_k})_k$ una subsucesión convergente vagamente de (μ_n) (esta existe por el (teorema (3.24)) que converge a μ una m.s.p.

Para cada δ que satisface las condiciones anteriores y tal que $-2\delta^{-1}$ y $2\delta^{-1}$ no sean átomos de μ, tenemos, por la propiedad de convergencia vaga que

$$\mu(\mathbb{R}) \geq \mu([-2\delta^{-1},2\delta^{-1}]) = \lim_{n\to+\infty} \mu_n([-2\delta^{-1},2\delta^{-1}]) \geq 1-2\varepsilon$$

Como ε es arbitrario, entonces es una m.p.

Sea f la función característica de μ. Así, por el (teorema 5.21) $f_{n_k} \to f$ en todas partes y por la hipótesis (a), $f=f_\infty$. Así, todo límite de convergencia vaga considerado tiene la misma función característica y por teorema de unicidad, la misma m.p. notando esta última como μ_∞.

Por lo tanto, hemos probado (i) y (ii)

Sección 5.4. Aplicaciones simples

Teorema 5.24. Sea F una función de distribución si f tiene momentos absolutos finitos de orden k entonces su función característica tiene derivada de orden k continua y acotada, entonces.

$$\varphi^{(k)}(t) = \int_{-\infty}^{+\infty} (ix)^k e^{itx}\, dF(x) \tag{21}$$

Recíprocamente, si f tiene derivada finita par de orden k en $t=0$, entonces F tiene momento finito de orden k

Demostración
Para $k=1$, la primera afirmación se obtiene de la fórmula

$$\frac{f(t+h)-f(t)}{h} = \int_{-\infty}^{+\infty} \frac{e^{i(t+h)x} - e^{itx}}{h}\, dF(x)$$

Usando una desigualdad elemental usada en la demostración del (teorema 5.12), se demuestra que la integral anterior está dominada por $|x|$. Así, si

$$\int |x|\, dF(x) < +\infty$$

podemos tomar $h \to 0$ bajo el signo integral para obtener (21). La continuidad uniforme forma parte de las propiedades demostradas al inicio del capítulo.
El caso general k se obtiene por inducción

Para la segunda afirmación, sea $k=2$ y suponga que $f''(0)$ existe y es finito. Así, tenemos que

$$\begin{aligned} f''(0) &= \lim_{h\to 0} \frac{f(h)-2f(0)+f(-h)}{h^2} \\ &= \lim_{h\to 0} \int_{-\infty}^{+\infty} \frac{e^{ihx} - 2 + e^{-ihx}}{h^2}\, dF(x) \\ &= -2 \lim_{h\to 0} \int_{-\infty}^{+\infty} \frac{1-\cos(hx)}{h^2}\, dF(x) \end{aligned} \tag{22}$$

Luego, como

$$2 \lim_{h\to 0} \frac{1-\cos(hx)}{h^2} = x^2,$$

se tiene que

$$\begin{aligned} \int_{-\infty}^{+\infty} x^2 F(dx) &= 2 \int_{-\infty}^{+\infty} \lim_{h\to 0} \frac{1-\cos(hx)}{h^2} F(dx) \\ &\leq \liminf_{h\to 0} 2 \int_{-\infty}^{+\infty} \frac{1-\cos(hx)}{h^2} F(dx) = -f''(0) \end{aligned}$$

Por lo tanto, F tiene segundo momento finito.
El caso general puede ser reducido por inducción, como se muestra a continuación.

Supongamos que f tiene derivada finita par de orden $2k-2$ en $t=0$ y que $f^{(2k)}(0)$ es finita, entonces $f^{(2k-2)}(t)$ existe y es continua en alguna vecindad de $t=0$ y luego por la hipótesis inductiva, tenemos que en particular

$$f^{(2k-2)}(0) = (-1)^{k-1} \int x^{2k-2}\, dF(x) = f^{(2k-2)}(0)$$

Tomando $G(x) = \int_{-\infty}^{x} y^{2k-2}\, dF(y)$ para cada x, entonces $G(\cdot)/G(\infty)$ es una función de distribución con función característica

$$\psi(t) = \frac{1}{G(\infty)} \int e^{itx} x^{2k-2} dF(x) = \frac{(-1)^{k-1} f^{(2k-2)}(t)}{G(\infty)}$$

Por lo tanto, se sigue que ψ'' existe y por el caso $k=2$ probado anteriormente, tenemos que

$$-\psi^{(2)}(0) = \frac{1}{G(\infty)} \int x^2 \, dG(x) = \frac{1}{G(\infty)} \int x^{2k} dF(x)$$

De donde, se concluye que

$$(-1)^k f^{(2k)}(0) = \int x^{2k} dF(x),$$

lo que demuestra que el $2k$-ésimo momento es finito. El argumento anterior falla si $G(\infty)=0$ pero en ese caso, se tiene que $F = \delta_0$, $f=1$ y el teorema es trivial.

Teorema 5.25. Si F tiene momento absoluto de orden k, con k un entero de orden $k \geq 1$, es decir

teo 6.4.2 [7]

$$\int |x|^k F(dx) < +\infty \qquad k \geq 1. \tag{23}$$

Entonces, en una vecindad de $t=0$:

$$f(t) = \sum_{j=0}^{k} \frac{i^j}{j!} m^{(j)} t^j + o(|t|^k) \tag{24}$$

o

$$f(t) = \sum_{j=0}^{k-1} \frac{i^j}{j!} m^{(j)} t^j + \frac{\theta_k}{k!} \mu^{(k)} |t|^k; \tag{25}$$

donde $m^{(j)}$ es el momento de orden j, $\mu^{(k)}$ es el momento absoluto de orden k y $|\theta_k| \leq 1$.

Demostración. (ver [7], pág 290)

De acuerdo a la expansión en series de Taylor, si f es k veces diferenciable en $t=0$ entonces

$$f(t) = \sum_{j=0}^{k} \frac{f^{(j)}(0)}{j!} t^j + o(|t|^k) \tag{26}$$

Si f tiene la k-ésima derivada finita en una vecindad de $t=0$, entonces

$$f(t) = \sum_{j=0}^{k-1} \frac{f^{(j)}(0)}{j!} t^j + \frac{f^{(k)}(\theta t)}{k!} t^k, \qquad |\theta| \leq 1 \tag{27}$$

Luego, como el momento absoluto de orden j con $1 \leq j \leq k$ es finito y

$$f^{(j)}(0) = i^j m^{(j)}, \qquad |f^{(k)}(\theta t)| \leq \mu^{(k)}.$$

Así, por (21), (26) y (27), se deduce el resultado deseado.

