Notas para el curso de Probabilidad II
Licenciatura en Estadística

Alejandro Cholaquidis

Facultad de Ciencias
Universidad de la República

Estas notas fueron hechas para el curso de Probabilidad 2 de la Licencitura en Estadística, dictado en el año 2015. Las erratas que hubieren, se agradece comunicarlas a acholaquidis@hotmail.com.

Índice general

Capítulo 1

Repaso

Definiremos a modo de repaso los principales conceptos y enunciaremos los resultados necesarios para la lectura de los siguientes capítulos, en general se dan sin demostración. Las demostraciónes de los mismos pueden encontrarse en [5], [4] o en cualquier libro introductorio a la probabilidad.

1.1. Espacio de Probabilidad

1.1.1. σ-álgebra

Definición 1.1. Dado un conjunto $\Omega \neq \emptyset$, diremos que $\mathcal{A} \subset 2^{\Omega}$ es una σ-álgebra de subconjuntos de Ω si se cumple:

- $\Omega \in \mathcal{A}$.
- Si $A \in \mathcal{A}$ entonces $A^c \in \mathcal{A}$.
- Si $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ entonces $\cup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Las siguientes propiedades son de fácil demostración, se sugiere hacerlas como ejercicio.

Proposición 1.2. *Si $\mathcal{A}$ es una σ-álgebra,*

1) $\emptyset \in \mathcal{A}$.

2) $A_1, A_2, \ldots, A_n \in \mathcal{A}$ *entonces* $\cup_{i=1}^{n} A_i \in \mathcal{A}$

3) Si $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ *entonces* $\cap_{n=1}^{+\infty} A_n \in \mathcal{A}$.

4) Si $A, B \in \mathcal{A}$ *entonces* $A - B \in \mathcal{A}$

5) Si $\mathcal{A}_\alpha$ *es* σ*-álgebra de conjuntos sobre* Ω *para todo* $\alpha \in I$*, siendo* I *un conjunto cualquiera de índices, entonces* $\cap_{\alpha\in I}\mathcal{A}_\alpha$ *es* σ*-álgebra de conjuntos sobre* Ω.

Ejemplo 1.3. Ejemplos de sígma álgebra:

1) $\{\emptyset, \Omega\}$.

2) 2^{Ω}.

3) $\{\emptyset, \Omega, A, A^c\}$, siendo $\emptyset \subsetneq A \subsetneq \Omega$.

Definición 1.4. Dada una famila F de subconjuntos de Ω, el conjunto $\cap_{\mathcal{A}:F\subset\mathcal{A}}\mathcal{A}$ (es decir la intersección de todas las σ-álgebra que contienen a F) se llama σ-álgebra generada por F y se denota $\sigma(F)$.

Definición 1.5. La σ-álgebra de Borel en $\mathbb{R}$ es $\sigma(F)$ con $F = \{A \subset \mathbb{R} : A \text{ es abierto}\}$.

Se deja como ejercicio la siguiente proposición.

Proposición 1.6. *Sea $f : \Omega \to \Omega'$, y D una familia de subconjuntos de Ω', entonces $\sigma(f^{-1}(D)) = f^{-1}(\sigma(D))$, donde denotamos, para $B \in D$ $f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\}$ la preimagen de B por f.*

Teorema 1.7. *Si*

1) $I_1 = \{(a,b) \subset \mathbb{R} : a < b\}$

2) $I_2 = \{[a,b) \subset \mathbb{R} : a < b\}$

3) $I_3 = \{(a,b] \subset \mathbb{R} : a < b\}$

4) $I_4 = \{(a,+\infty) \subset \mathbb{R} : a \in \mathbb{R}\}$

5) $I_5 = \{[a,\infty) \subset \mathbb{R} : a \in \mathbb{R}\}$

6) $I_6 = \{(-\infty,a) \subset \mathbb{R} : a \in \mathbb{R}\}$

7) $I_7 = \{(-\infty,a] \subset \mathbb{R} : a \in \mathbb{R}\}$

entonces la σ-álgebra de Borel definida anteriormente es igual a la σ-álgebra generada por cualquiera de las familias de conjuntos $I_1, \ldots, I_7$.

1.1.2. Espacio de probabilidad

Definición 1.8. Dado $\Omega \neq \emptyset$, diremos que la terna $(\Omega, \mathcal{A}, P)$ es un espacio de probabilidad sobre Ω si y sólo si $\mathcal{A}$ es una σ-álgebra de conjuntos sobre Ω y P es una función $P : \mathcal{A} \to [0,1]$ que cumple

i) $P(\Omega) = 1$

ii) si la familia de sucesos $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ son disjuntos dos a dos ($A_i \cap A_j = \emptyset$ para todo $i \neq j$), entonces

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

Proposición 1.9. *1)* $P(\emptyset) = 0$

2) Si $A_1, \ldots, A_n \in \mathcal{A}$ son disjuntos dos a dos $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.

3) Si $A, B \in \mathcal{A}$, entonces $P(B - A) = P(B) - P(A \cap B)$.

4) Si $A, B \in \mathcal{A}$, $A \subset B$, $P(B - A) = P(B) - P(A)$ y $P(A) \leq P(B)$

5) Si $A, B \in \mathcal{A}$, entonces $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

6) Si $A_1, \ldots, A_n \in \mathcal{A}$, entonces $P(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$.

7) Si $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ es tal que $A_1 \subset A_2 \subset A_3 \subset \ldots$, entonces $P(\cup_{n\in\mathbb{N}}) = \lim P(A_n)$.

8) Si $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ es tal que $A_1 \supset A_2 \supset A_3 \supset \ldots$, entonces $P(\cap_{n\in\mathbb{N}} A_n) = \lim P(A_n)$.

9) Si $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ es tal que $P(A_n) = 1$ para todo n, entonces $P(\cap_{n\in\mathbb{N}} A_n) = 1$.

1.1.3. Límite superior e inferior

Recordemos que si $\{a_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ es una sucesión de números reales, el límite superior e inferior de la misma se define como

$$\limsup_n a_n = \lim_{n\to+\infty} \sup_{k\geq n} a_k \quad \text{y} \quad \liminf_n a_n = \lim_{n\to\infty} \inf_{k\geq n} a_k,$$

respectivamente. Se puede probar fácilmente que $\limsup_n a_n = \inf_n \sup_{k\geq n} a_k$ y $\liminf a_n = \sup_n \inf_{k\geq n} a_n$. Este concepto se generaliza a conjuntos de la siguiente manera:

Definición 1.10. Dada $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$, se definen el límite superior e inferior de la sucesión de sucesos como

$$\limsup_n A_n = \bigcap_{n=1}^{+\infty} \bigcup_{k=n}^{+\infty} A_k \quad \text{y} \quad \liminf_n A_n = \bigcup_{n=1}^{+\infty} \bigcap_{k=n}^{+\infty} A_k$$

respectivamente.

Algunas propiedades importantes:

Proposición 1.11. *1)*

$$\limsup_n A_n = \{\omega \in \Omega : \omega \in A_n \textit{ para infinitos valores de } n\}$$

(ocurren infinitos A_n).

2)

$$\liminf_n A_n = \{\omega \in \Omega : \omega \in A_n \textit{ para todo } n \textit{ salvo a lo sumo una cantidad finita de índices}\}$$

(ocurre A_n para todos los n salvo una cantidad finita).

3) $\liminf A_n \subset \limsup A_n$.

4) Como la sucesión $B_n = \bigcup_{k=n}^{+\infty} A_k$ es decreciente, entonces

$$P(\limsup A_n) = \lim_{n\to\infty} P\left(\bigcup_{k=n}^{+\infty} A_k\right).$$

5) Como la sucesión $B_n = \bigcap_{k=n}^{+\infty} A_k$ es creciente, entonces

$$P(\liminf A_n) = \lim_{n\to\infty} P\left(\bigcap_{k=n}^{+\infty} A_k\right).$$

6) Si $\{A_n\}_{n\in\mathbb{N}}$ es creciente, entonces

$$\liminf_n A_n = \limsup_n A_n = \bigcup_{n=1}^{+\infty} A_n.$$

7) Si $\{A_n\}_{n\in\mathbb{N}}$ es decreciente, entonces

$$\liminf A_n = \limsup A_n = \bigcap_{n=1}^{+\infty} A_n.$$

8)

$$P(\liminf A_n) \leq \liminf P(A_n) \leq \limsup P(A_n) \leq P(\limsup A_n).$$

1.2. Independencia

Definición 1.12. Dada una familia de sucesos $\{A_\alpha\}_{\alpha\in I}$, donde I es un conjunto cualquiera de índices, se dice que son independientes si y sólo si, para todo $F \subset I$ finito, se cumple que

$$P\left(\bigcap_{\alpha\in F} A_\alpha\right) = \prod_{\alpha\in F} P(A_\alpha).$$

Teorema 1.13. ***(Lema de Borel-Cantelli).*** *Dada una sucesión $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$,*

1) Si $\sum_{n=1}^{+\infty} P(A_n) < +\infty$ entonces

$$P(\limsup A_n) = 0$$

2) Si $\sum_{i=1}^{+\infty} P(A_n) = \infty$ y además $\{A_n\}_{n\in\mathbb{N}}$ son independientes, entonces

$$P(\limsup A_n) = 1.$$

Demostración.

1. $P(\limsup A_n) = \lim_n P(\cup_{k=n}^{+\infty} A_k) \leq \sum_{k=n}^{+\infty} P(A_n) \to_n 0$, donde el último límite es 0 porque la serie converge.

2. Como $P\left(\limsup A_n\right) = \lim P\left(\cup_{k=n}^{+\infty} A_k\right)$ basta probar que $P\left(\cap_{k=n}^{+\infty} A_k^c\right) \to 0$. Para cada $m > n$ tenemos que

$$P\left(\bigcap_{k=n}^{+\infty} A_k^c\right) \leq P\left(\bigcap_{k=n}^{m} A_k^c\right) = \prod_{k=n}^{m} P(A_k^c) = \prod_{k=n}^{m} \big(1 - P(A_k)\big) \leq \prod_{k=n}^{m} e^{-P(A_k)} = e^{-\sum_{k=n}^{m} P(A_k)} \to_m 0,$$

donde en la penúltima desigualdad hemos usado que $1 - x \leq e^{-x}$ para todo x.

$\square$

1.3. Variables aleatorias

Vamos a introducir ahora uno de los principales conceptos de la probabilidad, la noción de variable aleatoria. Dado que en varias aplicaciones hay que considerar variables aleatorias como funciones no sólo a valores en $\mathbb{R}^d$, sino más en general, en un espacio métrico, definiremos algunas nociones topológicas previas.

1.3.1. Espacio métrico

Definición 1.14. Una métrica definida en un conjunto F es una función $d : F \times F \to \mathbb{R}$ tal que

1) $d(x, y) \geq 0$ y $d(x, y) = 0$ si y sólo si $x = y$.

2) $d(x, y) = d(y, x)$

3) para todo $z \in F$, $d(x, y) \leq d(x, z) + d(z, y)$

Ejemplo 1.15. Ejemplos clásicos de espacios métricos son:

1) $d(x, y) = 1$ si $x \neq y$ y $d(x, y) = 0$ si y sólo si $x = y$.

2) $F = \mathbb{R}^d$ y $d(x, y) = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$.

3) $F = \mathbb{R}^d$ y $d(x,y) = \sum_{i=1}^d |x_i - y_i|$.
4) $F = \mathbb{R}^d$ y $d(x,y) = \max_{i=1,\dots,d} |x_i - y_i|$.
5) $F = \mathbb{C}([0,1])$ funciones continuas de $[0,1]$ en $\mathbb{R}$ y $d_\infty(f,g) = \max_{x\in[0,1]} |f(x) - g(x)|$.
6) $F = L^p([0,1]) = \{f : [0,1] \to \mathbb{R}, \int_0^1 f(x)^p dx < \infty\}$ y $d(f,g) = \sqrt[p]{\int_0^1 |f(x) - g(x)|^p dx}$.
7) Si $D = \big\{f : [0,1] \to \mathbb{R}, \forall t \in (0,1), \exists \lim_{x\to t^-} f(x),\ \text{y} \ \lim_{x\to t^+} f(x) = f(t)\big\}$, la métrica de Skorohod en D se define como

$$d_{Sk}(f,g) = \inf\Big\{\varepsilon > 0 : \exists\lambda \in \Lambda : \sup_{t\in[0,1]} |f(t) - g(\lambda(t))| + \sup_{t\in[0,1]} |t - \lambda(t)| \leq \varepsilon\Big\},$$

donde Λ es el conjunto de funciones λ definidas en $[0,1]$ a valores en $[0,1]$ continuas, estrictamente crecientes, tal que $\lambda(0) = 0$ y $\lambda(1) = 1$. Observar que $\mathbb{C}([0,1]) \subset D$ y $B_{d_\infty}(f,r) \subset B_{d_{Sk}}(f,r)$.

Definición 1.16. Dado un espacio métrico (F,d) se dice que un conjunto $A \subset F$ es abierto si para todo $a \in A$, existe $r > 0$ tal que $B(a,r) \subset A$, siendo $B(a,r) = \{x \in F : d(a,x) < r\}$. Se dice que A es cerrado si A^c es abierto.

Proposición 1.17. *Dado un espacio métrico (F,d), se verifica*

1. *$\emptyset$ y F son conjuntos abiertos.*
2. *Si $\{F_\alpha\}_{\alpha\in I} \subset F$ es una familia cualquiera de conjuntos abiertos entonces $\cup_\alpha F_\alpha$ es abierto.*
3. *Para todo $k > 0$ $F_1,\dots,F_k$ son abiertos $F_1 \cap \dots \cap F_k$ es abierto.*

Observemos que en un espacio métrico (F,d) se puede definir convergencia de sucesiones de la misma manera que se hace en $\mathbb{R}^d$, es decir si $\{x_n\}_{n\in\mathbb{N}} \subset F$ es una sucesión, $x_n \to x$ si $d(x_n,x) \to 0$. Decimos que una sucesión $\{x_n\}_{n\in\mathbb{N}} \subset F$ es de Cauchy si $d(x_n,x_m) \to 0$ cuando $n,m \to \infty$. Se dice que el espacio métrico es completo si para toda $\{x_n\}_{n\in\mathbb{N}} \subset F$ de Cauchy existe $x \in F$ tal que $x_n \to x$. Se dice que $f : F \to G$ (siendo (G,ρ) otro espacio métrico) es continua en $x \in F$ si para toda $x_n \to x$, $\rho(f(x_n),f(x)) \to 0$. Se puede probar que $f : F \to G$ es continua si y sólo si para todo $B \in G$ abierto, $f^{-1}(B) = \{a \in F : f(a) \in B\}$ es abierto en F.

Definición 1.18. Dado un espacio métrico (F,d) la σ-álgebra de Borel de F es la intersección de todas las σ-álgebras que contienen a los conjuntos abiertos. En general se denota $\mathcal{B}(F)$.

Definición 1.19. Dado un espacio de probabilidad $(\Omega,\mathcal{A},P)$ y (F,d) un espacio métrico, diremos que $X : \Omega \to F$ es una variable aleatoria si para todo $B \in \mathcal{B}(F)$, se cumple que

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}.$$

Proposición 1.20. *Para verificar que X es una variable aleatoria, basta probar que $X^{-1}(B) \in \mathcal{A}$, para todo $B \subset F$, abierto.*

Demostración. Sabemos que, por definición $\mathcal{B}(F)$ es la σ-álgebra generada por los abiertos. Queremos probar que $X^{-1}(B) \in \mathcal{A}$ para todo $B \in \mathcal{B}(F)$. Observemos que la familia $\mathcal{D}$ de Borelianos para los cuales $X^{-1}(D) \in \mathcal{A}$, $D \in \mathcal{D}$, es una σ-álgebra (esto se sigue de la proposición 1.6) que contiene por hipótesis a los abiertos, por lo tanto contiene a la σ-álgebra generada por los abiertos, es decir a los Borelianos. En definitiva $\mathcal{B}(F) \subset \mathcal{D} \subset \mathcal{B}(F)$. □

Teorema 1.21. *Sean (F,d) y (G,ρ) espacios métricos. Si $X : \Omega \to F$ es una variable aleatoria y $g : F \to G$ es una función continua entonces $Y = g(X)$ es una variable aleatoria a valores en G.*

Demostración. Es una consecuencia inmediata de la observación anterior ya que, para todo $A \subset G$ abierto, $g^{-1}(A)$ es abierto y, por otra parte

$$Y^{-1}(A) = X^{-1}(g^{-1}(A)) \in \mathcal{A}.$$

$\square$

Corolario 1.22. *Si $X, Y : \Omega \to \mathbb{R}$ son variables aleatorias y $\alpha \in \mathbb{R}$ entonces $X + \alpha Y$, XY son variables aleatorias.*

Observación 1.23. *En algunos casos es necesario considerar variables aleatorias a valores no sólo en $\mathbb{R}$ sino en la recta extendida $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$, para eso, si denotamos como $\mathcal{B}$ la σ-álgebra de Borel en $\mathbb{R}$, definimos una σ-álgebra $\mathcal{B}(\overline{\mathbb{R}})$ como*

$$\mathcal{B} \cup \big\{A \cup \{+\infty, -\infty\}\} : A \in \mathcal{B}\big\} \cup \big\{A \cup \{+\infty\} : A \in \mathcal{B}\big\} \cup \big\{A \cup \{-\infty\} : A \in \mathcal{B}\big\}.$$

Muchas veces usaremos la notación $X \in A$ para indicar el conjunto: $\{\omega \in \Omega : X(\omega) \in A\}$. Por ejemplo $X^{-1}((-\infty, a])$ es $\{X \leq a\}$.