Lema 5.26 Sea $(c_n)_n$ una sucesión de números complejos tal que $c_n \to c$, entonces

$$\lim_{n \to +\infty} \left(1 + \frac{c_n}{n}\right)^n = e^c$$

Sea $(X_n)_n$ una sucesión de v.a. i.i.d con distribución F y $S_n = \sum_{k=1}^{n} X_k$.

Teorema 5.27. Si F tiene media (esperanza) finita, entonces

$$\frac{S_n}{n} \xrightarrow{P} m$$

Demostración.

Puesto que la convergencia a una constante m es equivalente a la convergencia en distribución a δ_m, es suficiente, por el (teorema 5.22) probar que la f.c. de S_n/n converge a e^{imt} (la cual es continua). Así, como

$$f_{S_n} = \prod_{j=1}^{n} f_{X_j},$$

tenemos que
$$E[e^{it(S_n/n)}] = E[e^{i(t/n)S_n}] = \left[f\left(\tfrac{t}{n}\right)\right]^n.$$
Por el (teorema 5.??), el último término se puede escribir como
$$\left(1 + im\frac{t}{n} + o\left(\frac{t}{n}\right)\right)^n.$$

Luego, para t fijo, y cuando $n \to +\infty$, se sigue por el (lema 5.26) que
$$f_{S_n/n}(t) \to e^{imt},$$

Teorema 5.28. Si F tiene esperanza m y varianza finita $\sigma^2 > 0$, entonces
$$\frac{S_n - mn}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

Demostración

Sin pérdida de generalidad, podemos suponer $m=0$ si consideramos $X_j - m$ con segundo momento σ^2. Análogamente a la demostración anterior, tenemos que
$$E\left[\exp\left(it\frac{S_n}{\sigma\sqrt{n}}\right)\right] = f\left(\frac{t}{\sigma\sqrt{n}}\right)^n$$
$$= \left\{1 + \frac{i^2\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + o\left(\frac{|t|}{\sigma\sqrt{n}}\right)^2\right\}^n$$
$$= \left\{1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right\}^n \to e^{-t^2/2}.$$
Luego, por el (teorema 5.13), se sigue el resultado deseado

Teorema 5.29. Sean X y Y v.a. i.i.d. con media 0 y varianza 1. Si $X+Y$ y $X-Y$ son independientes, entonces la distribución común de X y Y es $N(0,1)$.

Demostración

Sea f la función característica, entonces por (21),
$$f'(0) = 0 \quad y \quad f''(0) = -1.$$
La función característica de $X+Y$ es $f(t)^2$ y para $X-Y$ es $f(t)f(-t)$. Y, luego, como estas variables son independientes, la función característica $f(2t)$ de $2X$ satisface la siguiente relación
$$f(2t) = f(t)^3 f(-t). \tag{28}$$
Así, se sigue por (28) que la función f no se anula. En efecto, si se anulara para algún t_0, entonces también lo haría en $t_0/2$ y por inducción en $t_0/2^n$ para todo $n \in \mathbb{N}$. Pero esto no es posible pues
$$\lim_{n\to+\infty} f(t_0/2^n) = f(0) = 1.$$
Definiendo para cada t
$$\rho(t) = \frac{f(t)}{f(-t)},$$
tenemos que
$$\rho(2t) = \rho(t)^2.$$
Luego, iterativamente se tiene que
$$\rho(t) = \rho\left(\frac{t}{2^n}\right)^{2^n} = \left\{1 + o\left(\frac{t}{2^n}\right)\right\}^{2^n} \to 1$$
por el (teorema 5.25) y (lema 5.26). Así $\rho(t) \equiv 1$, $f(t) \equiv f(-t)$ y (28), entonces
$$f(2t) = f(t)^4. \tag{29}$$
Repitiendo el argumento anterior en (29), tenemos que
$$f(t) = f\left(\frac{t}{2^n}\right)^{4^n} = \left\{1 - \frac{1}{2}\left(\frac{t}{2^n}\right)^2 + o\left(\frac{t}{2^n}\right)^2\right\}^{4^n} \to e^{-t^2/2}$$

Capítulo 6: Teorema del límite central.

Sección 6.1 Liapounov

El teorema del límite central se refiere a un resultado que asegura la convergencia en distribución de una suma «normada» de variables aleatorias $(S_n - a_n)/b_n$ a una normal $N(0,1)$.

Una versión especial del teorema fue probada en el (teorema 5.28).

Para generalizar el teorema, notemos que

$$\frac{S_n - a_n}{b_n} = \left(\sum_{j=1}^{n} \frac{X_j}{b_n}\right) - \frac{a_n}{b_n}. \tag{1}$$

así, veamos a continuación una manera para manejar un arreglo bidimensional.

Para cada $n \geq 1$, tenemos el siguiente arreglo de v.a. $\{X_{n,j} : 1 \leq j \leq k_n\}$, donde $k_n \to +\infty$ cuando $n \to +\infty$:

$$\begin{matrix} X_{11} & X_{12} & \ldots, & X_{1k_1}; \\ X_{21} & X_{22} & \ldots, & X_{2k}; \\ \vdots & \vdots & & \\ X_{n1} & X_{n2} & \ldots, & X_{nk_n}; \\ \vdots & \vdots & & \end{matrix} \tag{2}$$

Las v.a. con n como primer subíndice se refieren a la n-ésima fila.

Sea F_{nj} la función de distribución, f_{nj} la f.c. y notemos

$$S_n = S_{n,k_n} = \sum_{j=1}^{k_n} X_{nj}.$$

El caso particular en el que $k_n = n$, para cada n, nos da como resultado un arreglo triangular y más aún $X_{nj} = X_j$, para cada n, entonces se reduce a la sucesión $(X_j)_{j \geq 1}$.