Teorema 1.24. *Si $X_n : \Omega \to \mathbb{R}$ es variable aleatoria para todo $n \in \mathbb{N}$ entonces también lo son las variables a valores en $\overline{\mathbb{R}}$, $\sup_n X_n$ y $\operatorname{ínf}_n X_n$.*

Demostración.

$$\Big(\sup_n X_n\Big)^{-1}\big((-\infty, a]\big) = \bigcap_{n=1}^{+\infty} X_n^{-1}\big((-\infty, a]\big) \in \mathcal{A}.$$

$\square$

Corolario 1.25. *Si $X_n : \Omega \to \mathbb{R}$ es variable aleatoria para todo $n \in \mathbb{N}$ entonces también lo son las variables a valores en $\overline{\mathbb{R}}$*

$$\limsup_n X_n \quad y \quad \liminf_n X_n.$$

Demostración. Se sigue del teorema anterior, usando que $\limsup_n X_n = \operatorname{ínf}_n \sup_{k \geq n} X_n$. Y lo mismo para $\liminf_n X_n$ $\square$

Definición 1.26. Dadas dos variables aleatorias $X, Y : \Omega \to F$ decimos que

1. X es igual a Y en distribución, (denotamos $X \stackrel{d}{=} Y$) si $\forall A \in \mathcal{B}(F)$ $P(X \in A) = P(Y \in A)$
2. X es igual a Y casi seguramente, (denotamos $X \stackrel{c.s}{=} Y$) si existe $C \subset \Omega$ con $P(C) = 1$ tal que para todo $\omega \in C$, $X(\omega) = Y(\omega)$.

Observación 1.27. *Observemos que si $X \stackrel{c.s}{=} Y$ entonces $X \stackrel{d}{=} Y$, probar con un contraejemplo que el recíproco no es cierto.*

1.3.2. Ley 0 o 1 de Kolmogorov

Dado $(\Omega, \mathcal{A}, P)$ y $X_1, \dots, X_n, \dots$, una sucesión de variables aleatorias, sea $\mathcal{A}_n^\infty = \sigma(X_n, X_{n+1}, \dots)$ la σ-álgebra generada por $X_n, X_{n+1}, \dots$ (es decir la menor σ-álgebra que las hace variable aleatoria), denotamos

$$\mathcal{A}^\infty = \bigcap_{n=1}^{\infty} \mathcal{A}_n^\infty,$$

es claro que $\mathcal{A}^\infty$ es una σ-álgebra. Los eventos de $\mathcal{A}^\infty$ se llaman *eventos de cola*, ya que no dependen de lo que sucede en la sucesión $\{X_n\}_{n \in \mathbb{N}}$ en una cantidad finita de índices. Algunos ejemplos de eventos en $\mathcal{A}^\infty$ son

1. Para todo $k \geq 1$, $A_1 = \left\{\sum_{n=1}^{\infty} \frac{\xi_n}{n} \text{ converge } \right\} = \left\{\sum_{n=k}^{\infty} \frac{\xi_n}{n} \text{ converge } \right\} \in \mathcal{A}_k^{\infty}$. Por lo tanto $A_1 \in \mathcal{A}^{\infty}$.
2. $A_2 = \{X_n \in I_n \text{ para infinitos } n\}$, donde $I_n \in \mathcal{B}(\mathbb{R})$
3. $A_3 = \{\limsup_n X_n < \infty\}$
4. $A_4 = \{\limsup_n \frac{X_1+\cdots+X_n}{n} < c\}$.

Ejemplos de eventos que no son eventos de cola son: $B_1 = \{X_n = 0 \text{ para todo } n\}$, $B_2 = \{\lim_n (X_1 + \ldots X_n)$existe y es menor que $c\}$. Observemos que si las variables aleatorias $X_1, X_2, \ldots$ son independientes, por el Lema de Borell Cantelli el suceso A_3 sólo puede tomar los valores 0 o 1. Esto es cierto en general no sólo para A_3 sino para cualquier evento de cola, como se demuestra en el siguiente teorema.

Teorema 1.28. *Sean $X_1, X_2, \ldots$ una sucesión de variables independientes. Sea $A \in \mathcal{A}^{\infty}$ entonces $P(A)$ sólo puede valer 0 o 1.*

Demostración. Sea $A \in \mathcal{A}^{\infty}$ observemos que por definición de $\mathcal{A}^{\infty}$, $A \in \mathcal{A}_1^{\infty} = \sigma\{X_1, X_2, \ldots\} = \sigma(\cup_n \mathcal{A}_1^n)$ siendo $\mathcal{A}_1^n = \sigma\{X_1, \ldots, X_n\}$. Se puede demostrar que es posible encontrar una sucesión de conjuntos $A_n \in \mathcal{A}_1^n$ que aproximen a A, es decir tal que $P(A \triangle A_n) \to 0$, siendo $\triangle$ la diferencia simétrica entre A y A_n. Por lo tanto

$$P(A_n) \to P(A) \quad P(A_n \cap A) \to P(A),$$

Por otra parte como $A \in \mathcal{A}^{\infty}$ se tiene que $A \in \mathcal{A}_{n+1}^{\infty}$ y por lo tanto A_n y A son independientes. Por lo tanto $P(A) = P^2(A)$ de donde $P(A)$ es 0 o 1. □

1.3.3. Desigualdad de Kolmogorov

Teorema 1.29. *Sean $X_1, X_2, \ldots, X_n$ variables independientes tal que $\mathbb{E}(X_i) = 0$ y $\mathbb{E}(X_i^2) < \infty$ para todo i. Entonces, si denotamos $S_k = X_1 + \cdots + X_k$, para todo $\varepsilon > 0$*

1.
$$P\left(\max_{1\leq k\leq n} |S_k| \geq \varepsilon\right) \leq \frac{\mathbb{E}(S_n^2)}{\varepsilon^2}, \tag{1.1}$$

2. si además existe $c > 0$ tal que $P(|X_i| \leq c) = 1$ para todo i entonces
$$P\left(\max_{1\leq k\leq n} |S_k| \geq \varepsilon\right) \geq 1 - \frac{(c+\varepsilon)^2}{\mathbb{E}(S_n^2)}. \tag{1.2}$$

Demostración. 1. Denotemos $A = \{\max_{1\leq k\leq n} |S_k| \geq \varepsilon\}$ y $A_k = \{|S_i| < \varepsilon, i = 1, \ldots, k-1, |S_k| \geq \varepsilon\}$ para $k = 1, \ldots, n$ donde $S_0 = 0$. Observemos que A es la unión disjunta de los A_k y por lo tanto

$$\mathbb{E}(S_n^2) \geq \mathbb{E}(S_n^2 \mathbb{I}_A) = \sum_{k=1}^{n} \mathbb{E}(S_n^2 \mathbb{I}_{A_k}).$$

Observemos que

$$\begin{aligned}\mathbb{E}(S_n^2 \mathbb{I}_{A_k}) =& \mathbb{E}\big[(S_k + (X_{k+1} + \cdots + X_n))^2 \mathbb{I}_{A_k}\big] \\ =& \mathbb{E}(S_k^2 \mathbb{I}_{A_k}) + 2\mathbb{E}[S_k(X_{k+1} + \cdots + X_n)\mathbb{I}_{A_k}] + \mathbb{E}[(X_{k+1} + \cdots + X_n)^2 \mathbb{I}_{A_k}].\end{aligned}$$

Como las variables $X_{k+1}, \ldots, X_n$ son centradas e independientes de la variable aleatoria $S_k \mathbb{I}_{A_k}$, se sigue que $\mathbb{E}(S_k(X_{k+1} + \cdots + X_n)\mathbb{I}_{A_k}) = 0$. Por lo tanto

$$\begin{aligned}\mathbb{E}(S_n^2 \mathbb{I}_{A_k}) =& \mathbb{E}(S_k^2 \mathbb{I}_{A_k}) + \mathbb{E}[(X_{k+1} + \cdots + X_n)^2 \mathbb{I}_{A_k}] \\ \geq& \mathbb{E}(S_k^2 \mathbb{I}_{A_k}).\end{aligned} \tag{1.3}$$

La última desigualdad se debe a que el segundo término es no negativo. Si $\omega \in A_k$ entonces por definición $|S_k(\omega)| \geq \varepsilon$, es decir $\mathbb{E}(S_k^2 \mathbb{I}_{A_k}) \geq \varepsilon^2 P(A_k)$. Finalmente

$$\mathbb{E}(S_n^2) \geq \sum_{k=1}^{n} \mathbb{E}(S_k^2 \mathbb{I}_{A_k}) \geq \varepsilon^2 \sum_{k=1}^{n} P(A_k) = \varepsilon^2 P(A) = \varepsilon^2 P\left(\max_{1\leq k\leq n} |S_k| \geq \varepsilon\right).$$

2. Primero acotamos por abajo $\mathbb{E}(S_n^2 \mathbb{I}_A)$: $\mathbb{E}(S_n^2 \mathbb{I}_A) = \mathbb{E}(S_n^2) - \mathbb{E}(S_n^2 \mathbb{I}_{A^c}) \geq \mathbb{E}(S_n^2) - \varepsilon^2 P(A^c)$, ya que si $\omega \in A^c$ entonces $S_n^2(\omega) < \varepsilon^2$. Por lo tanto

$$\mathbb{E}(S_n^2 \mathbb{I}_A) \geq \mathbb{E}(S_n^2) - \varepsilon^2 + \varepsilon^2 P(A). \tag{1.4}$$

Luego acotamos $\mathbb{E}(S_n^2 \mathbb{I}_A)$ por arriba: observemos que si $\omega \in A_k$ entonces $|S_{k-1}| \leq \varepsilon$ y por lo tanto

$$|S_k| = |S_{k-1} + X_k| \leq |S_{k-1}| + |X_k| \leq \varepsilon + c \tag{1.5}$$

Si ahora sumamos en k en la ecuación (1.3):

$$\mathbb{E}(S_n^2 \mathbb{I}_A) = \sum_{k=1}^{n} \mathbb{E}(S_k^2 \mathbb{I}_{A_k}) + \sum_{k=1}^{n} \mathbb{E}(\mathbb{I}_{A_k}(S_n - S_k)^2) = \sum_{k=1}^{n} \mathbb{E}(S_k^2 \mathbb{I}_{A_k}) + \sum_{k=1}^{n} P(A_k)\mathbb{E}(S_n - S_k)^2.$$

Donde en la última igualdad hemos usado que $\mathbb{I}_{A_k}$ y $S_n - S_k$ son independientes. Observemos que $\mathbb{E}(S_n - S_k)^2 = \mathbb{E}(X_n + \cdots + X_{k+1})^2 = \sum_{j=k+1}^{n} \mathbb{E}(X_j^2)$ ya que las variables X_j son independientes y centradas. Usando (1.5) podemos escribir

$$\begin{aligned}\sum_{k=1}^{n} \mathbb{E}(S_k^2 \mathbb{I}_{A_k}) + \sum_{k=1}^{n} P(A_k)\mathbb{E}((S_n - S_k)^2) \leq & (\varepsilon + c)^2 \sum_{k=1}^{n} P(A_k) + \sum_{k=1}^{n} P(A_k) \sum_{j=k+1}^{n} \mathbb{E}(X_j^2) \\ = & \sum_{k=1}^{n} P(A_k) \left[(\varepsilon + c)^2 + \sum_{j=k+1}^{n} \mathbb{E}(X_j^2)\right]\end{aligned}$$

Observemos que

$$\sum_{j=k+1}^{n} \mathbb{E}(X_j^2) \leq \sum_{j=1}^{n} \mathbb{E}(X_j^2)$$

por lo tanto

$$\begin{aligned}\sum_{k=1}^{n} P(A_k) \left[(\varepsilon + c)^2 + \sum_{j=k+1}^{n} \mathbb{E}(X_j^2)\right] \leq & \sum_{k=1}^{n} P(A_k) \left[(\varepsilon + c)^2 + \sum_{j=1}^{n} \mathbb{E}(X_j^2)\right] \\ = & P(A) \left[(\varepsilon + c)^2 + \sum_{j=1}^{n} \mathbb{E}(X_j^2)\right]\end{aligned}$$

Si usamos ahora que $\mathbb{E}(X_n) = 0$ para todo n, y que X_i es independiente de X_j para $i \neq j$

$$\mathbb{E}(S_n^2) = \sum_{j=1}^{n} \mathbb{E}(X_j^2)$$

es decir
$$P(A)\left[(\varepsilon+c)^2+\sum_{j=1}^{n}\mathbb{E}(X_j^2)\right]=P(A)[(\varepsilon+c)^2+\mathbb{E}(S_n^2)].$$
Usando (1.4) obtenemos
$$P(A)\geq\frac{\mathbb{E}(S_n^2)-\varepsilon^2}{(\varepsilon+c)^2+\mathbb{E}(S_n^2)-\varepsilon^2}.$$
Finalmente escribimos
$$\frac{\mathbb{E}(S_n^2)-\varepsilon^2}{(\varepsilon+c)^2+\mathbb{E}(S_n^2)-\varepsilon^2}=1-\frac{(\varepsilon+c)^2}{(\varepsilon+c)^2+\mathbb{E}(S_n^2)-\varepsilon^2}\geq 1-\frac{(\varepsilon+c)^2}{\mathbb{E}(S_n^2)},$$
donde en la última desigualdad hemos usado que $(\varepsilon+c)^2-\varepsilon^2>0$.

□

Corolario 1.30. *De la desigualdad* (1.1) *se sigue de manera inmediata la desigualdad de Markov:* $P(|X-\mathbb{E}(X)|>\varepsilon)\leq\frac{Var(X)}{\varepsilon^2}$.

1.4. Integral de Riemann-Stieltjes

Daremos un breve repaso, sin demostraciones, de la definición de integral de Riemann-Stieltjes así como alguna de sus propiedades.

1.4.1. Definición

Definición 1.31. Dadas $g,F:[a,b]\to\mathbb{R}$ y $P=\{a=c_0<c_1<c_2<\ldots<c_n=b\}$:

- definimos la suma parcial de Riemann-Stieltjes como
$$S(P,g,F)=\sum_{i=1}^{n}g(x_i)\big(F(c_i)-F(c_{i-1})\big),$$
siendo $x_i\in[c_{i-1},c_i]$.
- definimos la norma de la partición P como $\|P\|=\max\{c_i-c_{i-1}:i=1,\ldots,n\}$.
- decimos que $\lim_{\|P\|\to 0}S(P,g,F)=I$ si y sólo si para todo $\varepsilon>0$ existe $\delta>0$ tal que para toda partición P de $[a,b]$ con $\|P\|<\delta$ se tiene que $|S(P,g,F)-I|<\varepsilon$, en este caso decimos que la integral de Riemann-Stieltjes de g respecto de F en el intervalo $[a,b]$ existe y vale I, y denotamos
$$\int_a^b g(x)dF(x).$$

Observación 1.32.

- *Si* $F(x)=x$ *obtenemos la integral de Riemann.*
- *Si* $F(x)=k$, $\int_a^b gdF=0$ *para toda* g.
- *Si* $g:[a,b]\to\mathbb{R}$ *es continua y* $F(x)=\mathbb{I}_{[c,b]}(x)$ *con* $c\in(a,b)$ *entonces* $\int_a^b gdF=g(c)$.
- *Si* $g(x)=k$, *para toda* F $\int_a^b gdF=k(F(b)-F(a))$.

Teorema 1.33. *Los siguientes enunciados son equivalentes*

- *existe* $\lim_{\|P\|\to 0} S(P, g, F) = I < \infty$
- *dado* $\varepsilon > 0$ *existe* δ *tal que si* $\|P\| < \delta$ *y* $\|Q\| < \delta$ *se cumple* $|S(P, g, F) - S(Q, g, F)| < \varepsilon$
- *para toda sucesión de particiones* $\{P_n\}_{n\in\mathbb{N}}$ *de* $[a, b]$ *tal que* $\|P_n\| \to 0$ *se cumple que* $\lim_{n\to\infty} S(P_n, g, F) = I$.

Teorema 1.34. *Si* $g : [a, b] \to \mathbb{R}$ *es continua y* $F : [a, b] \to \mathbb{R}$ *es monótona, entonces existe* $\int_a^b gdF$.

Teorema 1.35. *Si* $g : [a, b] \to \mathbb{R}$ *es continua y* $F : [a, b] \to \mathbb{R}$ *es monótona y derivable tal que* $F'(x) = f(x)$ *siendo* f *integrable Riemann en* $[a, b]$, *entonces*

$$\int_a^b g(x)dF(x) = \int_a^b g(x)f(x)dx$$

Proposición 1.36. *Si* $g, h, F : [a, b] \to \mathbb{R}$ *son tales que existen* $\int_a^b gdF$ *y* $\int_a^b hdF$ *entonces también existe* $\int_a^b(\alpha g + \beta h)dF$ *para todo* $\alpha, \beta \in \mathbb{R}$ *y además*

$$\int_a^b (\alpha g + \beta h)dF = \alpha \int_a^b gdF + \beta \int_a^b hdF$$

Proposición 1.37. *Si* $h, F, G : [a, b] \to \mathbb{R}$ *son tales que existen* $\int_a^b hdF$ *y* $\int_a^b hdG$ *entonces también existe* $\int_a^b hd(\alpha F + \beta G)$ *para todo* $\alpha, \beta \in \mathbb{R}$ *y además*

$$\int_a^b hd(\alpha F + \beta G) = \alpha \int_a^b hdF + \beta \int_a^b dG$$

Proposición 1.38. *Si* $g, F : [a, b] \to \mathbb{R}$ *son tales que existe* $\int_a^b gdF$ *entonces para todo* $c \in (a, b)$ *existe* $\int_a^c gdF$ *y* $\int_c^b gdF$ *y vale*

$$\int_a^b gdF = \int_a^c gdF + \int_c^b gdF$$

Proposición 1.39. *Si* $g, F : [a, b] \to \mathbb{R}$ *son tal que* $\alpha \leq g \leq \beta$ *para todo* $x \in [a, b]$, *y* F *es monótona no decreciente tal existe* $\int_a^b gdF$ *entonces*

$$\alpha(F(b) - F(a)) \leq \int_a^b gdF \leq \beta(F(b) - F(a))$$

Proposición 1.40. *Si* $g : [a, b] \to \mathbb{R}$ *es continua y* $F : [a, b] \to \mathbb{R}$ *es monótona no decreciente, entonces*

$$\left| \int_a^b g(x)dF(x) \leq \int_a^b |g(x)|dF(x) \right|$$

Proposición 1.41. ***(Teorema del valor medio.)*** *Si* $g, F : [a, b] \to \mathbb{R}$ *son tales que* g *es continua,* F *es monótona creciente, entonces existe* $c \in [a, b]$ *tal que* $\int_a^b gdF = g(c)(F(b) - F(a))$.

1.4.2. Métodos de integración

Teorema 1.42. ***(Fórmula de integración por partes.)*** *Si $g, F : [a, b] \to \mathbb{R}$ son tales que existe $\int_a^b gdF$, entonces existe existe $\int_a^b Fdg$ y además*

$$\int_a^b Fdg = gF\big|_a^b - \int_a^b gdF$$

Teorema 1.43. ***(Cambio de variable.)*** *Si $g, F : [a, b] \to \mathbb{R}$ son tales que $\int_a^b gdF$ existe, $h : [c, d] \to [a, b]$ es continua y biyectiva, entonces existe $\int_a^b g \circ hd(F \circ h)$ y además*

$$\int_c^d g(h(t))dF(h(t)) = \int_a^b g(x)dF(x) \tag{1.6}$$

1.4.3. Aplicaciones a la teoría de la probabilidad

Proposición 1.44. *Si F_X es la función de distribución de una variable aleatoria X, entonces*

$$\int_a^b dF_X(x) = P(a < X \leq b)$$

Proposición 1.45. *Sea X una variable aleatoria cuya función de distribución es F_X,*

- *Si su recorrido es $A = \{x_1, x_2, \dots\}$ y $g : [a, b] \to \mathbb{R}$ es continua, entonces*

$$\int_a^b g(x)dF_X(x) = \sum_{x \in (a,b] \cap A} g(x)p_X(x) = \mathbb{E}(g(X)\mathbb{I}_{X \in (a,b]})$$

- *Si X es absolutamente continua con densidad f_X y $g[a, b] \to \mathbb{R}$ es continua, entonces*

$$\int_a^b g(x)dF_X(x) = \int_a^b g(x)f_X(x)dx = \mathbb{E}(g(X)\mathbb{I}_{\{X \in [a,b]\}})$$

Observemos que, de la proposición anterior se sigue que podemos definir la esperanza de una variable aleatoria X con distribución F_X, como

$$\mathbb{E}(X) = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b xdF_X(x) = \int_{-\infty}^{+\infty} xdF_X(x)$$

siempre que dicho límite doble exista, (la segunda igualdad es simplemente notación, para el caso en que dicho límite existe).

Ejercicio 1.46. Probar que si F es una función de distribución, $a < b$ son puntos de continuidad de F, definimos

$$\psi(x) = \begin{cases} 0, & x < a, x > b \\ 1/2, & x = a, x = b \\ 1, & a < x < b \end{cases},$$

entonces $\int_{-\infty}^{+\infty} \psi(x)dF(x) = F(b) - F(a)$

Capítulo 2

Convergencia de variables aleatorias

2.1. Convergencia casi segura, sucesiones de Cauchy, convergencia de series

Definición 2.1. Sean X y $\{X_n\}_{n\in\mathbb{N}}$ variables aleatorias definidas en el mismo espacio $(\Omega, \mathcal{A}, P)$, decimos que X_n converge casi seguramente a X si y sólo si

$$P\big(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}\big) = 1,$$

denotamos $X_n \stackrel{c.s.}{\to} X$.

Observación 2.2.

- *por el Corolario 1.25, si $X_n \stackrel{c.s.}{\to} X$, X es una variable aleatoria.*
- *Si $g : \mathbb{R}^d \to \mathbb{R}$ es continua y $X_n^i \stackrel{c.s.}{\to} X^i$ para todo $1 \leq i \leq d$ entonces $g(X_n^1, \ldots, X_n^d) \stackrel{c.s.}{\to} g(X^1, \ldots, X^d)$. En particular tomando $d = 2$ obtenemos que si $X_n \stackrel{c.s.}{\to} X$ y $Y_n \stackrel{c.s.}{\to} Y$ entonces para todo a, b $aX_n + bY_n \stackrel{c.s.}{\to} aX + bY$ y $X_nY_n \stackrel{c.s.}{\to} XY$.*
- *Si $X_n \stackrel{c.s.}{\to} 0$ y existe $k > 0$ tal que $P(|Y_n| > k) = 0$ para todo n entonces $X_nY_n \stackrel{c.s.}{\to} 0$.*

Teorema 2.3. *$X_n \stackrel{c.s.}{\to} X$ si y sólo si $\lim_{k\to+\infty} P(\cap_{n=k}^{+\infty}\{|X_n - X| < \varepsilon\}) = 1$ para todo $\varepsilon > 0$.*

Demostración. Observemos que $P\left(\lim_{n\to+\infty} X_n = X\right) = 1$ si y sólo si

$$P\left(\bigcap_{\varepsilon\in\mathbb{Q}^+} \bigcup_{k=1}^{+\infty} \bigcap_{n=k}^{+\infty} \Big\{|X_n - X| < \varepsilon\Big\}\right) = 1 \Leftrightarrow P\left(\bigcup_{k=1}^{+\infty} \bigcap_{n=k}^{+\infty} \Big\{|X_n - X| < \varepsilon\Big\}\right) = 1, \text{ para todo } \varepsilon \in \mathbb{Q}^+.$$

Como los conjuntos $B_k = \cap_{n=k}^{+\infty}\{|X_n - X| < \varepsilon\}$ forman una sucesión creciente (es decir $B_k \subset B_{k+1}$) tenemos que $P(\cup_{k=1}^{+\infty} B_k) = \lim_{k\to+\infty} P(B_k)$. □

Corolario 2.4. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables tal que para todo $\varepsilon > 0$, $\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty$, entonces $X_n \to X$ c.s.*

Demostración. Por el teorema anterior basta probar que $\lim_{k\to+\infty} P(\cup_{n=k}^{+\infty}\{|X_n - X| > \varepsilon\}) = 0$,

$$P(\cup_{n=k}^{+\infty}\{|X_n - X| > \varepsilon\}) \leq \sum_{n=k}^{\infty} P(|X_n - X| > \varepsilon)$$

y como la serie es convergente

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

□

Teorema 2.5. *$X_n \overset{c.s.}{\to} X$ si y sólo si $\forall\varepsilon > 0$, $\lim_{k\to+\infty} P(\sup_{n\geq k}|X_n - X| > \varepsilon) = 0$.*

Demostración. Basta observar que

$$\left\{\bigcup_{n=k}^{+\infty} |X_n - X| > \varepsilon\right\} = \left\{\sup_{n\geq k}|X_n - X| > \varepsilon\right\},$$

por lo tanto

$$P\left(\sup_{n\geq k}|X_n - X| > \varepsilon\right) \to 0 \Leftrightarrow P\left(\bigcup_{n=k}^{+\infty}|X_n - X| > \varepsilon\right) \to 0,$$

y por el Teorema 2.3 esto se da si y sólo si $X_n \overset{c.s.}{\to} X$. □

Vamos a introducir ahora un concepto análogo al de sucesión de Cauchy, recordemos que una sucesión $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ es de Cauchy si $|x_n - x_m| \to 0$ cuando $n \to \infty$ y $m \to \infty$.

Definición 2.6. Sean X y $\{X_n\}_{n\in\mathbb{N}}$ variables aleatorias definidas en el mismo espacio $(\Omega, \mathcal{A}, P)$, a valores en $\mathbb{R}^d$, decimos que X_n es de Cauchy con probabilidad 1 si y sólo si

$$P\big(\{\omega : X_n(\omega) \text{ es de Cauchy}\}\big) = 1.$$

Teorema 2.7. *La sucesión $\{X_n\}_{n\in\mathbb{N}}$ es de Cauchy con probabilidad 1 si y sólo si*

1. *para todo $\varepsilon > 0$*

$$P\left(\sup_{\substack{k\geq n\\ l\geq n}}|X_k - X_l| > \varepsilon\right) \to 0 \quad \textit{cuando } n \to \infty, \tag{2.1}$$

o de forma equivalente:

2. *para todo $\varepsilon > 0$*

$$P\left(\sup_{k\geq 0}|X_{n+k} - X_n| > \varepsilon\right) \to 0 \quad \textit{cuando } n \to \infty. \tag{2.2}$$

Demostración. La demostración de que $\{X_n\}_{n\in\mathbb{N}}$ es de Cauchy en probabilidad si y sólo si (2.1) es análoga a la prueba del Teorema 2.3. En este caso, fijado $\varepsilon > 0$ tomamos los conjuntos

$$B^{\varepsilon}_{k,l} = \big\{\omega : |X_k - X_l| > \varepsilon\big\} \quad y \quad B^{\varepsilon} = \bigcap_{n=1}^{\infty}\bigcup_{\substack{k\geq n\\ l\geq n}} B^{\varepsilon}_{k,l}$$

Observemos que si $X_n(\omega)$ no es de Cauchy existe $\varepsilon > 0$ (el cual podemos suponer sin perdida de generalidad que es un número racional), tal para infinitos valores de l y k se cumple que que $|X_l(\omega) - X_k(\omega)| > \epsilon$ es decir:

$$\{\omega : X_n(\omega) \text{ no es de Cauchy}\} = \bigcup_{\varepsilon\in\mathbb{Q}} B^{\varepsilon}.$$

Si acotamos la probabilidad de la unión por la suma:

$$P(\{\omega : X_n(\omega) \text{ no es de Cauchy}\}) \leq \sum_{\varepsilon \in \mathbb{Q}} P(B^\varepsilon).$$

Por lo tanto basta probar que $P(B^\varepsilon) = 0$ si y sólo si se cumple (2.1). Denotemos

$$B_n^\varepsilon = \bigcup_{\substack{k \geq n \\ l \geq n}} B_{k,l}^\varepsilon$$

Observemos que, para todo $\varepsilon > 0$, B_n^ε es una sucesión decreciente de conjuntos, por lo tanto, para todo $\varepsilon > 0$,

$$P(B^\varepsilon) = \lim_{n \to +\infty} P\left(\bigcup_{\substack{k \geq n \\ l \geq n}} B_{k,l}^\varepsilon \right), \tag{2.3}$$

y

$$\bigcup_{\substack{k \geq n \\ l \geq n}} B_{k,l}^\varepsilon = \sup_{\substack{k \geq n \\ l \geq n}} |X_k - X_l| > \varepsilon, \text{ implica } P\left(\bigcup_{\substack{k \geq n \\ l \geq n}} B_{k,l}^\varepsilon \right) = P\left(\sup_{\substack{k \geq n \\ l \geq n}} |X_k - X_l| > \varepsilon \right).$$

Si usamos (3.5) obtenemos que

$$P(B^\varepsilon) = \lim_{n \to +\infty} P\left(\sup_{\substack{k \geq n \\ l \geq n}} |X_k - X_l| > \varepsilon \right). \tag{2.4}$$

La equivalencia entre (2.1) y (2.2) se sigue de

$$\sup_{k \geq 0} |X_{n+k} - X_n| \leq \sup_{\substack{k \geq 0 \\ l \geq 0}} |X_{n+k} - X_{n+l}| \leq \sup_{k \geq 0} |X_{n+k} - X_n| + \sup_{l \geq 0} |X_n - X_{n+l}| = 2 \sup_{k \geq 0} |X_{n+k} - X_n|$$

donde la primera desigualdad es porque estamos tomando supremo en un conjunto mas grande. De la primera desigualdad obtenemos que

$$\left\{\omega : \sup_{k \geq 0} |X_{n+k}(\omega) - X_n(\omega)| > \varepsilon\right\} \subset \left\{\omega : \sup_{\substack{k \geq 0 \\ l \geq 0}} |X_{n+k}(\omega) - X_{n+l}(\omega)| > \varepsilon\right\}$$

y por lo tanto

$$P\left\{\omega : \sup_{k \geq 0} |X_{n+k}(\omega) - X_n(\omega)| > \varepsilon\right\} \leq P\left\{\omega : \sup_{\substack{k \geq 0 \\ l \geq 0}} |X_{n+k}(\omega) - X_{n+l}(\omega)| > \varepsilon\right\}$$

de donde se sigue que (2.1) implica (2.2). Para ver que (2.2) implica (2.1), observemos que de

$$\sup_{\substack{k \geq 0 \\ l \geq 0}} |X_{n+k} - X_{n+l}| \leq 2 \sup_{k \geq 0} |X_{n+k} - X_n|$$

se sigue

$$P\left\{\omega : \sup_{\substack{k \geq 0 \\ l \geq 0}} |X_{n+k}(\omega) - X_{n+l}(\omega)| > \varepsilon\right\} \leq P\left\{\omega : \sup_{k \geq 0} |X_{n+k}(\omega) - X_n(\omega)| > \varepsilon/2\right\}$$

□

Teorema 2.8. *Un sucesión $\{X_n\}_{n\in\mathbb{N}}$ es de Cauchy con probabilidad 1 si y sólo si existe X tal que $X_n \overset{c.s.}{\to} X$.*