Asumiremos que las v.a. en cada fila de (2) son independientes; más aún pueden estar definidas en distintos espacios de probabilidad sin ninguna relación entre sí. Además, introduciremos las siguientes notaciones

$$\begin{aligned} E[X_{nj}] &= \alpha_{nj} & \sigma^2(X_{nj}) &= \sigma_{nj}^2 \\ E[S_n] &= \sum_{j=1}^{k_n} \alpha_{nj} = \alpha_n & \sigma^2(S_n) &= \sum_{j=1}^{k} \sigma_{nj}^2 = s_n^2 \\ E|X_{nj}|^3 &= \gamma_{nj} & \Gamma_n &= \sum_{j=1}^{k_n} \gamma_{nj} \end{aligned} \tag{3}$$

En el caso especial de (1), tenemos que

$$X_{nj} = \frac{X_j}{b_n} \qquad \sigma^2(X_{nj}) = \frac{\sigma^2(X_j)}{b_n^2}.$$

Si tomamos $b_n = s_n$, entonces

$$\sum_{j=1}^{k_n} \sigma^2(X_{nj}) = 1 \tag{4}$$

Considerando $X_{nj} - \alpha_{nj}$, entonces

$$\forall n, \forall j : \alpha_{nj} = 0 \tag{5}$$

siempre que la media exista. La reducción « en ocasiones llamada normalización» de (4) y (5) es posible siempre que X_{nj} tenga segundo momento finito.

Para manejar arreglos como el presentado en (2) es importante suponer que los términos en la suma

$$S_n = \sum_{j=1}^{k_n} X_{nj}$$

son «insignificantes» en comparación con la suma en sí misma.

Para clarificar la noción intuitiva de insignificancia, consideremos la siguiente jerarquía de condiciones que se satisfacen para todo $\varepsilon>0$

(a) $\forall j: \lim_{n\to+\infty} P(|X_{nj}|>\varepsilon)=0$;

(b) $\lim_{n\to+\infty} \max_{1\le j\le k_n} P(|X_{nj}|>\varepsilon)=0$;

(c) $\lim_{n\to+\infty} P(\max_{1\le j\le k_n} |X_{nj}|>\varepsilon)=0$;

(d) $\lim_{n\to+\infty} \sum_{j=1}^{k_n} P(|X_{nj}|>\varepsilon)=0$.

Es claro que (d) $\Rightarrow$ (c) $\Rightarrow$ (b) $\Rightarrow$ (a).

Teorema 6.1 La condición (b) es equivalente a

$$\forall t\in\mathbb{R}: \lim_{n\to+\infty} \max_{1\le j\le k_n} |f_{nj}(t)-1|=0. \tag{6}$$

Demostración:

($\Rightarrow$) Supongamos que (b) es cierta, entonces

$$|f_{nj}(t)-1| \le \int |e^{itx}-1|\,dF_{nj}(x) = \int_{|x|>\varepsilon} + \int_{|x|\le\varepsilon}$$

$$\le \int_{|x|>\varepsilon} 2\,dF_{nj}(x) + |t|\int_{|x|\le\varepsilon} |x|\,dF_{nj}(x)$$

$$\le 2\int_{|x|>\varepsilon} dF_{nj}(x) + \varepsilon|t|;$$

y, consecuentemente

$$\max_j |f_{nj}(t)-1| \le 2\max_j P(|X_{nj}|>\varepsilon)+\varepsilon|t|,$$

tomando $n\to+\infty$ y $\varepsilon\to0$, se sigue el resultado.

Recíprocamente, del (lema 5.23)

$$\int_{|x|>\varepsilon} dF_{nj}(x) \le 2 - \left|\frac{\varepsilon}{2}\int_{|t|\le 2/\varepsilon} f_{nj}(t)\,dt\right| \le \frac{\varepsilon}{2}\int_{|t|\le 2/\varepsilon} |1-f_{nj}(t)|\,dt$$

y en consecuencia

$$\max_j P(|X_{nj}|>\varepsilon) \le \frac{\varepsilon}{2}\int_{|t|\le 2/\varepsilon} \max_j |1-f_{nj}(t)|\,dt$$

tomando $n\to+\infty$, por hipótesis, y convergencia acotada, se sigue el resultado.

Lema 6.2. Sea $\{\theta_{n,j},\ 1\le j\le k_n,\ n\ge1\}$ un arreglo doble de números complejos tales que

i) $\max_{1\le j\le k_n} |\theta_{nj}|\to0$;

ii) $\sum_{j=1}^{k_n} |\theta_{nj}|\le M<+\infty$ con M independiente de n;

iii) $\sum_{j=1}^{k_n} \theta_{nj}\to\theta$, con θ un número complejo finito.

Entonces, tenemos que

$$\prod_{j=1}^{k_n}(1+\theta_{nj})\to e^{\theta} \tag{7}$$

Demostración

Por (i), existe n_0 tal que si $n\ge n_0$, entonces

$$|\theta_{nj}|\le 1/2 \qquad \forall j$$

y así $1+\theta_{nj}\ne0$. Ahora, considerando valores de n suficientemente grandes y denotando por $\log(1+\theta_{nj})$ con un ángulo entre $]-\pi,\pi]$. Entonces

$$\log(1+\theta_{nj})=\theta_{nj}+\Lambda|\theta_{nj}|^2, \tag{8}$$

donde Λ es un número complejo dependiente de varias variables pero acotado por alguna constante absoluta independiente de todo.

Así, se sigue que
$$|\log(1+\theta_{nj})-\theta_{nj}| = \left|\sum_{m=2}^{\infty}\frac{(-1)^{m-1}}{m}\theta_{nj}^m\right| \leq \sum_{m=2}^{\infty}\frac{|\theta_{nj}|^m}{m}$$
$$\leq \frac{|\theta_{nj}|^2}{2}\sum_{m=2}^{+\infty}\left(\frac{1}{2}\right)^{m-2} = |\theta_{nj}|^2 \leq 1$$

Por lo tanto, la constante mencionada puede ser tomada como 1. Luego
$$\sum_{j=1}^{k_\theta}\log(1+\theta_{nj}) = \sum_{j=1}^{k_n}\theta_{nj} + \Lambda\sum_{j=1}^{k_n}|\theta_{nj}|^2$$

Note que este Λ no es el mismo que el anterior pero igual está acotado por 1. Así, por (ii) y (i)
$$\sum_{j=1}^{k_n}|\theta_{nj}|^2 \leq \max_{1\leq j\leq k_n}|\theta_{nj}|\sum_{j=1}^{k_n}|\theta_{nj}| \leq M\max_{1\leq j\leq k_n}|\theta_{nj}| \longrightarrow 0 \tag{9}$$

y en consecuencia, por (iii)
$$\sum_{j=1}^{k}\log(1+\theta_{nj}) \longrightarrow \theta$$

De donde se deduce (7), como se quería.

Teorema 6.3. Sean $\{X_{nj}: 1\leq j\leq k_n, n\geq 1\}$ variables aleatorias independientes por filas con $\sum_{j=1}^{k_n}\sigma^2(X_{nj}) = 1$ y $\forall n: \alpha_{nj} = 0$.