Demostración. El recíproco se sigue de que:

$$\sup_{\substack{k\geq n\\ l\geq n}} |X_l - X_k| \leq \sup_{k\geq n} |X_k - X| + \sup_{l\geq n} |X - X_l|,$$

esto implica que

$$\Big\{\omega : \sup_{\substack{k\geq n\\ l\geq n}} |X_l(\omega) - X_k(\omega)| > \varepsilon\Big\} \subset \Big\{\omega : \sup_{k\geq n} |X_k(\omega) - X(\omega)| > \varepsilon/2\Big\} \bigcup \Big\{\omega : \sup_{l\geq n} |X(\omega) - X_l(\omega)| > \varepsilon/2\Big\}$$

es decir

$$P\left(\sup_{\substack{k\geq n\\ l\geq n}} |X_l - X_k| > \varepsilon\right) \leq P\left(\sup_{k\geq n} |X_k - X| > \frac{\varepsilon}{2}\right) + P\left(\sup_{l\geq n} |X - X_l| > \frac{\varepsilon}{2}\right),$$

si $X_n \overset{c.s.}{\to} X$, por el Teorema 2.5, cada uno de los dos sumandos tiende a cero, y por el Teorema 2.7 la sucesión es de Cauchy con probabilidad 1. Para ver el directo consideremos el conjunto $N = \{\omega : X_n(\omega) \text{ es de Cauchy}\}$, dicho conjunto tiene probabilidad 1, además, si $\omega \in N$ entonces $\{X_n(\omega)\}_n$ es una sucesión de Cauchy en $\mathbb{R}$ y por lo tanto converge. Definimos $X(\omega) = \text{lím}_n X_n(\omega)$ si $\omega \in N$ y cero en otro caso. □

2.1.1. Teoremas de convergencia monótona y dominada

Teorema 2.9. ***(Teorema de convergencia monótona.)*** *Sean X y $\{X_n\}_{n\in\mathbb{N}}$ variables aleatorias definidas en el mismo espacio $(\Omega, \mathcal{A}, P)$ tal que existe $\mathbb{E}(X)$, $X_n(\omega) \geq 0$ y $X_n(\omega) \uparrow X(\omega)$ para casi todo $\omega \in \Omega$ (es decir existe $D \subset \Omega$ con $P(D) = 1$ tal que $X_n(\omega) \uparrow X(\omega)$ para todo $\omega \in D$) entonces existe $\mathbb{E}(X_n)$ para todo n y además*

$$\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X). \tag{2.5}$$

Demostración. Observemos que como $0 \leq X_n \leq X$ existe $\mathbb{E}(X_n)$ y además es una sucesión creciente y acotada por $\mathbb{E}(X)$, por lo tanto existe $\text{lím}_n \mathbb{E}(X_n) \leq \mathbb{E}(X)$. Para probar (2.5) basta probar que para todo $\varepsilon > 0$ $\text{lím}_n \mathbb{E}(X_n) \geq \mathbb{E}(X) - \varepsilon$. Fijado $\varepsilon > 0$ consideremos las variables aleatorias

$$Y_\varepsilon = \sum_{n=0}^{+\infty} n\varepsilon \mathbb{I}_{\{n\varepsilon < X \leq (n+1)\varepsilon\}},$$

es claro que $X - \varepsilon \leq Y_\varepsilon < X$ y por lo tanto $\mathbb{E}(X) - \varepsilon \leq \mathbb{E}(Y_\varepsilon) < \mathbb{E}(X)$. Veamos que $\text{lím}_n \mathbb{E}(X_n) \geq \mathbb{E}(Y_\varepsilon)$. En lo que resta los ω que consideremos serán tomados en el conjunto de probabilidad 1 tal que $X_n(\omega) \uparrow X(\omega)$. Consideremos los sucesos $A_k = \{X_k \geq Y_\varepsilon\}$, como X_k es creciente los A_k son una sucesión creciente. Como $X_k(\omega) \to X(\omega) > Y_\varepsilon(\omega)$ tenemos que $\cup_{k=1}^{+\infty} A_k = \Omega$. Por lo tanto $A_k \cap B_n \uparrow_k B_n$ siendo $B_n = \{n\varepsilon < X \leq (n+1)\varepsilon\}$, por lo tanto

$$\mathbb{E}(Y_\varepsilon \mathbb{I}_{A_k}) = \sum_{n=0}^{+\infty} n\varepsilon P(Y_\varepsilon \mathbb{I}_{A_k} = n\varepsilon) = \sum_{n=0}^{+\infty} n\varepsilon P(A_k \cap B_n) \geq \sum_{n=0}^{m} n\varepsilon P(A_k \cap B_n),$$

para todo m, como $A_k \cap B_n \uparrow_k B_n$, $P(A_k \cap B_n) \to_k P(B_n)$, es decir

$$\lim_k \mathbb{E}(Y_\varepsilon \mathbb{I}_{A_k}) \geq \sum_{n=0}^{m} n\varepsilon P(B_n)$$

tomando límite en m,

$$\lim_{k} \mathbb{E}(Y_\varepsilon \mathbb{I}_{A_k}) \geq \sum_{n=0}^{\infty} n\varepsilon P(B_n) = \mathbb{E}(Y_\varepsilon).$$

Como $Y_\varepsilon \mathbb{I}_{A_k} \leq X_k$ se tiene que $\mathbb{E}(Y_\varepsilon \mathbb{I}_{A_k}) \leq \mathbb{E}(X_k)$. □

Observación 2.10. *Se deja como ejercicio observar que, siguiendo las mismas ideas que antes, el teorema anterior sigue siendo válido si estamos en el caso en que* $\mathbb{E}(X) = \infty$, *no obstante lo único que se concluye es que* $\mathbb{E}(X_n) \to \infty$.

Teorema 2.11. ***(Teorema de convergencia dominada.)*** *Sean* X, Y *y* $\{X_n\}_{n\in\mathbb{N}}$ *variables aleatorias definidas en el mismo espacio* $(\Omega, \mathcal{A}, P)$ *tal que* $X_n(\omega) \xrightarrow{c.s} X(\omega)$ *y* $|X_n(\omega)| \leq Y$ *para casi todo* $\omega \in \Omega$, *supongamos que existe* $\mathbb{E}(Y)$ *entonces existen las esperanzas de las* X_n *para todo* n *y la de* X *y además*

$$\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X).$$

Demostración. Que existen $\mathbb{E}(X_n)$ y $\mathbb{E}(X)$ se sigue de que $|X_n| \leq Y$ para todo n y por lo tanto $|X| \leq Y$. Definimos $Y_n = \inf_{k\geq n} X_k$, es claro que $Y_n \uparrow X$ y por lo tanto $Y_n + Y \uparrow X + Y$. Además $0 \leq Y_n + Y$ ya que $|X_n| \leq Y$, por lo tanto por el teorema de convergencia monótona $\lim_n \mathbb{E}(Y_n + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$, de donde $\mathbb{E}(Y_n) \to \mathbb{E}(X)$. Análogamente, si tomamos $Z_n = \sup_{k\geq n} X_k$ tenemos que $Z_n \downarrow X$ y como $0 \leq Y - Z_n \uparrow Y - X$ obtenemos, aplicando nuevamente el teorema de convergencia monótona, que $\mathbb{E}(Z_n) \to \mathbb{E}(X)$. La tesis se sigue de que $Y_n \leq X_n \leq Z_n$ y por lo tanto $\mathbb{E}(Y_n) \leq \mathbb{E}(X_n) \leq \mathbb{E}(Z_n)$. □

Corolario 2.12. *Si* $f_n : [a, b] \to \mathbb{R}$ *y* $g : [a, b] \to \mathbb{R}$ *son Riemann integrables tal que* $\lim_n f_n(x) = f(x)$ *y* $|f_n(x)| \leq g(x)$ *para todo* $x \in [a, b]$ *entonces*

$$\lim_{n} \int_a^b f_n(x)dx = \int_a^b f(x)dx.$$

El siguiente Teorema, permite "derivar bajo el signo de la integral". Lo enunciamos ya que será usado mas adelante, una demostración del mismo puede encontrarse en el [2].

Teorema 2.13. *Sea* $f : \mathbb{R} \times [a, b] \to \mathbb{R}$ *y* $F : \mathbb{R} \to \mathbb{R}$ *una función monótona no decreciente, con* $-\infty < a < b < +\infty$, *tal que para todo* $t \in [a, b]$ *la función* $f(\cdot, t) : \mathbb{R} \to \mathbb{R}$ *es integrable respecto de* $dF(x)$ *(es decir existe y es finita* $\int_{\mathbb{R}} f(t, x)dF(x)$*), denotemos*

$$G(t) = \int_{\mathbb{R}} f(x, t)dF(x),$$

supongamos que $\frac{\partial f}{\partial t}(x, t)$ *existe para todo* x *y además existe* g *tal que* $\int_{\mathbb{R}} g(x)F(dx) < \infty$ *y* $\left|\frac{\partial f}{\partial t}(x, t)\right| < g(x)$ *para todo* t, x, *entonces* G *es derivable y vale*

$$G'(t) = \int_{\mathbb{R}} \frac{\partial f}{\partial t}(x, t)dF(x).$$

2.1.2. Convergencia casi segura de series

Teorema 2.14. ***(Kolmogorov-Khinchin.)*** *Sea* $\{X_n\}_{n\in\mathbb{N}}$ *una sucesión de variables aleatorias independientes tal que* $\mathbb{E}(X_n) = 0$ *para* $n \geq 1$, *si*

$$\sum_{n=1}^{\infty} \mathbb{E}(X_n^2) < \infty,$$

entonces, con probabilidad 1

$$\sum_{n=1}^{\infty} X_n < \infty.$$

Si las variables X_k son uniformemente acotadas, es decir existe $c < \infty$ tal que $P(|X_k| < c) = 1$ para todo k, el recíproco es cierto: $\sum_{n=1}^{\infty} X_n < \infty$ implica que $\sum_{n=1}^{\infty} \mathbb{E}(X_n^2) < \infty$.

Demostración. Por el Teorema 2.8 basta probar que $\sum_{n=1}^{\infty} X_n$ es de Cauchy con probabilidad 1. Denotemos $S_n = \sum_{j=1}^{n} X_j$, por el Teorema 2.7 2., esto es equivalente a demostrar que, para todo $\varepsilon > 0$,

$$P\left(\sup_{k\geq 1}\left|S_{n+k} - S_n\right| > \varepsilon\right) \to 0 \quad \text{cuando } n \to \infty,$$

o lo que es lo mismo, si escribimos $S_{n+k} - S_n = \sum_{j=1}^{n+k} X_j - \sum_{j=1}^{n} X_j$,

$$P\left(\sup_{k\geq 1}\left|\sum_{j=1}^{n+k} X_j - \sum_{j=1}^{n} X_j\right| > \varepsilon\right) \to 0 \quad \text{cuando } n \to \infty.$$

Observemos que, para todo $n > 0$,

$$\left\{\omega \in \Omega : \max_{1\leq k\leq N}\left|\sum_{j=1}^{n+k} X_j - \sum_{j=1}^{n} X_j\right| > \varepsilon\right\} \uparrow \left\{\omega \in \Omega : \sup_{k\geq 1}\left|\sum_{j=1}^{n+k} X_j - \sum_{j=1}^{n} X_j\right| > \varepsilon\right\} \quad \text{cuando } N \to \infty,$$

por lo tanto,

$$P\left(\sup_{k\geq 1}\left|\sum_{j=1}^{n+k} X_j - \sum_{j=1}^{n} X_j\right| > \varepsilon\right) = \lim_{N\to\infty} P\left(\max_{1\leq k\leq N} |S_{n+k} - S_n| > \varepsilon\right),$$

para acotar la probabilidad del máximo usamos la desigualdad de Kolmogorov (1.1)

$$\lim_{N\to\infty} P\left(\max_{1\leq k\leq N} |S_{n+k} - S_n| > \varepsilon\right) \leq \lim_{N\to\infty} \frac{\sum_{k=n+1}^{n+N} \mathbb{E}(X_k^2)}{\varepsilon^2} = \frac{\sum_{k=n+1}^{\infty} \mathbb{E}(X_k^2)}{\varepsilon^2} \to_n 0.$$

donde el último límite se debe a que, como estamos suponiendo que $\sum_{n=1}^{\infty} \mathbb{E}(X_n^2) < \infty$, se tiene que $\lim_{n\to\infty} \sum_{k=n+1}^{\infty} \mathbb{E}(X_k^2) = 0$.

Vamos a demostrar el recíproco por absurdo, es decir suponemos que $\sum X_k$ converge, y $\sum \mathbb{E}(X_k^2) = \infty$. Como $\sum X_k$ converge, por el Teorema 2.7 podemos tomar n suficientemente grande tal que

$$P\left(\sup_{k\geq 1} |S_{n+k} - S_n| > \varepsilon\right) < 1/2, \tag{2.6}$$

pero, por otra parte, si usamos (1.2) obtenemos que, para todo N

$$P\left(\max_{1\leq k\leq N} |S_{n+k} - S_n| > \varepsilon\right) \geq 1 - \frac{(\varepsilon + c)^2}{\sum_{k=n+1}^{N+n} \mathbb{E}(X_k^2)},$$

como estamos suponiendo que $\sum \mathbb{E}(X_k^2) = \infty$ podemos tomar N tal que

$$P\left(\max_{1\leq k\leq N} |S_{n+k} - S_n| > \varepsilon\right) > 1/2$$

lo que contradice (2.6). $\square$

2.2. Convergencia en Probabilidad

Definición 2.15. Sean X y $\{X_n\}_{n\in\mathbb{N}}$ variables aleatorias definidas en el mismo espacio $(\Omega, \mathcal{A}, P)$, decimos que X_n converge en probabilidad a X si y sólo si para todo $\varepsilon > 0$

$$P\Big(\big\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\big\}\Big) \to 0 \quad \text{cuando } n \to \infty$$

denotamos $X_n \xrightarrow{P} X$.

Proposición 2.16. *Si $X_n \xrightarrow{P} X$ y $X_n \xrightarrow{P} Y$ entonces $P(X = Y) = 1$*

Demostración. Sea $k > 0$, denotemos $\Omega_k = \{\omega : |X(\omega) - Y(\omega)| < 1/k\}$, observemos que

$$\Omega_k^c \subset \big\{|X_n - X| \geq 1/(2k)\big\} \cup \big\{|X_n - Y| \geq 1/(2k)\big\},$$

por lo tanto

$$P(\Omega_k^c) \leq P\big(|X_n - Y| \geq 1/(2k)\big) + P\big(|X_n - X| \geq 1/(2k)\big) \to_n 0,$$

de donde $P(\Omega_k) = 1$ para todo k y entonces $P(\cap_k \Omega_k) = 1$. □

Teorema 2.17. *Sean X y $\{X_n\}_{n\in\mathbb{N}}$ variables aleatorias definidas en el mismo espacio $(\Omega, \mathcal{A}, P)$, si $X_n \xrightarrow{c.s.} X$ entonces $X_n \xrightarrow{P} X$.*

Demostración. Observemos que, por el Teorema 2.3 sabemos que, para todo $\varepsilon > 0$,

$$\lim_{k\to+\infty} P\left(\bigcup_{n=k}^{+\infty} \big\{|X_n - X| > \varepsilon\big\}\right) = 0,$$

la convergencia en probabilidad se sigue de que $\{|X_k - X| > \varepsilon\} \subset \cup_{n=k}^{+\infty}\{|X_n - X| > \varepsilon\}$. □

Ejemplo 2.18. Veamos con un ejemplo que el recíproco no es cierto. Consideremos la sucesión de intervalos $I_n^i = [\frac{i-1}{n}, \frac{i}{n}]$ y las variables $X_n^i = \mathbb{I}_{I_n^i}$, con $i = 1, 2, \ldots, n$; $n \geq 1$. Observemos que en este caso el espacio de probabilidad Ω es el intervalo $[0, 1]$, la σ-álgebra es la de Borel en $[0, 1]$, y la probabilidad es la que asigna a un intervalo $[a, b]$ con $0 \leq a \leq b \leq 1$ el valor $b - a$. La sucesión $\{X_1^1, X_2^1, X_2^2, X_3^1, X_3^2, X_3^3, \ldots\}$ converge en probabilidad a 0, pero no converge c.s.