Si
$$\Gamma_n = \sum_{j=1}^{k_n}\gamma_{nj} \xrightarrow[n\to+\infty]{} 0 \qquad \text{con } \gamma_{nj} = E|X_{nj}|^3 \tag{10}$$

entonces S_n converge en distribución a $N(0,1)$.

Demostración

Para cada $n\in\mathbb{N}$, j está comprendido entre 1 y k_n. Así, se sigue que por (10) y por la desigualdad de Liapounov (ver capítulo 2 - desigualdades), se tiene que
$$\max_j \sigma_{nj}^3 \leq \max_j \gamma_{nj} \leq \Gamma_n \longrightarrow 0. \tag{11}$$

Luego, por (25 - capítulo 5)
$$f_{nj}(t) = 1 - \frac{1}{2}\sigma_{nj}^2t^2 + \Lambda_{nj}\gamma_{nj}|t|^3,$$

donde $|\Lambda_{nj}|\leq 1/6$. Así, aplicando el (lema 6.2) para un t fijo
$$\theta_{nj} = -\frac{1}{2}\sigma_{nj}^2t^2 + \Lambda_{nj}\gamma_{nj}|t|^3$$

Para ello, la condición (i) se satisface pues
$$\max_j|\theta_{nj}| \leq \frac{t^2}{2}\max_j\sigma_{nj}^2 + \Lambda|t|^3\max_j\gamma_{nj} \longrightarrow 0$$

por (11). La condición (ii) se sigue pues
$$\sum_j|\theta_{nj}| \leq \frac{t^2}{2} + \Lambda|t|^3\Gamma_n$$

es acotada por (11). Similarmente se satisface (iii) pues
$$\sum_j\theta_{nj} = -\frac{t^2}{2} + \Lambda|t|^3\Gamma_n \longrightarrow \frac{t^2}{2}$$

Así, se sigue que,
$$\prod_{j=1}^{k_n}f_{nj}(t) \longrightarrow e^{-t^2/2}$$

Corolario 6.4. Sin suponer que $E[X_{nj}]=0$, supongamos que para cada n y j existe una constante H_{nj} tal que $|X_{nj}|\leq H_{nj}$ casi seguramente y,

$$\max_{1\leq j\leq k_n} H_{nj}\longrightarrow 0 \tag{12}$$

Entonces $S_n - E(S_n)$ converge en distribución a $N(0,1)$.

Demostración

En primer lugar, notemos que

$$\sum_{j=1}^{k_n} E|X_{nj}-E[X_{nj}]|^3 \leq 2\max_{1\leq j\leq k_n} H_{nj}\sum_{j=1}^{k_n}\sigma^2(X_{nj}) = 2\max_{1\leq j\leq k_n} H_{nj}$$

Luego, consideremos la formulación unidimensional para una sucesión $(X_n)_n$ de v.a.i. tales que:

$$E[X_j]=0,\quad \sigma^2(X_j)=\sigma_j^2<+\infty,\quad E|X_j|^3<+\infty,$$

Así, tomando

$$S_n=\sum_{j=1}^n X_j,\qquad s_n^2=\sum_{j=1}^n \sigma_j^2 \qquad \text{y}\qquad \Gamma_n=\sum_{j=1}^n \gamma_j.$$

Si $\Gamma_n/s_n^3\longrightarrow 0$, entonces $S_n/s_n \xrightarrow{d} N(0,1)$

Teorema 6.5 - Condición de Lindeberg- Sea $(X_n)_n$ una s.v.a. con segundo momento finito, sea $m_k=E[X_k]$, $Var(X_k)=\sigma_k^2>0$, $S_n=X_1+\cdots+X_n$, $D^2=\sum_{k=1}^n\sigma_k^2$ y $F_k=F_k(x)$ la función de distribución de X_k.

Suponga además que se tiene la condición de Lindeberg; para todo $\varepsilon>0$:

$$(L)\quad \frac{1}{D_n^2}\sum_{k=1}^n\int_{\{x:|x-m_k|\geq\varepsilon D_n\}}(x-m_k)^2\,dF_k(x)\longrightarrow 0,\quad n\longrightarrow+\infty \tag{13}$$

Entonces

$$\frac{S_n-E[S_n]}{\sqrt{Var\ S_n}}\xrightarrow{d} N(0,1)$$

Demostración:

Sin pérdida de generalidad, podemos asumir que $m_k=0$, $k\geq 1$. Denotemos

$$\varphi_k(t)=E[e^{itX_k}],\quad T_n=\frac{S_n}{\sqrt{Var\ S_n}}=\frac{S_n}{D_n},\quad \varphi_{S_n}(t)=E[e^{itS_n}],\quad \varphi_{T_n}(t)=E[e^{itT_n}]. \tag{14}$$

entonces

$$\varphi_{T_n}(t)=E[e^{itT_n}]=E[e^{itS_n/D_n}]=\varphi_{S_n}\left(\frac{t}{D_n}\right)=\prod_{k=1}^n\varphi_k\left(\frac{t}{D_k}\right) \tag{15}$$

Por el (teorema 5.22) basta probar que para $t\in\mathbb{R}$

$$\varphi_{T_n}(t)\longrightarrow e^{-t^2/2},\qquad n\longrightarrow+\infty$$

Tomemos $t\in\mathbb{R}$ arbitrario pero fijo. Y usando las expansiones

$$e^{iy}=1+iy+\frac{\theta_1 y^2}{2},\qquad e^{iy}=1+iy-\frac{y^2}{2}+\frac{\theta_2|y|^3}{3!}$$

las cuales son ciertas para todo $y\in\mathbb{R}$ con $\theta_1=\theta_1(y)$ y $\theta_2=\theta_2(y)$ tales que $|\theta_1|\leq 1$ y $|\theta_2|\leq 1$ entonces

$$\varphi_k(t)=E[e^{itX_k}]=\int_{\mathbb{R}}e^{itx}\,dF_k(x)=\int_{|x|\geq\varepsilon D_n}\left(1+itx+\frac{\theta_1(tx)^2}{2}\right)dF_k(x)+\int_{|x|<\varepsilon D_n}\left(1+itx-\frac{t^2x^2}{2}+\frac{\theta_2|tx|^3}{6}\right)dF_k(x)$$

$$=1+\frac{t^2}{2}\int_{|x|\geq\varepsilon D_n}\theta_1 x^2\,dF_k(x)-\frac{t^2}{2}\int_{|x|<\varepsilon D_n}x^2\,dF_k(x)+\frac{|t|^3}{6}\int_{|x|<\varepsilon D_n}\theta_2|x|^3\,dF_k(x)$$

Note que se usó que

$$m_K = \int_{-\infty}^{+\infty} x \, dF_K(x) = 0$$

Así, se tiene que

$$\varphi_K\left(\frac{1}{D_n}\right) = 1 - \frac{t^2}{2D_n^2}\int_{|x|<\varepsilon D_n} x^2 dF_K(x) + \frac{t^2}{2D_n^2}\int_{|x|\geq \varepsilon D_n} \theta_1 x^2 dF_K(x) + \frac{|t|^3}{6D_n^3}\int_{|x|<\varepsilon D_n} \theta_2 |x|^3 dF_K(x). \qquad (16)$$

Luego, como

$$\left|\frac{1}{2}\int_{|x|\geq\varepsilon D_n} \theta_1 x^2 dF_K(x)\right| \leq \frac{1}{2}\int_{|x|\geq \varepsilon D_n} x^2 dF_K(x),$$

entonces

$$\frac{1}{2}\int_{|x|\geq\varepsilon D_n} \theta_1 x^2 dF_K(x) = \tilde{\theta}_1 \int_{|x|\geq\varepsilon D_n} x^2 dF_K(x), \qquad (17)$$

donde $\tilde{\theta}_1 = \tilde{\theta}_1(t,K,n)$ con $|\tilde{\theta}_1| \leq 1/2$.

Análogamente

$$\left|\frac{1}{6}\int_{|x|<\varepsilon D_n} \theta_2 |x|^3 dF_K(x)\right| \leq \frac{1}{6}\int_{|x|\leq \varepsilon D_n} \frac{\varepsilon D_n}{|x|} |x|^3 dF_K(x) \leq \frac{1}{6}\int_{|x|<\varepsilon D_n} \varepsilon D_n x^2 dF_K(x)$$

y así

$$\frac{1}{6}\int_{|x|<\varepsilon D_n} \theta_2 |x|^3 dF_K(x) = \tilde{\theta}_2 \int_{|x|<\varepsilon D_n} \varepsilon D_n x^2 dF_K(x), \qquad (18)$$

donde $\tilde{\theta}_2 = \tilde{\theta}_2(t,K,n)$ y $|\tilde{\theta}_2| \leq 1/6$.

Ahora, definiendo

$$A_{Kn} = \frac{1}{D_n^2}\int_{|x|<\varepsilon D_n} x^2 dF_K(x), \qquad B_{Kn} = \frac{1}{D_n^2}\int_{|x|\geq\varepsilon D_n} x^2 dF_K(x).$$

Finalmente, por (16)-(18)

$$\varphi_K\left(\frac{t}{D_n}\right) = 1 - \frac{t^2 A_{Kn}}{2} + t^2\tilde{\theta}_1 B_{Kn} + |t|^3 \varepsilon \tilde{\theta}_2 A_{Kn} \; (= 1 + C_{Kn}) \qquad (19)$$

Por otro lado, note que

$$\sum_{K=1}^{n} (A_{Kn} + B_{Kn}) = 1 \qquad (20)$$

y por hipótesis tenemos que

$$\sum_{K=1}^{n} B_{Kn} \longrightarrow 0 \qquad n \to +\infty. \qquad (21)$$

Por lo tanto, para n suficientemente grande

$$\max_{1\leq K\leq n} |C_{Kn}| \leq t^2\varepsilon^2 + \varepsilon|t|^3 \qquad (22)$$

y

$$\sum_{K=1}^{n} |C_{Kn}| \leq t^2 + \varepsilon|t|^3. \qquad (23)$$

Ahora, como para todo complejo z con $|z| \leq 1/2$

$$\log(1+z) = z + \theta|z|^2$$

donde $\theta = \theta(z)$, $|\theta| \leq 1$ y $\log$ es el valor principal del logaritmo ($\log z = \log|z| + i\arg z$, $-\pi < \arg z \leq \pi$). Así por (19) y (22) para $\varepsilon > 0$ suficientemente pequeño y n suficientemente grande

$$\log \varphi_K\left(\frac{t}{D_n}\right) = \log(1 + C_{Kn}) = C_{Kn} + \theta_{Kn}|C_{Kn}|^2$$

donde $|C_{Kn}| \leq 1$, entonces de (15) se sigue que

$$\frac{t^2}{2} + \log \varphi_{T_n}(t) = \frac{t^2}{2} + \sum_{K=1}^{n} \log\varphi_K\left(\frac{t}{D_n}\right) = \frac{t^2}{2} + \sum_{K=1}^{n} C_{Kn} + \sum_{K=1}^{n} \theta_{Kn}|C_{Kn}|^2.$$

Pero

$$\frac{t^2}{2} + \sum_{K=1}^{n} C_{Kn} = \frac{t^2}{2}\left(1 - \sum_{K=1}^{n} A_{Kn}\right) + t^2\sum_{K=1}^{n} \tilde{\theta}_1(t,K,n) B_{Kn} + \varepsilon|t|^3 \sum_{K=1}^{n} \tilde{\theta}_2(t,K,n) A_{Kn},$$

Ahora, por (20) y (21), para todo $\delta > 0$ podemos encontrar n_0 suficientemente grande y $\varepsilon > 0$ tal que si $n \geq n_0$

$$\left| \frac{t^2}{n^2} + \sum_{k=1}^{n} C_{kn} \right| \leq \frac{\delta}{2}.$$

Luego, por (22) y (23)

$$\left| \sum_{k=1}^{n} \theta_{kn} \, |C_{kn}|^2 \right| \leq \max_{1 \leq k \leq n} |C_{kn}| \cdot \sum_{k=1}^{n} |C_{kn}| \leq (t^2 \varepsilon^2 + \varepsilon |t|^3)(t^2 + \varepsilon |t|^3)$$

Así, para n suficientemente grande, podemos tomar $\varepsilon > 0$ tal que se cumpla que

$$\left| \sum_{k=1}^{n} \theta_{kn} \, |C_{kn}|^2 \right| \leq \frac{\delta}{2}$$

y así

$$\left| \frac{t^2}{2} + \log \varphi_{T_n}(t) \right| \leq \delta.$$

Así, aplicando el (lema 6.2)

$$\varphi_{T_n}(t) \, e^{t^2/2} \longrightarrow 1, \quad n \to +\infty$$

es decir

$$\varphi_{T_n}(t) \to e^{-t^2/2}, \quad n \to +\infty.$$

Teorema 6.6. - Condición de Liapunov- Una condición menos restrictiva nos dice lo siguiente:
pág. 399 [1] si para todo $\delta > 0$:

$$\frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} E|X_k - m_k|^{2+\delta} \underset{n \to +\infty}{\longrightarrow} 0.$$

entonces se tiene los resultados del teorema 6.5 (24)