Proposición 2.19. *Si $g : \mathbb{R}^d \to \mathbb{R}$ es continua y $X_n^i \xrightarrow{P} X^i$ para todo $1 \leq i \leq d$ entonces $g(X_n^1, \ldots, X_n^d) \xrightarrow{P} g(X^1, \ldots, X^d)$. En particular tomando $d = 2$ obtenemos que si $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} Y$ entonces para todo a, b $aX_n + bY_n \xrightarrow{P} aX + bY$ y $X_nY_n \xrightarrow{P} XY$.*

Proposición 2.20. *Si $X_n \xrightarrow{P} 0$ y existe $k > 0$ tal que $P(|Y_n| > k) = 0$ para todo n entonces $X_nY_n \xrightarrow{P} 0$.*

Teorema 2.21. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias. Entonces $X_n \xrightarrow{P} X$ si y sólo si para toda subsucesión $\{X_{n_k}\}_{k\in\mathbb{N}}$ existe $\{X_{n_{k_j}}\}_{j\in\mathbb{N}}$ una subsucesión de $\{X_{n_k}\}_{k\in\mathbb{N}}$ tal que $X_{n_{k_j}} \xrightarrow{c.s.} X$ cuando $j \to \infty$.*

Demostración. Para demostrar el directo tomemos una subsucesión $\{X_{n_k}\}_{k\in\mathbb{N}}$ de $\{X_n\}_{n\in\mathbb{N}}$, es claro que $X_{n_k} \xrightarrow{P} X$ cuando $k \to \infty$, por lo tanto para simplificar la notación podemos suponer $X_{n_k} = X_n$. Observemos que para todo $j > 0$, $P(|X_n - X| > 1/2^j) \to 0$, cuando $n \to \infty$. Definimos n_j para $j = 1, 2, \ldots$ tal que $P(|X_{n_j} - X| > 1/2^j) < 1/2^j$. Veamos que $X_{n_j} \xrightarrow{c.s} X$ cuando $j \to \infty$. Para eso observemos que por el

Teorema 2.3 basta probar que para todo $\varepsilon > 0$, $P(\cup_{j\geq m}|X_{n_j} - X| > \varepsilon) \to 0$ cuando $m \to \infty$, para eso tomemos m_0 tal que $1/2^{m_0} < \varepsilon$ entonces

$$P\left(\bigcup_{j\geq m_0} |X_{n_j} - X| > \varepsilon\right) \leq P\left(\bigcup_{j\geq m_0} |X_{n_j} - X| > 1/2^j\right) \leq \sum_{j\geq m_0} 1/2^j \to 0 \quad \text{cuando } m_0 \to \infty.$$

Para demostrar el recíproco supongamos por absurdo que existe $\varepsilon > 0$ tal que $P(|X_n - X| > \varepsilon)$ no converge a 0. Esto significa que existe $\delta > 0$ y n_k un conjunto de indices con $n_k \to \infty$ cuando $k \to \infty$ tal que $P(|X_{n_k} - X| > \varepsilon) > \delta$ para todo k. Por hipótesis existe $\{X_{n_{k_j}}\}_{j\in\mathbb{N}}$ una subsucesión de $\{X_{n_k}\}_{k\in\mathbb{N}}$ que converge c.s. a X. Por el Teorema 2.17 $X_{n_{k_j}} \xrightarrow{P} X$ cuando $j \to \infty$ es decir $P(|X_{n_{k_j}} - X| > \varepsilon) \to_j 0$ lo cual contradice que $P(|X_{n_k} - X| > \varepsilon) > \delta$ para todo k, ya que $\{X_{n_{k_j}}\}_j$ es una subsucesión de $\{X_{n_k}\}_k$. □

Se deja como ejercicio el siguiente teorema:

Teorema 2.22. *$X_n \xrightarrow{P} X$ si y sólo si $\mathbb{E}\left(\frac{|X_n-X|}{1+|X_n-X|}\right) \to 0$.*

2.3. Ley débil de los grandes números.

Teorema 2.23. *Sea $\{X_n\}_n$ una sucesión de variables aleatorias con esperanza finita. Supongamos que*

$$\frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \to_n 0, \tag{2.7}$$

entonces

$$\frac{1}{n}\left[\sum_{i=1}^{n} \big(X_i - \mathbb{E}(X_i)\big)\right] \xrightarrow{P} 0, \tag{2.8}$$

en particular si las X_i son independientes dos a dos y existe C tal que $Var(X_i) < C$ para todo i, se cumple (2.8)

Demostración. Denotemos $Z_n = \frac{1}{n}\sum_{i=1}^n X_i$, (2.8) es equivalente a probar que para todo $\varepsilon > 0$, $P(|Z_n - \mathbb{E}(Z_n)| > \varepsilon) \to_n 0$, pero esto se sigue de la desigualdad de Markov ya que

$$P\big(|Z_n - \mathbb{E}(Z_n)| > \varepsilon\big) \leq \frac{1}{n^2\varepsilon^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \to_n 0.$$

□

Veamos ahora otra versión del Teorema anterior en el el cual pedimos una condición más fuerte que (2.7) pero no tan restrictiva como la independencia. Para eso vamos a definir dos conceptos importantes.

Definición 2.24. Dadas X, Y variables aleatorias definidas en un mismo espacio de probabilidad tal que $0 < \mathrm{Var}(X) < \infty$ y $0 < \mathrm{Var}(Y) < \infty$, el coeficiente de correlación entre X e Y se define como

$$\rho(X,Y) = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}.$$

Observación 2.25. *Si X e Y son independientes, $\rho(X,Y) = 0$, no obstante el recíproco no es cierto en general. Se puede ver que $-1 \leq \rho(X,Y) \leq 1$ y que $|\rho(X,Y)| = 1$ si y sólo si $X = aY + b$ siendo a y b constantes.*

Definición 2.26. Decimos que una sucesión de variables aleatorias $\{X_n\}$ es débilmente estacionaria si la esperanza de cada variable es constante ($\mathbb{E}(X_n) = c < \infty$ para todo n), $\mathbb{E}(X_n^2) < \infty$ (segundo momento finito) y $\mathbb{E}(X_n X_m)$ depende únicamente de la diferencia $n - m$.

Ejercicio 2.27. Probar que si $\{X_n\}_{n\in\mathbb{N}}$ es iid con $\mathbb{E}(X_n) = a$ y $\mathrm{Var}(X_n) = \sigma^2$ para todo n entonces es débilmente estacionaria.

Teorema 2.28. *Consideremos $\{X_n\}_{n\in\mathbb{N}}$ una sucesón débilmente estacionaria de variables aleatorias. Supongamos que $\rho(X_n, X_m) \to 0$ cuando $|n-m| \to \infty$, entonces si denotamos $\mathbb{E}(X_i) = a$*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} a.$$

Demostración. Vamos a demostrar que se verifica (2.7). Observemos que como X_n es débilmente estacionaria $\mathbb{E}(X_n^2)$ no depende de n y por lo tanto tampoco $\mathrm{Var}(X_n)$, denotemos $\mathrm{Var}(X_n) = \sigma^2$.

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\mathrm{Var}(X_i) + 2\sigma^2 \sum_{1\leq i<k\leq n} \rho(X_i, X_k) = n\sigma^2 + 2\sigma^2(T_1 + T_2),$$

donde

$$T_1 = \sum_{\substack{1\leq i<k\leq n\\ |i-k|\leq M}} \rho(X_i, X_k) \qquad T_2 = \sum_{\substack{1\leq i<k\leq n\\ |i-k|> M}} \rho(X_i, X_k),$$

sea $\varepsilon > 0$ y M tal que $|\rho(X_i, X_k)| < \varepsilon$ si $|i-k| > M$, por lo tanto $T_2 \leq \varepsilon n^2$. Para acotar T_1 vamos a calcular la cantidad de sumandos que intervienen en dicha sumatoria. Este número se puede calcular como la antidad de entradas de una matriz de tamaño $n \times n$ cuyos índices verifican que $0 < |i-j| \leq M$. Esto nos da

$$2\sum_{j=1}^{M}(n-j) = 2\Big(nM - \sum_{j=1}^{M} j\Big) = M(2n - M - 1)$$

donde hemos usado la formula $\sum_{j=1}^{M} j = M(M+1)/2$. Cada sumando $\rho(X_i, X_k)$ de esta suma está acotado superiormente por 1, por lo tanto

$$0 \leq \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \leq \frac{\sigma^2}{n} + \sigma^2\left(\frac{M(2n-M-1)}{n^2} + \varepsilon\right),$$

si tomamoslím sup

$$0 \leq \limsup_{n\to\infty} \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \leq \sigma^2\varepsilon.$$

Como ε es arbitrario obtenemos (2.7). $\square$

2.4. Ley fuerte de los grandes números

Antes de enunciar y demostrar distintas versiones de la Ley fuerte de los grandes números vamos a demostrar unos lemas sobre sucesiones (no aleatorias) que serán necesarios.

Lema 2.29. ***(Kronecker.)*** *Sea $\{x_n\}_{n\in\mathbb{N}}$ una sucesión de números tal que $\sum_{n=1}^{\infty} x_n = s$, sea $b_n \uparrow +\infty$ una sucesión no decreciente de números reales positivos, entonces*

$$\lim_{n\to\infty} \frac{1}{b_n}\sum_{k=1}^{n} b_k x_k = 0,$$

en particular si $b_n = n$ *y* $x_n = y_n/n$ *tal que* $\sum y_n/n$ *converge, entonces*

$$\frac{y_1 + \cdots + y_n}{n} \to 0.$$

Demostración. Denotemos $S_k = \sum_{n=1}^{k} x_n$, tomemos $b_0 = 0$ y $S_0 = 0$, se prueba por inducción completa que:

$$\sum_{k=1}^{n} b_k x_k = b_n S_n - \sum_{k=1}^{n} (b_k - b_{k-1}) S_{k-1}.$$

Dado $\varepsilon > 0$ sea N tal que $|S_{k-1} - s| < \varepsilon$ si $k \geq N$. Por lo tanto, tomando $n > N$, primero separamos la suma en dos y luego sumamos y restamos $\frac{1}{b_n} \sum_{k=N}^{n} (b_k - b_{k-1}) s$,

$$\begin{aligned} \frac{1}{b_n} \sum_{k=1}^{n} b_k x_k =& S_n - \frac{1}{b_n} \sum_{k=1}^{N-1} (b_k - b_{k-1}) S_{k-1} - \frac{1}{b_n} \sum_{k=N}^{n} (b_k - b_{k-1}) S_{k-1} \\ =& S_n - \frac{1}{b_n} \sum_{k=1}^{N-1} (b_k - b_{k-1}) S_{k-1} - \frac{1}{b_n} \sum_{k=N}^{n} (b_k - b_{k-1}) s - \frac{1}{b_n} \sum_{k=N}^{n} (b_k - b_{k-1})(S_{k-1} - s). \end{aligned} \tag{2.9}$$

Observemos que

$$\frac{1}{b_n} \sum_{k=N}^{n} (b_k - b_{k-1}) s = \frac{s}{b_n} \sum_{k=N}^{n} (b_k - b_{k-1}) = \frac{s}{b_n} (b_n - b_{N-1}),$$

por lo tanto, sustituyendo en (2.9)

$$\frac{1}{b_n} \sum_{k=1}^{n} b_k x_k = S_n - \frac{1}{b_n} \sum_{k=1}^{N-1} (b_k - b_{k-1}) S_{k-1} - \frac{s}{b_n} (b_n - b_{N-1}) - \frac{1}{b_n} \sum_{k=N}^{n} (b_k - b_{k-1})(S_{k-1} - s).$$

Entonces,

$$\left| \frac{1}{b_n} \sum_{k=1}^{n} b_k x_k \right| \leq \left| S_n - (b_n - b_{N-1}) \frac{s}{b_n} \right| + \left| \frac{1}{b_n} \sum_{k=1}^{N-1} (b_k - b_{k-1}) S_{k-1} \right| + \left| \frac{1}{b_n} \sum_{k=N}^{n} (b_k - b_{k-1})(S_{k-1} - s) \right|. \tag{2.10}$$

Como $S_n \to s$, $b_n \uparrow \infty$ y N es fijo

$$\left| S_n - \frac{b_n - b_{N-1}}{b_n} s \right| \to 0 \text{ cuando } n \to \infty. \tag{2.11}$$

Nuevamente, como N es fijo y $b_n \to \infty$,

$$\left| \frac{1}{b_n} \sum_{k=1}^{N-1} (b_k - b_{k-1}) S_{k-1} \right| \to 0 \text{ cuando } n \to \infty. \tag{2.12}$$

Finalmente, como $0 < b_{k-1} \leq b_k$ para todo k

$$\left| \frac{1}{b_n} \sum_{k=N}^{n} (b_k - b_{k-1})(S_{k-1} - s) \right| \leq \frac{1}{b_n} \sum_{k=N}^{n} (b_k - b_{k-1}) |S_{k-1} - s| \leq \varepsilon (b_n - b_{N-1})/b_n \leq \varepsilon. \tag{2.13}$$

Usando (2.11), (2.12) y (2.13) en (2.10) obtenemos que, para todo $\varepsilon > 0$,

$$\lim_{n \to \infty} \left| \frac{1}{b_n} \sum_{k=1}^{n} b_k x_k \right| \leq \varepsilon,$$

por lo tanto, como ε es arbitrario,

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^{n} b_k x_k = 0.$$

$\square$

Teorema 2.30. *(**Kolmogorov.**) Sea $\{X_n\}_{n\in\mathbb{N}}$ un sucesión de variables aleatorias independientes con segundo momento finito, supongamos que existen números positivos b_n tal que $b_n \uparrow \infty$ y*

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{b_n^2} < \infty \tag{2.14}$$

entonces

$$\frac{S_n - \mathbb{E}(S_n)}{b_n} \xrightarrow{c.s.} 0$$

en particular si

$$\sum_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty,$$

entonces

$$\frac{S_n - \mathbb{E}(S_n)}{n} \xrightarrow{c.s.} 0,$$

Demostración. Observemos que

$$\frac{S_n - \mathbb{E}(S_n)}{b_n} = \frac{1}{b_n} \sum_{k=1}^{n} b_k \left(\frac{X_k - \mathbb{E}(X_k)}{b_k} \right),$$

Fijado ω, $\{(X_k(\omega) - \mathbb{E}(X_k))/b_k\}_k$ es una sucesión de números reales que podemos llamar x_k, Por lo tanto por el Lema anterior basta probar que existe $D \subset \Omega$ tal que $P(D) = 1$ y para todo $\omega \in D$,

$$\sum_{k=1}^{n} \left(\frac{X_k(\omega) - \mathbb{E}(X_k)}{b_k} \right) = s(\omega) < \infty,$$

pero esto se sigue del Teorema 2.14, usando (2.14). $\square$

Veamos ahora una versión del teorema anterior, para el caso de variables idénticamente distribuidas, sin la hipótesis de finitud del segundo momento. Para eso vamos a necesitar dos lemas:

Lema 2.31. *Sea X una variable no negativa, entonces*

$$\sum_{n=1}^{\infty} P(X \geq n) \leq \mathbb{E}(X) \leq 1 + \sum_{n=1}^{\infty} P(X \geq n). \tag{2.15}$$

Demostración.

$$\begin{aligned}
\sum_{n=1}^{\infty} P(X \geq n) &= \sum_{n=1}^{\infty} \sum_{k\geq n} P(k \leq X < k+1) \\
&= \sum_{k=1}^{\infty} kP(k \leq X < k+1) = \sum_{k=0}^{\infty} \mathbb{E}\Big[k\mathbb{I}_{\{k\leq X<k+1\}}\Big]. \qquad (2.16)
\end{aligned}$$

Observemos que, $k\mathbb{I}_{\{k\leq X<k+1\}} \leq X\mathbb{I}_{\{k\leq X<k+1\}} < (k+1)\mathbb{I}_{\{k\leq X<k+1\}}$, por lo tanto,

$$\mathbb{E}(k\mathbb{I}_{\{k\leq X<k+1\}}) \leq \mathbb{E}(X\mathbb{I}_{\{k\leq X<k+1\}}) < \mathbb{E}((k+1)\mathbb{I}_{\{k\leq X<k+1\}}), \tag{2.17}$$

Por el Teorema 2.9 (ver también Observación 2.10 para el caso $\mathbb{E}(X) = \infty$), obtenemos

$$\sum_{k=0}^{\infty} \mathbb{E}\Big[X\mathbb{I}_{\{k\leq X<k+1\}}\Big] = \mathbb{E}(X). \tag{2.18}$$

Finalmente

$$\sum_{n=1}^{\infty} P(X \geq n) \leq \sum_{k=0}^{\infty} \mathbb{E}\Big[X\mathbb{I}_{\{k\leq X<k+1\}}\Big] = \mathbb{E}(X) \tag{2.19}$$

Esto prueba la primera desigualdad en (2.15). Para probar la segunda observemos que de (2.17) y (2.18) se sigue que

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{\infty} \mathbb{E}\Big[X\mathbb{I}_{\{k\leq X<k+1\}}\Big] \qquad (2.20)\\
&< \sum_{k=0}^{\infty} \mathbb{E}\Big[(k+1)\mathbb{I}_{\{k\leq X<k+1\}}\Big]\\
&= \sum_{k=0}^{\infty} (k+1)P(k \leq X < k+1)\\
&= \sum_{k=0}^{\infty} kP(k \leq X < k+1) + \sum_{k=0}^{\infty} P(k \leq X < k+1).
\end{aligned}$$