Demostración

Sea $\varepsilon > 0$, cualquiera; entonces

$$E|X_k - m_k|^{2+\delta} = \int_{-\infty}^{+\infty} |x - m_k|^{2+\delta} \, dF_k(x)$$

$$\geq \int_{\{x : |x - m_k| \geq \varepsilon D_n\}} |x - m_k|^{2+\delta} \, dF_k(x)$$

$$\geq \varepsilon^{\delta} D_n^{\delta} \int_{\{x : |x - m_k| \geq \varepsilon D_n\}} (x - m_k)^2 \, dF_k(x).$$

De donde

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \int_{\{x : |x - m_k| \geq \varepsilon D_n\}} (x - m_k)^2 \, dF_k(x) \leq \frac{1}{\varepsilon^{\delta}} \cdot \frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} E|X_k - m_k|^{2+\delta},$$

es decir, la condición de Liapunov implica la condición de Lindenberg.

Observación. Sea $T_n = (S_n - E[S_n])/D_n$ y $F_{T_n} = P(T_n \leq x)$, entonces la conclusión del teorema 6.5, nos dice que

$$F_{T_n}(x) \underset{n \to +\infty}{\longrightarrow} \Phi(x)$$

donde $\Phi \sim N(0,1)$. Luego, como $\Phi(x)$ es continua, entonces la convergencia es uniforme, es decir

$$\| F_{T_n} - \Phi \| = \sup_{x \in \mathbb{R}} | F_{T_n}(x) - \Phi(x) | \underset{n \to +\infty}{\longrightarrow} 0.$$

En particular, esto implica que

$$P(S_n \leq x) - \Phi\left((x - E[S_n]) / D_n \right) \underset{n \to +\infty}{\longrightarrow} 0$$

Esto se obtiene pues para n suficientemente grande, S_n es aproximadamente $N(E[S_n], Var(S_n) = D_n^2)$.

Capítulo 7: Condicionamiento.

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Si Λ es cualquier conjunto en $\mathcal{F}$ con $P(\Lambda)>0$, podemos definir

$$P_\Lambda(\cdot): \mathcal{F} \longrightarrow [0,1]$$
$$E \longmapsto P_\Lambda(E) = \frac{P(\Lambda \cap E)}{P(\Lambda)}. \tag{1}$$

Así, claramente, P_Λ define una medida de probabilidad en $\mathcal{F}$ denominada «probabilidad condicional relacionada a Λ». La integral con respecto a esta medida de probabilidad se denomina «experanza condicional relativa a Λ»

$$E_\Lambda[Y] = \int_\Omega Y(\omega)\, P_\Lambda(d\omega) = \frac{1}{P(\Lambda)} \int_\Lambda Y(\omega)\, P(d\omega) \tag{2}$$

Si $P(\Lambda)=0$, definimos que $P_\Lambda(E)=0$ para todo $E \in \mathcal{F}$.

Ahora, sea $\{\Lambda_n : n \geq 1\}$ una partición medible contable de Ω, notada

$$\Omega = \bigcup_{n=1}^{+\infty} \Lambda_n \qquad \Lambda_n \in \mathcal{F} \qquad \Lambda_m \cap \Lambda_n = \emptyset \quad \text{si } n \neq m.$$

Entonces, tenemos que

$$P(E) = \sum_{n=1}^{+\infty} P(\Lambda_n \cap E) = \sum_{n=1}^{\infty} P(\Lambda_n)\, P_{\Lambda_n}(E) \tag{3}$$

$$E[Y] = \sum_{n=1}^{+\infty} \int_{\Lambda_n} Y(\omega) P(d\omega) = \sum_{n=1}^{+\infty} P(\Lambda_n)\, E_{\Lambda_n}[Y] \tag{4}$$

siempre que $E[Y]$ esté bien definida.

Sea $\mathcal{G}$ la sub σ-álgebra generada por una partición contable $\{\Lambda_n\}$. (Por ejemplo si consideramos la variable aleatoria discreta X, con $\Lambda_n = \{X = a\}$). Dada Y una v.a. integrable, definimos la función $E_\mathcal{G}[Y]$ en Ω por

$$E_\mathcal{G}[Y] = \sum_n E_{\Lambda_n}[Y]\, \mathbb{1}_{\Lambda_n}(\cdot) \tag{5}$$

Por lo tanto, $E_\mathcal{G}[Y]$ es una v.a. discreta que devuelve el valor de $E_{\Lambda_n}[Y]$ en el conjunto Λ_n, para cada n. Así, podemos reescribir (4) como

$$E[Y] = \sum_n \int_{\Lambda_n} E_\mathcal{G}[Y]\, dP = \int_\Omega E_\mathcal{G}(Y)\, dP.$$

Más aún, para $\Lambda \in \mathcal{G}$, con Λ la unión de una subcolección de Λ_n y la misma manipulación, se tiene que

$$\forall \Lambda \in \mathcal{G}: \int_\Lambda Y\, dP = \int_\Lambda E_\mathcal{G}[Y]\, dP \tag{6}$$

Sean φ_1 y φ_2 funciones $\mathcal{G}$-medibles tales que

$$\forall \Lambda \in \mathcal{G}: \int_\Lambda Y\, dP = \int_\Lambda \varphi_i\, dP \qquad i=1,2.$$

Definamos $\Lambda = \{\omega : \varphi_1(\omega) > \varphi_2(\omega)\}$, entonces $\Lambda \in \mathcal{G}$ y por lo tanto

$$\int_\Lambda (\varphi_1 - \varphi_2)\, dP = 0.$$

Así, $P(\Lambda)=0$ e intercambiando φ_1 y φ_2, tenemos que $\varphi_1=\varphi_2$ c.s.
Notaremos $E_{\mathcal{G}}[Y]=E[Y\mid\mathcal{G}]$ para las clases de equivalencia y llamaremos a todo miembro de dicha clase lo denominaremos una «versión» de la esperanza condicional.

Teorema 7.1 Si $E[Y]<+\infty$ y $\mathcal{G}$ es una sub σ-álgebra de $\mathcal{F}$, entonces existe una única clase de equivalencias de variables aleatorias $E[Y,\mathcal{G}]$, $\mathcal{G}$-medibles tales que

$$\forall \Lambda\in\mathcal{G},\ \int_\Lambda Y\,dP=\int_\Lambda E[Y\mid\Lambda]\,dP$$

Demostración
Sea $\Lambda\in\mathcal{G}$, cualquiera; definamos

$$v(\Lambda)=\int_\Lambda Y\,dP$$

Notemos que es finita y contablemente aditiva; por lo tanto es una «medida con signo» en $\mathcal{G}$. Si $P(\Lambda)=0$, entonces $v(\Delta)=0$; y así, es absolutamente continua con respecto a P: $v\ll P$. Así, el teorema se sigue del teorema de Radon-Nikodym y así, la «derivada resultante» cumple que

$$\frac{dv}{dP}=E[Y\mid\mathcal{G}]$$

Definición 7.2. -Esperanza condicional-
Sea Y una v.a. integrable y $\mathcal{G}\subseteq\mathcal{F}$ una sub σ-álgebra, la esperanza condicional $E[Y\mid\mathcal{G}]$ de Y relativa a $\mathcal{G}$ es cualquiera de las clases de equivalencia de las v.a. en Ω tales que
a) Pertenecen a $\mathcal{G}$ (son $\mathcal{G}$-medible)
b) Para todo $\Lambda\in\mathcal{G}$

$$\int_\Lambda Y\,dP=\int_\Lambda z\,dP$$

Notemos que (6) es equivalente a (b)

En la definición 7.2 si $Y=\mathbb{1}_\Delta$, donde $\Delta\in\mathcal{F}$, entonces

$$P(\Delta\mid\mathcal{G})=E[\mathbb{1}_\Delta\mid\mathcal{G}].$$

y la denominaremos la esperanza condicional de Δ relativa a $\mathcal{G}$. Específicamente $P(\Delta\mid\mathcal{G})$ es una de las clases de equivalencia de v.a. pertenecientes a $\mathcal{G}$ y tales que

$$\forall\Lambda\in\mathcal{G}:\ P(\Delta\cap\Lambda)=\int_\Lambda P(\Delta\mid\mathcal{G})\,dP \tag{7}$$

Así, se sigue de la definición de v.a. de Y (integrable) y una sub σ-álgebra $\mathcal{G}$, entonces

$$\int_\Lambda\{[Y-E[Y\mid G]]\,z\}=0.$$

para todo $\Lambda \in \mathcal{G}$, y en consecuencia

$$E[\{Y - E[Y|\mathcal{G}]\}Z] = 0$$

para todo $Z \in \mathcal{G}$, acotado. Así, una variable aleatoria integrable Y se descompone de forma «ortogonal» de la siguiente forma

$$Y = Y' + Y''$$

donde $Y' = E[Y|\mathcal{G}]$ y Y'' tal que $Y'' \perp \mathcal{G}$; es decir $E[Y''Z] = 0$ para toda $Z \in \mathcal{G}$ medible y acotada.
En otras palabras Y' es la proyección de Y en $\mathcal{G}$ y Y'' es su complemento ortogonal

Lema 7.3 Si Z es $\sigma(X)$ medible, entonces $Z = \varphi(X)$ para alguna función φ Borel medible.

Demostración:
Para probar el resultado, es suficiente probarlo para una función acotada y positiva. Así, sea Z una función acotada y positiva, entonces existe una sucesión de funciones simples (Z_m) creciente a Z en todas partes. y cada Z_m es de la forma

$$Z_m = \sum_{j=1}^{\ell} c_j \mathbb{1}_{\Lambda_j}$$

donde $\Lambda \in \sigma(X)$. Por lo tanto, $\Lambda_j = X^{-1}(B_j)$ para algún $B_j \in Bor(\mathbb{R})$ (ver ejer 11 sec 3.1 [3]).
Así, si tomamos

$$\varphi_m = \sum_{j=1}^{\ell} c_j \mathbb{1}_{B_j},$$

tenemos que $Z_m = \varphi_m(X)$ con φ_m Borel medible, luego como $\varphi_m(X) \longrightarrow Z$, entonces φ_m converge en el rango de X; no obstante, el límite no es necesariamente Borel o Lebesgue medible; para solucionar esto, notemos que

$$\forall x \in \mathbb{R}: \quad \varphi(x) = \limsup_{m \to +\infty} \varphi_m(x)$$

es Borel medible; así

$$Z = \limsup_{m \to \infty} \varphi_m(X) = \varphi(X).$$

Para probar la segunda afirmación, dado $B \in Bor(\mathbb{R})$ y sea $\Lambda = X^{-1}(B)$, entonces por el (teorema 14) del capítulo 2

$$\int_\Lambda E[Y|X]\, dP = \int_\Omega \mathbb{1}_B(X)\varphi(X)\, dP = \int_{\mathbb{R}} 1_B(x)\varphi(x)\, d\mu = \int_B \varphi(x)\, d\mu.$$

y finalmente por (6)

$$\lambda(B) = \int_\Lambda Y\, dP = \int_B \varphi(x)\, d\mu. \qquad \forall B \in Bor(\mathbb{R})$$

Así se sigue que φ es una versión de la derivada $d\lambda/d\mu$ lo que completa la demostración.

Teorema 7.4 Una versión de la esperanza condicional $E[Y|X]$ está dada por $\varphi(X)$, donde φ es una función Borel medible en $\mathbb{R}$. Más aún, si definimos la medida con signo λ en $Bor(\mathbb{R})$ como

$$\forall B \in Bor(\mathbb{R}): \quad \lambda(B) = \int_{X^{-1}(B)} Y\, dP$$

y la medida de probabilidad μ de X, entonces φ es una versión de la derivada de Radon-Nikodym, $d\lambda/d\mu$.

Demostración: La demostración del teorema se obtiene del lema anterior.
Como consecuencia del teorema, la función $E[X|Y](\omega)$ es constante c.s. en cada conjunto en el que $X(\omega)$ lo sea. Así, como un abuso de notación podemos decir que $\varphi(x)$ se puede escribir como $E[Y|X=x]$. Por ejemplo, para un real c

$$\int_{\{X\leq c\}} Y\,dP = \int_{]-\infty,c]} E[Y|X=x]\;P(X\leq x)$$

Teorema 7.5 Sean Y y YZ v.a. y $Z\in\mathcal{G}$ (i.e. Y y YZ son $\mathcal{G}$-medibles), entonces

$$E[YZ|\mathcal{G}] = Z\,E[Y|\mathcal{G}] \quad \text{c.s.} \qquad (8)$$

Demostración:
Supongamos que $Y\geq 0$, $Z\geq 0$ la demostración consiste en observar que el lado derecho de (8) es $\mathcal{G}$-medible y satisface la ecuación de la definición

$$\forall \Lambda\in\mathcal{G}: \int_\Lambda Z\,E[Y|\mathcal{G}]\,dP = \int_\Lambda ZY\,dP \qquad (9)$$

Para (9) es cierto si $Z=\mathbb{1}_\Delta$ con $\Delta\in\mathcal{G}$, por lo tanto, es cierto si Z es una v.a simple perteneciente a $\mathcal{G}$ (i.e. es $\mathcal{G}$-medible) y en consecuencia para cada Z en $\mathcal{G}$ por convergencia monótona donde los límites son finitos o infinitos positivos. Note que la integrabilidad de $Z\,E[Y|\mathcal{G}]$ es parte de las hipótesis del teorema.