Observemos que, como la variable es no negativa, el segundo sumando es

$$\sum_{k=0}^{\infty} P(k \leq X < k+1) = P(X \geq 0) = 1, \tag{2.21}$$

y ya hemos probado en (2.16) que el primero es

$$\sum_{n=1}^{\infty} P(X \geq n) = \sum_{k=0}^{\infty} kP(k \leq X < k+1). \tag{2.22}$$

Por lo tanto con (2.21) y (2.22) se concluye (2.15). □

Lema 2.32. *Sea X una variable no negativa, entonces*

$$\mathbb{E}(X) = \int_0^{+\infty} P(X > x)dx. \tag{2.23}$$

Demostración. La demostración es muy similar a la anterior, basta tomar $A_{k,n} = \{\omega : X(\omega) \in [k/2^n, (k+1)/2^n)\}$ para $k \geq 0$, y la variable aleatoria

$$S_n = \sum_{k=1}^{\infty} \frac{k}{2^n}\mathbb{I}_{A_{k,n}}$$

observemos que $S_n \uparrow X$ y por lo tanto usando el teorema de convergencia monótona, $\mathbb{E}(S_n) \uparrow \mathbb{E}(X)$. Vamos a acostar superiormente $\mathbb{E}(S_n)$ por $\int_0^\infty P(X > x)dx$ para eso:

$$\mathbb{E}(S_n) = \sum_{k=1}^{\infty} \sum_{i=1}^{k} \frac{1}{2^n} P(A_{k,n}) = \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} \frac{1}{2^n} P(A_{k,n}) = \frac{1}{2^n} \sum_{i=1}^{\infty} P\left(X \geq \frac{i}{2^n}\right) \tag{2.24}$$

Observemos que

$$\frac{1}{2^n} \sum_{i=1}^{\infty} P\left(X \geq \frac{i}{2^n}\right) = \frac{1}{2^n} \sum_{i=0}^{\infty} P\left(X \geq \frac{i+1}{2^n}\right)$$

Como la función $P(X > x)$ es decreciente (en sentido amplio), la serie anterior corresponde a la suma inferior de Riemann de $P(X > x)$ en la partición $k/2^n$ con $k = 0, 1, 2, \ldots$. Por lo tanto es menor o igual que $\int_0^{+\infty} P(X > x)dx$.

Para acotar inferiormente $\mathbb{E}(S_n)$ observemos que, como $P(X > x)$ es decreciente

$$\int_{1/2^n}^{+\infty} P(X > x)dx = \sum_{i=1}^{\infty} \int_{i/2^n}^{(i+1)/2^n} P(X > x)dx \leq \sum_{i=1}^{\infty} \int_{i/2^n}^{(i+1)/2^n} P(X > i/2^n)dx,$$

como

$$\int_{i/2^n}^{(i+1)/2^n} P\left(X > \frac{i}{2^n}\right) dx = \frac{1}{2^n} P\left(X > \frac{i}{2^n}\right),$$

obtenemos finalmente que

$$\int_{1/2^n}^{+\infty} P(X > x)dx \leq \sum_{i=1}^{\infty} \frac{1}{2^n} P\left(X > \frac{i}{2^n}\right) = \mathbb{E}(S_n) \tag{2.25}$$

donde en la última igualdad hemos usado la ecuación (2.24). Finalmente, de (2.25) y (2.24) obtenemos que

$$\int_{1/2^n}^{+\infty} P(X > x)dx \leq \mathbb{E}(S_n) \leq \int_0^{+\infty} P(X > x)dx,$$

domando límite cuando $n \to +\infty$ se concluye (2.23). $\square$

Lema 2.33. ***(Toeplitz.)*** *Sea $\{a_n\}_{n\in\mathbb{N}}$ una sucesión de números no negativos tal que $a_1 > 0$, sea $b_n = \sum_{i=1}^n a_n$, supongamos que $b_n \uparrow \infty$. Sea $\{x_n\}_{n\in\mathbb{N}}$ tal que $x_n \to x$, entonces*

$$\frac{1}{b_n} \sum_{j=1}^{n} a_j x_j \to x,$$

en particular si $a_n = 1$ para todo n entonces

$$\frac{x_1 + \cdots + x_n}{n} \to x$$

Demostración. Sea $\varepsilon > 0$ y $n_0 = n_0(\varepsilon)$ tal que $|x_n - x| \leq \varepsilon/2$ para todo $n \geq n_0$. Tomemos $n_1 > n_0$ tal que

$$\frac{1}{b_{n_1}} \sum_{j=1}^{n_0} |x_j - x| < \varepsilon/2.$$

Para todo $n > n_1$ tenemos que

$$\begin{aligned}
\left|\frac{1}{b_n}\sum_{j=1}^{n} a_j x_j - x\right| =& \frac{1}{b_n}\left|\sum_{j=1}^{n} a_j(x_j - x)\right| \\
\leq& \frac{1}{b_n}\sum_{j=1}^{n} a_j|x_j - x| \\
=& \frac{1}{b_n}\sum_{j=1}^{n_0} a_j|x_j - x| + \frac{1}{b_n}\sum_{j=n_0+1}^{n} a_j|x_j - x| \\
\leq& \frac{1}{b_{n_1}}\sum_{j=1}^{n_0} a_j|x_j - x| + \frac{1}{b_n}\sum_{j=n_0+1}^{n} a_j|x_j - x| \\
\leq& \frac{\varepsilon}{2} + \frac{b_n - b_{n_0}}{b_n}\frac{\varepsilon}{2} \leq \varepsilon
\end{aligned}$$

□

Teorema 2.34. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas tal que $\mathbb{E}|X_i| < \infty$, entonces*

$$\frac{S_n}{n} \overset{c.s.}{\to} \mathbb{E}(X_1) \tag{2.26}$$

Demostración. Supongamos sin pérdida de generalidad que $\mathbb{E}(X_n) = 0$ (si no fuese el caso tomamos las variables $X_i - \mathbb{E}(X_i)$). Observemos que, por el Lema 2.31, $\mathbb{E}(|X_1|) < \infty$ si y sólo si $\sum_n P(|X_1| \geq n) < \infty$. Como las X_i son idénticamente distribuidas tenemos que

$$\sum_n P(|X_1| \geq n) < \infty \Leftrightarrow \sum_n P(|X_n| \geq n) < \infty,$$

Por el Lema de Borel Cantelli tenemos que, si definimos $A_n = \{\omega : |X_n| \geq n\}$, $P(\limsup A_n) = 0$, es decir, por la Proposición 1.11 1. $|X_n| < n$ excepto una cantidad finita de n, con probabilidad 1. Denotemos

$$\tilde{X}_n = \begin{cases} X_n, & |X_n| < n \\ 0, & |X_n| \geq n \end{cases}$$

Veamos que, como $|X_n| < n$ excepto una cantidad finita de n, con probabilidad 1, entonces:

$$\frac{X_1 + \cdots + X_n}{n} \overset{c.s}{\to} 0 \Leftrightarrow \frac{\tilde{X}_1 + \cdots + \tilde{X}_n}{n} \overset{c.s}{\to} 0. \tag{2.27}$$

Sea $D \subset \Omega$, $P(D) = 1$ tal que si $\omega \in D$, existe $n_0(\omega)$ tal que para todo $n > n_0$ $|X_n(\omega)| < n$, por definición de $\tilde{X}_n$ tenemos que $\tilde{X}_n(\omega) = X_n(\omega)$ para $n > n_0$, sea $n > n_0$

$$\frac{\tilde{X}_1(\omega) + \cdots + \tilde{X}_n(\omega)}{n} = \frac{\tilde{X}_1(\omega) + \cdots + \tilde{X}_{n_0}(\omega)}{n} + \frac{X_{n_0+1}(\omega) + \cdots + X_n(\omega)}{n}.$$

Como n_0 es fijo, el primer sumando tiende a 0 cuando $n \to \infty$, de donde se deduce (2.27). Observemos que $\mathbb{E}(\tilde{X}_n)$ no es necesariamente 0, no obstante

$$\mathbb{E}(\tilde{X}_n) = \mathbb{E}(X_n \mathbb{I}_{\{|X_n|<n\}}) = \mathbb{E}(X_1 \mathbb{I}_{\{|X_1|<n\}}).$$

Por el Teorema de convergencia dominada (se puede aplicar ya que $X_1\mathbb{I}_{\{|X_1|<n\}} \leq X_1$ y $\mathbb{E}|X_1| < \infty$) obtenemos que $\mathbb{E}(X_1\mathbb{I}_{\{|X_1|<n\}}) \to \mathbb{E}(X_1) = 0$. Si aplicamos el Lema 2.33 con $a_i = 1$ y $x_n = \mathbb{E}(\tilde{X}_n)$ obtenemos que

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(\tilde{X}_n) \to \mathbb{E}(X_1) = 0, \quad n \to \infty,$$

como consecuencia

$$\frac{X_1 + \cdots + X_n}{n} \overset{c.s.}{\to} 0 \Leftrightarrow \frac{(\tilde{X}_1 - \mathbb{E}(\tilde{X}_1)) + \cdots + (\tilde{X}_n - \mathbb{E}(\tilde{X}_n))}{n} \overset{c.s.}{\to} 0.$$

Para probar que se verifica el último límite, basta probar, por el Lema 2.29, que la serie $\sum \hat{X}_n/n$ es convergente, siendo $\hat{X}_n = \tilde{X}_n - \mathbb{E}(\tilde{X}_n)$. Por el Teorema 2.14 para ver que $\sum \hat{X}_n/n$ converge basta ver que $\sum \mathrm{Var}(\hat{X}_n)/n^2$ converge. Si acotamos $\mathrm{Var}(\hat{X}_n) = \mathbb{E}(\tilde{X}_n^2) - (\mathbb{E}(\tilde{X}_n))^2 \leq \mathbb{E}(\tilde{X}_n^2)$,

$$\sum_{n=1}^{\infty}\frac{\mathrm{Var}(\hat{X}_n)}{n^2} \leq \sum_{n=1}^{\infty}\frac{\mathbb{E}(\tilde{X}_n^2)}{n^2} = \sum_{n=1}^{\infty}\frac{1}{n^2}\mathbb{E}\Big(\big[X_n\mathbb{I}_{\{|X_n|<n\}}\big]^2\Big) \tag{2.28}$$

Como $\mathbb{I}^2_{\{|X_n|<n\}} = \mathbb{I}_{\{|X_n|<n\}}$ tenemos que

$$\sum_{n=1}^{\infty}\frac{\mathrm{Var}(\hat{X}_n)}{n^2} \leq \sum_{n=1}^{\infty}\frac{1}{n^2}\mathbb{E}\big[X_n^2\mathbb{I}_{\{|X_n|<n\}}\big]$$

Como las X_i son idénticamente distribuidas podemos escribir

$$\mathbb{E}\big[X_n^2\mathbb{I}_{\{|X_n|<n\}}\big] = \mathbb{E}\big[X_1^2\mathbb{I}_{\{|X_1|<n\}}\big] = \sum_{k=1}^{n}\mathbb{E}\big[X_1^2\mathbb{I}_{\{(k-1)\leq|X_1|<k\}}\big]$$

de donde obtenemos que

$$\sum_{n=1}^{\infty}\frac{\mathrm{Var}(\hat{X}_n)}{n^2} \leq \sum_{n=1}^{\infty}\frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\big[X_1^2\mathbb{I}_{\{(k-1)\leq|X_1|<k\}}\big]$$

si damos vuelta la suma doble,

$$\sum_{n=1}^{\infty}\frac{\mathrm{Var}(\hat{X}_n)}{n^2} \leq \sum_{k=1}^{\infty}\mathbb{E}\big[X_1^2\mathbb{I}_{\{(k-1)\leq|X_1|<k\}}\big]\sum_{n=k}^{\infty}\frac{1}{n^2}.$$

Veamos que $\sum_{n=k}^{\infty}\frac{1}{n^2} \leq \frac{2}{k}$, para eso observemos que, como $1/n^2 \leq 2/(n(n+1))$,

$$\sum_{n=k}^{\infty}\frac{1}{n^2} \leq \sum_{n=k}^{\infty}\frac{2}{n(n+1)} = 2\sum_{n=k}^{\infty}\left[\frac{1}{n} - \frac{1}{n+1}\right]$$

La suma anterior es una *suma telescópica* y por lo tanto da $2/k$.

Por otra parte, como

$$X_1^2\mathbb{I}_{\{(k-1)\leq|X_1|<k\}} \leq k|X_1|\mathbb{I}_{\{(k-1)\leq|X_1|<k\}}$$

se obtiene que

$$\mathbb{E}\big[X_1^2\mathbb{I}_{\{(k-1)\leq|X_1|<k\}}\big] \leq k\mathbb{E}\big[|X_1|\mathbb{I}_{\{(k-1)\leq|X_1|<k\}}\big],$$

y finalmente

$$\sum_{n=1}^{\infty}\frac{\mathrm{Var}(\hat{X}_n)}{n^2} \leq 2\sum_{k=1}^{\infty}\mathbb{E}\big[|X_1|\mathbb{I}_{\{(k-1)\leq|X_1|<k\}}\big] = 2\mathbb{E}(|X_1|) < \infty.$$

□

Se deja como ejercicio probar el recríproco del teorema anterior:

Teorema 2.35. *Sea* $\{X_n\}_{n\in\mathbb{N}}$ *una sucesión de variables aleatorias independientes e idénticamente distribuidas tal que*

$$\frac{X_1 + \dots X_n}{n} \to C \quad c.s.$$

entonces: $\mathbb{E}|X_i| < \infty$ *y* $C = E(X_1)$.

Demostración. Escribir:

$$\frac{X_n}{n} = \frac{S_n}{n} - \left(\frac{n-1}{n}\right)\frac{S_{n-1}}{n-1} \tag{2.29}$$

Probar usando el Lema 1.13 2. y (2.29) que

$$\sum P(|X_1| > n) < \infty,$$

y concluir usando el Lema 2.31.

□

2.5. Convergencia en Distribución

Definición 2.36. Dada $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas en $(\Omega_n, \mathcal{A}_n, P_n)$ y X definida en $(\Omega, \mathcal{A}, P)$, se dice que $\{X_n\}_{n\in\mathbb{N}}$ converge en distribución a X si y sólo si

$$\lim_{n\to\infty} F_{X_n}(x) = F(x) \text{ para todo } x \text{ punto de continuidad de } F_X,$$

denotamos $X_n \xrightarrow{d} X$.

Observación 2.37. *Recordemos que una función de distribución* $F : \mathbb{R} \to [0,1]$ *es monótona no decreciente y verifica que* $F \in D$ *siendo* D *el espacio definido en el punto 7 del ejemplo 1.15. Por tal motivo se puede ver fácilmente que las discontinuidades de una función de distribución son de salto (el límite por izquierda y derecha no coinciden), con lo cual, como la suma de dichos saltos tiene que estar incluido en* $[0,1]$*, sólo pueden ser una cantidad numerable. Si* F *es continua la convergencia es uniforme como se demuestra en el siguiente teorema:*

Teorema 2.38. *Si* $X_n \xrightarrow{d} X$ *y* $F_X(x)$ *es continua para todo* x *entonces* $F_{X_n} \to F_X$ *uniformemente.*

Demostración. Denotemos $F_n = F_{X_n}$ y $F_X = F$. Sea $\varepsilon > 0$ existe $R = R(\varepsilon) > 0$ tal que $F(R) - F(-R) > 1 - \varepsilon$. Observar que entonces $F(-R) < \varepsilon$ y $F(R) > 1 - \varepsilon$. Sea N_0 tal que si $n > N_0$ $|F_n(R) - F(R)| < \varepsilon$ y $|F_n(-R) - F(-R)| < \varepsilon$ (N_0 existe ya que F es continua en R y en $-R$ y F_n converge puntualmente a F). Si $r \le -R$, como F_n y F son no decrecientes $|F_n(r) - F(r)| \le F_n(r) + F(r) \le F_n(-R) + F(-R) < (F(-R) + \varepsilon) + F(-R) \le 3\varepsilon$ si $n > N_0$. Si $r \ge R$, $1 - \varepsilon < F(R) \le F(r)$, además $1 - 2\varepsilon < F_n(r) < 1$, por lo tanto $|F_n(r) - F(r)| \le 2\varepsilon$. Por lo tanto hemos probado que si $n > N_0$ entonces $\sup_{r\in[-R,R]^c} |F_n(r) - F(r)| < 3\varepsilon$. Para ver el caso en que $r \in [-R, R]$ observemos que F es uniformemente continua en $[-R, R]$ y por lo tanto para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $|F(x) - F(y)| < \varepsilon$ si $|x - y| < \delta$. Sea $-R < x_1 < x_2 < \dots < x_k < R$ tal que $x_i - x_{i-1} < \delta$. Como $F_n(x_i) \to F(x)$ existe n_i tal que si $n > n_i$ $|F_n(x_i) - F(x_i)| < \varepsilon$, sea $N_1 = \max\{n_1, \dots, n_k\}$. Si $r \in [-R, R]$ sea i tal que $x_{i-1} < r < x_i$ y $n > N_1$, nuevamente usando que F_n y F son no decrecientes

$$F(x_{i-1}) - \varepsilon - F(x_i) \le F_n(x_{i-1}) - F(x_i) \le F_n(r) - F(r) \le F_n(x_i) - F(x_{i-1}) \le F(x_i) - \varepsilon - F(x_{i-1})$$

como $x_i - x_{i-1} < \delta$ se tiene que $|F(x_i) - \varepsilon - F(x_{i-1})| < 2\varepsilon$. Por lo tanto si tomamos $n > \max\{N_0, N_1\}$ tenemos que para todo $r \in \mathbb{R}$ $|F_n(r) - F(r)| \le 3\varepsilon$ y por lo tanto $\lim_{n\to\infty} \sup_{r\in\mathbb{R}} |F_n(r) - F(r)| = 0$, es decir F_n converge a F uniformemente. □

El siguiente teorema da una definición equivalente de convergencia en distribución que, a diferencia de la anterior, permite definir convergencia en distribución de variables aleatorias definidas en un espacio métrico cualquiera.

Teorema 2.39. *Dada $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas en $(\Omega_n, \mathcal{A}_n, P_n)$ y X definida en $(\Omega, \mathcal{A}, P)$, $\{X_n\}_{n\in\mathbb{N}}$ converge en distribución a X si y sólo si*

$$\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)) \qquad \text{para toda } f \text{ a valores reales, continua y acotada .}$$

Demostración. Aquí solo probaremos el directo, el recíproco se probará mas adelante (ver Teorema 3.11). Lo haremos para el caso de variables aleatorias a valores reales, el caso general es similar. Denotemos F_n la función de distribución de X_n y F la de X. Tomemos $g : \mathbb{R} \to \mathbb{R}$ tal que $|g(x)| \leq c$ para todo $x \in \mathbb{R}$. Para todo $a < b$ tenemos que

$$|\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))| = \Big| \int_{-\infty}^{+\infty} g(x)dF_n(x) - \int_{-\infty}^{+\infty} g(x)dF(x) \Big| \leq \Big| \int_{-\infty}^{+\infty} g(x)dF_n(x) - \int_a^b g(x)dF_n(x) \Big|$$

$$+ \Big| \int_a^b g(x)dF_n(x) - \int_a^b g(x)dF(x) \Big| + \Big| \int_a^b g(x)dF(x) - \int_{-\infty}^{+\infty} g(x)dF(x) \Big| = I_1 + I_2 + I_3$$

$$I_3 = \Big| \int_{-\infty}^{a} g(x)dF(x) + \int_b^{+\infty} g(x)dF(x) \Big| \leq c(F(a) + 1 - F(b))$$

dado que $(F(a) + 1 - F(b)) \to 0$ cuando $a \to -\infty$ y $b \to +\infty$, para todo $\varepsilon > 0$, podemos tomar $a < b$ tal que $c(F(a) + 1 - F(b)) < \varepsilon$. Como los puntos de discontinuidad de F son a lo sumo numerables podemos suponer que a y b son puntos de continuidad de F. Si procedemos con I_1 de la misma manera que hicimos con I_3 obtenemos que $I_1 \leq c(F_n(a) + 1 - F_n(b))$. Dado que a y b son puntos de continuidad de F y X_n converge en distribución a F, podemos tomar n suficientemente grande tal que $|c(F_n(a) + 1 - F_n(b)) - c(F(a) + 1 - F(b))| < \varepsilon$, es decir $I_1 < 2\varepsilon$. Para acotar I_2 tomemos $a = x_0 < x_1 < \ldots < x_N = b$ una partición de a, b tal que $|g(x) - g(x_i)| < \varepsilon$ para todo x_i, $i = 0, \ldots, N-1$ y para todo $x \in [x_i, x_{i+1}]$ (esto se puede hacer ya que g es continua). Podemos suponer ademas que los x_i son puntos de continuidad de F.

$$I_2 = \Big| \sum_{i=0}^{N} \int_{x_i}^{x_{i+1}} g(x)dF_n(x) - \int_{x_i}^{x_{i+1}} g(x)dF(x) \Big|.$$

Observemos que

$$m_{n_i} = (g(x_i) - \varepsilon)(F_n(x_{i+1}) - F_n(x_i)) \leq \int_{x_i}^{x_{i+1}} g(x)dF_n(x) \leq (g(x_i) + \varepsilon)(F_n(x_{i+1}) - F_n(x_i)) = M_{n_i}$$

y

$$m_i = (g(x_i) - \varepsilon)(F(x_{i+1}) - F(x_i)) \leq \int_{x_i}^{x_{i+1}} g(x)dF(x) \leq (g(x_i) + \varepsilon)(F(x_{i+1}) - F(x_i)) = M_{_i}.$$

Por lo tanto

$$m_{n_i} - M_i \leq \int_{x_i}^{x_{i+1}} g(x)dF_n(x) - \int_{x_i}^{x_{i+1}} g(x)dF(x) \leq M_{n_i} - m_i,$$

sumando en i,

$$\sum_{i=0}^{N-1} (m_{n_i} - M_i) \leq \int_{x_i}^{x_{i+1}} g(x)dF_n(x) - \int_{x_i}^{x_{i+1}} g(x)dF(x) \leq \sum_{i=0}^{N-1} (M_{n_i} - m_i). \tag{2.30}$$

Como los x_i son puntos de continuidad, $m_{n_i} \to m_i$, y $M_{n_i} \to M_i$, para todo i. Observar que $M_i - m_i = 2\varepsilon(F(x_{i+1}) - F(x_i))$ y por lo tanto

$$\lim_{n\to\infty} \sum_{i=0}^{N-1} (m_{n_i} - M_i) = -2\varepsilon \sum_{i=0}^{N-1} (F(x_{i+1}) - F(x_i)) = -2\varepsilon(F(b) - F(a)) > -2\varepsilon,$$

análogamente

$$\lim_{n\to\infty} \sum_{i=0}^{N-1} (M_{n_i} - m_i) < 2\varepsilon,$$

por lo tanto, tomando n suficientemente grande en (2.30) se tiene que $I_2 < 3\varepsilon$. $\square$

Proposición 2.40. *Sean* $\{X_n\}_{n\in\mathbb{N}}$ *y* X *variables aleatorias definidas en* $(\Omega, \mathcal{A}, P)$, *supongamos que* $X_n \xrightarrow{P} X$ *y sean* $a < x < b$ *números reales, entonces*

$$F_X(a) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F_X(b)$$

Demostración.

$$\begin{aligned} F_X(a) =&P(\{X \leq a\} \cap \{X_n - X \geq x - a\}) + P(\{X \leq a\} \cap \{X_n - X < x - a\}) \\ \leq&P(X_n - X \geq x - a) + P(\{X \leq a\} \cap \{X_n - X < x - a\}) \\ \leq&P(|X_n - X| \geq x - a) + P(X_n \leq x) \end{aligned}$$

Donde en la última desigualdad hemos usado que $\{X \leq a\} \cap \{X_n - X < x - a\} \subset \{X_n \leq x\}$ y $\{X_n - X \geq x - a\} \subset \{|X_n - X| \geq x - a\}$. El primer sumando tiende a 0 cuando $n \to \infty$ ya que $X_n \xrightarrow{P} X$. Por lo tanto tomando límite inferior obtenemos que $F_X(a) \leq \liminf F_n(x)$. Para demostrar que $\limsup F_n(x) \leq F_X(b)$, procedemos de forma análoga:

$$\begin{aligned} F_{X_n}(x) =&P(\{X_n \leq x\}) \\ =&P(\{X_n \leq x\} \cap \{X - X_n \leq b - x\}) + P(\{X_n \leq x\} \cap \{X - X_n > b - x\}) \\ \leq&P(X \leq b) + P(|X - X_n| > b - x), \end{aligned}$$

donde usamos que $\{X_n \leq x, X - X_n \leq b - x\} \subset \{X \leq b\}$ y $\{X_n \leq x, X - X_n > b - x\} \subset \{|X - X_n| > b - x\}$, luego se concluye tomándo límite superior. $\square$

Veamos que de la proposición anterior se sigue que la convergencia en probabilidad es mas fuerte (implica) la convergencia en distribución.

Corolario 2.41. *Sean* $\{X_n\}_{n\in\mathbb{N}}$ *y* X *variables aleatorias definidas en* $(\Omega, \mathcal{A}, P)$, *supongamos que* $X_n \xrightarrow{P} X$ *y sean* x *punto de continuidad de* F_X *entonces* $F_{X_n}(x) \to F_X(x)$.

Demostración. Tomemos $a_k = x - 1/k$ y $b_k = x + 1/k$, aplicando la proposición anterior tenemos que

$$F_X(a_k) \leq \liminf F_n(x) \leq \limsup F_n(x) \leq F_X(b_k),$$

el resultado se sigue de la continuidad de F en x. $\square$

El recíproco del corolario anterior es falso, basta tomar $X \sim N(0,1)$ y $X_n = (-1)^n X$. Como X es simétrica es claro que $F_X(x) = F_{X_n}(x)$ para todo x y n, no obstante $|X_{2n+1} - X| = |-X - X| = 2|X|$. Observemos que de este ejemplo se desprende una diferencia importante con la convergencia en probabilidad y c.s. Si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{d} Y$, no es cierto en general que $X_n + Y_n \xrightarrow{d} X + Y$. Basta tomar como $Y_n = X_{2n}$

y $X_n = X_{2n+1}$ entonces $X_n + Y_n = 0$ para todo n. Por otra parte es inmediato verificar a partir de la definición de convergencia en distribución que si $X_n \xrightarrow{d} X$ y $X_n \xrightarrow{d} Y$ entonces $F_X(x) = F_Y(x)$ para todo x punto de continuidad de X e Y (en este caso denotamos $X \stackrel{d}{=} Y$). De los ejemplos anteriores vemos que $X \stackrel{d}{=} Y$ no implica $X \stackrel{c.s.}{=} Y$. Veamos ahora un teorema que permite en algún sentido *pasar* de la convergencia en distribucón a la convergencia casi segura. No veremos la demostración ya que requiere de técnicas de analisis real, no obstante la misma puede encontrarse en el libro [1] página 70.

Teorema 2.42. ***(Skorokhod.)*** *Dada $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias definidas en $(\Omega_n, \mathcal{A}_n, P_n)$ y X definida en $(\Omega_0, \mathcal{A}_0, P_0)$, tal que $X_n \xrightarrow{d} X$, existen X_n^* y X^* definidas todas en cierto $(\Omega, \mathcal{A}, P)$ tal que $X_n \stackrel{d}{=} X_n^*$ para todo n y $X \stackrel{d}{=} X^*$ y*

$$X_n^* \xrightarrow{c.s.} X^*$$

El teorema anterior es muy útil para probar propiedades de la convergencia en distribución. Por ejemplo para probar que $X_n \xrightarrow{d} X$ entonces $g(X_n) \xrightarrow{d} g(X)$ siendo g continua, basta tomar X_n^* y X^* copias de X_n y X respectivamente (es decir $X_n \stackrel{d}{=} X_n^*$ y $X \stackrel{d}{=} X^*$), tal que $X_n^* \xrightarrow{c.s.} X^*$, por lo tanto sabemos que $g(X_n^*) \xrightarrow{c.s} g(X^*)$ y por lo tanto $g(X_n^*) \xrightarrow{d} g(X^*)$. Se puede demostrar que $g(X^*) \stackrel{d}{=} g(X)$ y $g(X_n^*) \stackrel{d}{=} g(X_n)$ (esto último se puede ver usando que si para todo A abierto $P(X \in A) = P(Y \in B) \Rightarrow F_X(x) = F_Y(x)$) por lo tanto $g(X_n) \xrightarrow{d} g(X)$.

Lema 2.43. ***(Lema de Slutsky)*** *Si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{P} c$ siendo c una constante, entonces*

- $X_n + Y_n \xrightarrow{d} X + c$
- $X_n Y_n \xrightarrow{d} cX$
- $X_n/Y_n \xrightarrow{d} X/c$ *siempre que* $c \neq 0$.

2.5.1. Métricas entre distribuciones

Definiremos distancias entre funciones de distribución F y G de dos variables aleatorias X e Y a valores reales, de modo que la convergencia en distribución de una sucesión de variables aleatorias sea equivalente a la convergencia en dicha distancia. Algunos de estos conceptos se pueden generalizar a variables aleatorias a valores en espacios métricos mas generales (ver, por ejemplo, Capítulo 2 del libro [3]).

Distancia de Levy

Definición 2.44. Sean F y G dos distribuciones de probabilidad en $\mathbb{R}$, la distancia de Lévy entre F y G se define como

$$d_L(F, G) = \inf \left\{\varepsilon > 0 : F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \quad \forall x \in \mathbb{R}\right\}.$$

Veamos que la convergencia en distribución definida en (2.36) es equivalente a la convergencia en la distancia de Levy.

Teorema 2.45. *Sean $\{F_n\}_{n\in\mathbb{N}}$ funciones de distribución de variables aleatorias a valores reales, y F otra función de distribución, entonces, son equivalentes*

1. $F_n(x) \to F(x)$ *para todo x de continuidad de F*
2. $d_L(F_n, F) \to 0$.

Demostración. Que 2) implica 1) se sigue de manera inmediata de que si x es un punto de continuidad de F, $F(x-\varepsilon)-\varepsilon \to F(x)$ y $F(x+\varepsilon) \to F(x)$ para $\varepsilon \to 0$. Para ver la otra implicancia tomemos $x_0 < x_1 < \ldots < x_N$ puntos de continuidad de F tal que $F(x_0) < \varepsilon/2$ y $F(x_N) > 1-\varepsilon/2$, y $x_{i+1}-x_i < \varepsilon$ para todo $i = 0, \ldots, N-1$. Sea n_0 tal que para todo $n > n_0$ y para todo i, $|F_n(x_i)-F(x_i)| < \varepsilon/2$. Entonces si $x_{i-1} < x < x_i$,

$$F_n(x) \leq F_n(x_i) < F(x_i) + \frac{\varepsilon}{2} \leq F(x+\varepsilon).$$

De la misma forma se prueba que $F_n(x) \geq F(x-\varepsilon)-\varepsilon$. □

Distancia de Wasserstein

Vamos a definir la distancia de Wasserstein entre dos distribuciones, la cual es el infimo de la distancia en L^2 de dos variables cuyas distribuciones son F y G.

Definición 2.46. La distancia de Wasserstein entre dos distribuciones de probabilidad F y G se define como

$$\mathcal{W}_2^2(F,G) = \text{ínf}\left\{E|X-Y|^2 : X \sim F, Y \sim G\right\}$$

Si $X \sim F$ y $Y \sim G$, de la desigualdad $cov(X,Y) \leq cov(F^{-1}(U), G^{-1}(U))$ siendo U una variable con distribución uniforme en $[0,1]$ (observar que $X \sim F$ e $Y \sim G$). se puede ver fácilmente que la distancia de Wasserstein se realiza tomando como variables $F^{-1}(U)$ y $G^{-1}(U)$. Se puede probar que $\mathcal{W}_2^2(F_n,F) \to 0$ si y sólo si F_n converge en distribución a F y el segundo momento de variables con distribución F_n converge al segundo momento de una variable con distribución F, es decir $\int x^2 dF_n(x) \to \int x^2 dF(x)$.

2.6. Convergencia en L^p

Definición 2.47. Sean X y $\{X_n\}_{n\in\mathbb{N}}$ variables aleatorias definidas en el mismo espacio $(\Omega, \mathcal{A}, P)$, decimos que X_n converge en L^p (siendo $p > 0$) o en media de orden p a X si y sólo si

$$\mathbb{E}|X_n - X|^p \to 0,$$

denotamos $X_n \overset{L^p}{\to} X$

Omitiremos la prueba de las propiedades de la convergencia en L^p que se resumen en el siguiente ejercicio

Ejercicio 2.48.