Propiedades de la esperanza condicional

1) Si X es $\mathcal{G}$-medible, entonces $E[X|\mathcal{G}]=X$ c.s., esto es cierto en particular si X es constante c.s.
2) $E[X_1+X_2|\mathcal{G}] = E[X_1|\mathcal{G}]+E[X_2|\mathcal{G}]$
3) Si $X_1\leq X_2$, entonces $E[X_1|\mathcal{G}]\leq E[X_2|\mathcal{G}]$
4) $|E[X|\mathcal{G}]|\leq E[|X||\mathcal{G}]$
5) Si $X_n\uparrow X$, entonces $E[X_n|\mathcal{G}]\uparrow E[X|\mathcal{G}]$
6) Si $X_n\downarrow X$, entonces $E[X_n|\mathcal{G}]\downarrow E[X|\mathcal{G}]$
7) Si $|X_n|\leq Y$ donde $E[Y]<+\infty$ y $X_n\to X$, entonces $E[X_n|\mathcal{G}]\to E[X|\mathcal{G}]$

Desigualdad de Cauchy-Schwarz-Bunyakovsky

$$E[|XY||\mathcal{G}]^2\leq E[X^2|\mathcal{G}]\,E[Y^2|\mathcal{G}]$$

Teorema 7.6. Sea φ una función convexa en $\mathbb{R}$ y X y $\varphi(X)$ v.a. integrable, entonces para cada $\mathcal{G}$

$$\varphi(E[X|\mathcal{G}])\leq E[\varphi(X)|\mathcal{G}]$$

La demostración de este resultado se encuentra en la hoja 11 - ejercicio 2.

Teorema 7.7. Si Y es integrable y $\mathcal{F}_1\subseteq\mathcal{F}_2$, entonces

$$E[Y|\mathcal{F}_1] = E[Y|\mathcal{F}_2] \quad \text{si y solo si } E_{\mathcal{F}_2}(Y)\in\mathcal{F}_1 \qquad (10)$$

y

$$E[E[Y|\mathcal{F}_2]|\mathcal{F}_1] = E[Y|\mathcal{F}_1] = E[E[Y|\mathcal{F}_1]|\mathcal{F}_2] \qquad (11)$$

Demostración
Es fácil ver que Y satisface trivialmente la relación para $E[Y|\mathcal{F}_1]$; esta será cierta si y solo si $Y\in\mathcal{F}_1$

Ahora, si reemplazamos $\mathcal{F}$ por $\mathcal{F}_2$ y Y por $E[Y|\mathcal{F}_2]$, se obtiene (10).
Por otro lado, como

$$E[Y|\mathcal{F}_1] \in \mathcal{F}_1 \subseteq \mathcal{F}_2$$

la segunda ecuación en (11) se obtiene por la misma observación. Así, basta probar la primera ecuación en (11). Sea $\Lambda \in \mathcal{F}_1$, entonces $\Lambda \in \mathcal{F}_2$ y aplicando la definición

$$\int_\Lambda E[E[Y|\mathcal{F}_2]|\mathcal{F}_1]\, dP = \int_\Lambda E[Y|\mathcal{F}_2]\, dP = \int_\Lambda Y\, dP$$

Por lo tanto, $E[E[Y|\mathcal{F}_2]|\mathcal{F}_1]$ satisface la relación para $E[Y|\mathcal{F}_1]$ y como esta es $\mathcal{F}_1$-medible se sigue el resultado.

Un caso particular que podemos notar es

$$E[E[Y|X_1,X_2]|X_1] = E[Y|X_1] = E[E[Y|X_1]|X_1,X_2]$$

Sección 7.1: Independencia condicional

Definición 7.8. Sea $\{\mathcal{F}_\alpha\}_{\alpha\in A}$ un conjunto de sub-álgebras de $\mathcal{F}$ con A un conjunto arbitrario de índices. Se dice que las sub-álgebras son condicionalmente independientes respecto a una sub-álgebra $\mathcal{G} \subseteq \mathcal{F}$ si y solo si para cualquier colección finita de conjuntos $\Lambda_1, \dots, \Lambda_n$, tal que $\Lambda_j \in \mathcal{F}_{\alpha_j}$ con $\alpha_j \in A$ distintos se tiene que

$$P\left(\bigcap_{j=1}^{n} \Lambda_j \,\Big|\, \mathcal{G}\right) = \prod_{j=1}^{n} P(\Lambda_j|\mathcal{G}) \qquad (12)$$

Notemos que cuando $\mathcal{G}$ es la sub-álgebra trivial, (12) se reduce a la independencia definida en capítulos anteriores.

Teorema 7.9. Para cada $\alpha\in A$ denotemos $\mathcal{F}^{(\alpha)}$ la sub-álgebra más pequeña que contiene todos los $\mathcal{F}_\beta$, con $\beta \in A-\{\alpha\}$, es decir

$$\mathcal{F}^{(\alpha)} = \sigma\left(\bigcup_{\beta\in A-\{\alpha\}} \mathcal{F}_\beta\right).$$

Entonces, los $\mathcal{F}_\alpha$ con $\alpha\in A$ son condicionalmente independientes relativo a $\mathcal{G}$ si y solo si para cada α y $\Lambda_\alpha \in \mathcal{F}_\alpha$, tenemos que

$$P(\Lambda_\alpha | \mathcal{F}^{(\alpha)} \vee \mathcal{G}) = P(\Lambda_\alpha|\mathcal{G}),$$

donde $\mathcal{F}^{(\alpha)} \vee \mathcal{G}$ denota el σ-álgebra más pequeña que contiene a $\mathcal{F}^{(\alpha)}$ y $\mathcal{G}$, es decir $\mathcal{F}^{(\alpha)}\vee\mathcal{G} = \sigma(\mathcal{F}^{(\alpha)} \cup \mathcal{G})$