- Probar usando la desigualdad de Chebyshev que si $X_n \overset{L^p}{\to} X$ entonces $X_n \overset{P}{\to} X$.
- Un ejemplo de que el recíproco no es cierto es el siguiente: Sea $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}([0,1])$ y P la probabilidad dada por la distribución de una variable uniforme en $[0,1]$, las variables

$$X_n(\omega) = \begin{cases} e^n, & 0 \leq \omega \leq 1/n \\ 0, & \omega > 1/n \end{cases}.$$

  Verificar que $X_n \overset{P}{\to} 0$ y $X_n \overset{L^p}{\to} +\infty$.
- El ejemplo 2.18 muestra que $X_n \overset{L^p}{\to} X$ no implica $X_n \overset{c.s.}{\to} X$. El recíproco también es falso, basta tomar X_n independientes tal que $P(X_n = 1) = 1/n$ y $P(X_n = 0) = 1-1/n$, y observar que si fuese $X_n \overset{c.s.}{\to} 0$, por el Lema de Borel-Cantelli, tendría que ser $\sum_{n=1}^{\infty} 1/n < \infty$ lo cual es falso.

Proposición 2.49. *Si* X_n *es una sucesión de variables no negativas tal que* $X_n \stackrel{c.s}{\to} X$ *y* $\mathbb{E}(X_n) \to \mathbb{E}(X) < \infty$ *entonces*

$$\mathbb{E}|X_n - X| \to 0,$$

Demostración. Observar que para $n > n_0$ $\mathbb{E}(X_n) < \infty$, que $0 \leq (X - X_n)\mathbb{I}_{X \geq X_n} \leq X$, concluir usando que

$$\begin{aligned}\mathbb{E}|X_n - X| =& \mathbb{E}\big[(X - X_n)\mathbb{I}_{\{X \geq X_n\}}\big] + \mathbb{E}\big[(X_n - X)\mathbb{I}_{\{X_n > X\}}\big] \\ =& 2\mathbb{E}\big[(X - X_n)\mathbb{I}_{\{X \geq X_n\}}\big] + \mathbb{E}(X_n - X)\end{aligned}$$

□

2.6.1. Desigualdades en espacios L^p

Teorema 2.50.

- ***Desigualdad de Jensen:*** *Si* g *es convexa y* $\mathbb{E}|X| < \infty$, $g(\mathbb{E}(X)) \leq \mathbb{E}(g(X))$
- ***Desigualdad de Lyapunov:*** *Si* $0 < s < t$ $(\mathbb{E}|X|^s)^{1/s} \leq (\mathbb{E}(X)^t)^{1/t}$
- ***Desigualdad de Hölder:*** *Si* $1 < p < \infty$ *y* $(1/p) + (1/q) = 1$, *si* $\mathbb{E}|X|^p < \infty$ *y* $\mathbb{E}|Y|^q < \infty$ *entonces* $\mathbb{E}|XY| < \infty$ *y*
$$\mathbb{E}|XY| \leq (\mathbb{E}(|X|^p)^{1/p}(\mathbb{E}|Y|^q)^{1/q}.$$
- ***Desigualdad de Minkowski:*** *Si* $\mathbb{E}|X|^p < \infty$ *y* $\mathbb{E}|Y|^p < \infty$, *con* $1 \leq p < \infty$, *entonces* $\mathbb{E}|X + Y|^p < \infty$ *y*
$$(\mathbb{E}|X + Y|^p)^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|Y|^p)^{1/p}$$

Capítulo 3

Funciones características y TCL

Vamos a introducir ahora una función que caracteriza la distribución de una variable, que permiitirá luego probar el Teorema Central del Límite. A lo largo del capítulo denotaremos $\mathbb{C}$ al conjunto de los números complejos y si $z = a + ib \in \mathbb{C}$ (con $a, b \in \mathbb{R}$), $|z|$ denota su ḿodulo, es decir $|z| = \sqrt{a^2 + b^2}$. El conjugado de z, que denotamos $\overline{z}$ es $\overline{z} = a - ib$. La exponencial compleja se define como $\exp(z) = \exp(a)\exp(ib) = \exp(a)(\cos(b) + i\operatorname{sen}(b))$.

3.1. Función característica

Definición 3.1. (Función característica.) Dado $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $X : \Omega \to \mathbb{R}$, definimos la función caracterśtica $\varphi_X : \mathbb{R} \to \mathbb{C}$ de X como $\varphi_X(t) = \mathbb{E}(e^{itX})$.

Uno de los objetivos fundamentales de este capítulo es probar que φ_X caracteriza la distribución de X, (es decir si $\varphi_X = \varphi_Y$ entonces $X \overset{d}{=} Y$) y que si $\varphi_{X_n}(t) \to \varphi_X(t)$ para todo t entonces $X_n \xrightarrow{d} X$. Esto último representa una herramienta importante a la hora de probar convergencia en distribución y en particular el Teorema Central del Límite.

Observación 3.2. *Observemos que dado que* $e^{itX} = \cos(tX) + i\operatorname{sen}(tX)$*, se tiene que*

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} dF_X(x) = \mathbb{E}(\cos(tX)) + i\mathbb{E}(\operatorname{sen}(tX)),$$

como $|e^{itX}| = 1$*, para todo* t*,* $\varphi(t)$ *existe y es finito para todo* t*.*

Veamos algunas propiedades que se siguen de manera inmediata de la definición, se dejan como ejercicio

Proposición 3.3.

1. $|\varphi_X(t)| \leq 1$ *para todo* $t \in \mathbb{R}$
2. $\varphi_X(0) = 1$
3. $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$.
4. *Si* X *e* Y *son independientes* $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$ *para todo* $t \in \mathbb{R}$*. El recíproco no es cierto, basta tomar* $X = Y$ *con distribución de Cauchy.*
5. $\varphi_X(t) = \overline{\varphi_X(-t)}$*, donde si* $z = a + ib$*,* $\overline{z} = a - ib$*.*

Proposición 3.4. *Si* $X \sim N(\mu, \sigma^2)$ *entonces*

$$\varphi_X(t) = e^{it\mu - \frac{1}{2}(\sigma t)^2}$$

Demostración. Observemos que como $\eta = \frac{X-\mu}{\sigma} \sim N(0,1)$, por la parte 3 de la proposición anterior, nos basta considerar el caso $X \sim N(0,1)$. Como $\operatorname{sen}(x)$ es impar y $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ es par,

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \int_{\mathbb{R}} \cos(tx) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx,$$

derivando respecto de t bajo el signo de integral (ver Teorema 2.13), e integrando por partes obtenemos que

$$\varphi_X'(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \operatorname{sen}(tx) \left(-xe^{-\frac{1}{2}x^2}\right) dx = \operatorname{sen}(tx) e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} - t \int_{-\infty}^{+\infty} \cos(tx) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

Observemos que como $\operatorname{sen}(tx)$ es acotada, el primer sumando es 0, mientras que el segundo es $-t\varphi_X(t)$, es decir φ_X verifica la ecuación $y' + ty = 0$ de donde $\varphi_X(t) = Ce^{-\frac{1}{2}t^2}$ y $C = 1$ ya que $\varphi_X(0) = 1$ $\square$

Proposición 3.5. *$\varphi_X(t)$ es uniformemente continua.*

Demostración. Observemos que $\varphi_X(t) - \varphi_X(s) = \mathbb{E}(e^{isX}(e^{i(t-s)X} - 1))$, por lo tanto $|\varphi_X(t) - \varphi_X(s)| \leq \mathbb{E}|(e^{i(t-s)X} - 1|$, como $|(e^{i(t-s)X} - 1| \leq 2 \in L^1$, basta observar que cuando $t, s \to 0$ entonces $e^{i(t-s)X} - 1 \xrightarrow{c.s} 0$, y luego aplicar convergencia dominada. $\square$

Proposición 3.6. *Si $\mathbb{E}|X|^n < \infty$ para algún $n \geq 1$ entonces existen las derivadas de φ_X de orden $1, \ldots, n$ y*

$$\varphi_X^{(r)}(t) = \int_{\mathbb{R}} (ix)^r e^{itx} dF(x) = i^r \mathbb{E}(X^r e^{itX}) \quad y \quad i^r \mathbb{E}(X^r) = \varphi_X^{(r)}(0) \tag{3.1}$$

además

$$\varphi_X(t) = \sum_{r=0}^{n} \frac{(it)^r}{r!} \mathbb{E}(X^r) + \frac{(it)^n}{n!} \varepsilon_n(t), \tag{3.2}$$

donde $|\varepsilon_n(t)| \leq 3\mathbb{E}(|X|^n)$ y $\varepsilon_n(t) \to 0$ cuando $t \to 0$.

Demostración. La prueba de 3.1 se hace por inducción completa, veamos el caso base $r = 1$. Observemos que por la desigualdad de Lyapunov $\mathbb{E}|X|^r < \infty$. Escribimos el cociente incremental:

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \mathbb{E}\left[e^{itX}\left(\frac{e^{ihX} - 1}{h}\right)\right],$$

como

$$\left|e^{itx}\left(\frac{e^{ihx} - 1}{h}\right)\right| = \left|\frac{e^{ihx} - 1}{h}\right| = \Big|\int_0^x e^{ihs} ds\Big| \leq \Big|\int_0^x |e^{ihs}| |ds| = |x|,$$

como $X \in L^1$ podemos usar convergencia dominada:

$$\lim_{h\to 0} \frac{\varphi(t+h) - \varphi(t)}{h} = \mathbb{E}\left[e^{itX} \lim_{h\to 0}\left(\frac{e^{ihX} - 1}{h}\right)\right] = \mathbb{E}\left[e^{itX} iX\right].$$

Para demostrar 3.2 observemos que , para $y \in \mathbb{R}$,

$$e^{iy} = \cos(y) + i\sin(y) = \sum_{k=0}^{n-1} \frac{(iy)^k}{k!} + \frac{(iy)^n}{n!}\left[\cos(\theta_1 y) + i\sin(\theta_2 y)\right],$$

con $|\theta_1| \leq 1$ y $|\theta_2| \leq 1$, entonces

$$e^{itX} = \sum_{k=0}^{n-1} \frac{(itX)^k}{k!} + \frac{(itX)^n}{n!}\big[\cos(\theta_1(\omega)tX) + i\sin(\theta_2(\omega)tX)\big],$$

tomando esperanza

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \sum_{k=0}^{n-1} \frac{(it)^k}{k!}\mathbb{E}(X^k) + \frac{(it)^n}{n!}\mathbb{E}\Big[X^n\big[\cos(\theta_1(\omega)tX) + i\sin(\theta_2(\omega)tX)\big]\Big],$$

Si escribimos:

$$\begin{aligned}\mathbb{E}\Big[X^n\big[\cos(\theta_1(\omega)tX) + i\sin(\theta_2(\omega)tX)\big]\Big] =& \mathbb{E}\Big[X^n\big[\cos(\theta_1(\omega)tX) + i\sin(\theta_2(\omega)tX) - 1 + 1\big]\Big]\\ =& \ \mathbb{E}\Big[X^n\big[\cos(\theta_1(\omega)tX) + i\sin(\theta_2(\omega)tX) - 1\big]\Big] + \mathbb{E}(X^n)\end{aligned}$$

si definimos

$$\varepsilon_n(t) = \mathbb{E}\Big[X^n\big[\cos(\theta_1(\omega)tX) + i\sin(\theta_2(\omega)tX) - 1\big]\Big]$$

obtenemos que

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \sum_{k=0}^{n-1} \frac{(it)^k}{k!}\mathbb{E}(X^k) + \frac{(it)^n}{n!}\mathbb{E}(X^n) + \frac{(it)^n}{n!}\varepsilon_n(t),$$

de donde se sigue (3.2).

$$|\varepsilon_n(t)| \leq \mathbb{E}\Big[|X^n|\big|\cos(\theta_1(\omega)tX) + i\sin(\theta_2(\omega)tX) - 1\big|\Big] \leq 3\mathbb{E}|X^n|$$

Para ver que $\varepsilon_n(t) \to 0$ observemos que como $\mathbb{E}|X^n| < \infty$ podemos usar convergencia dominada y

$$\lim_{t\to 0}\Big(\cos(\theta_1(\omega)tX) + i\sin(\theta_2(\omega)tX) - 1\Big) = 0.$$

□

Teorema 3.7. ***(Fórmula de inversión.)*** *Sea $F = F(x)$ una función de distribución y $\varphi(t) = \int_{-\infty}^{+\infty} e^{itx}dF(x)$ su función característica.*

1 Para todo par $a < b$ de puntos de continuidad de F,

$$F(b) - F(a) = \lim_{c\to+\infty} \frac{1}{2\pi}\int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it}\varphi(t)dt \tag{3.3}$$

2 Si $\int_{-\infty}^{+\infty} |\varphi(t)|dt < \infty$, F tiene densidad f dada por

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\varphi(t)dt \tag{3.4}$$

Demostración. Veamos la demostración de 1). Para eso observemos que

$$\begin{aligned}\frac{1}{2\pi}\int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it}\varphi(t)dt =& \frac{1}{2\pi}\int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it}\left[\int_{-\infty}^{+\infty} e^{itx}dF(x)\right]dt\\ =& \frac{1}{2\pi}\int_{-c}^{c}\left[\int_{-\infty}^{+\infty} \frac{e^{-ita} - e^{-itb}}{it}e^{itx}dF(x)\right]dt\end{aligned}$$

ahora vamos a aplicar el Teorema de Fubini de intercambio de integrales, para eso tenemos que probar que el integrando está en $L^1(dF \times dt)$, observemos que

$$\left|\frac{e^{-ita}-e^{-itb}}{it}e^{itx}\right| = \left|\frac{e^{-ita}-e^{-itb}}{it}\right| = \left|\int_a^b e^{-itx}dx\right| \leq b-a,$$

como $\int_{-\infty}^{+\infty}(b-a)dF(x) = b-a$ obtenemos que

$$\int_{-c}^{c}\left[\int_{-\infty}^{+\infty}\left|\frac{e^{-ita}-e^{-itb}}{it}e^{itx}\right|dF(x)\right]dt \leq \int_{-c}^{c}\int_{-\infty}^{+\infty}(b-a)dF(x)dt \leq 2c(b-a)$$

Si damos vuelta las integrales obtenemos que

$$\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)dt = \int_{-\infty}^{+\infty}\psi_c(x)dF(x) \tag{3.5}$$

siendo

$$\psi_c(x) = \frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}e^{itx}dt$$

en lo que resta de la prueba veremos que podemos usar convergencia dominada, para eso observemos que

$$\begin{aligned}\frac{e^{-ita}-e^{-itb}}{it}e^{itx} =& \frac{e^{it(x-a)}-e^{it(x-b)}}{it}\\ =& \frac{\cos(t(x-a))-\cos(t(x-b))}{it}+\frac{\operatorname{sen}(t(x-a))-\operatorname{sen}(t(x-b))}{t}\end{aligned}$$

Como

$$\int_{-c}^{c}\frac{\cos(t(x-a))}{t}dt = \int_{-c(x-a)}^{c(x-a)}\frac{\cos(u)}{u}du = 0$$

donde en la última igualdad usamos que $\frac{\cos(u)}{u}$ es impar (de igual manera $\int_{-c}^{c}\frac{\cos(t(x-b))}{t} = 0$), obtenemos que

$$\psi_c(x) = \frac{1}{2\pi}\int_{-c}^{c}\frac{\operatorname{sen}(t(x-a))-\operatorname{sen}(t(x-b))}{t}dt$$

si ahora hacemos los cambios de variable $v = t(x-a)$ (de donde $dv = (x-a)$) y $u = t(x-b)$ (de donde $du = (x-b)$) obtenemos que

$$\psi_c(x) = \frac{1}{2\pi}\int_{-c(x-a)}^{c(x-a)}\frac{\operatorname{sen}(v)}{v}dv - \frac{1}{2\pi}\int_{-c(x-b)}^{c(x-b)}\frac{\operatorname{sen}(u)}{u}du = \Psi_{1c}(x)+\Psi_{2c}(x)$$

consideremos $g(s,t) = \int_s^t \frac{\operatorname{sen}(v)}{v}dv$, se puede demostrar que $g(s,t)$ es unifórmemente continua y $g(s,t) \to \pi$ cuando $s \downarrow -\infty$ y $t \uparrow +\infty$ (esto último requiere de técnicas de análisis complejo y no lo veremos aquí). Veamos que para todo x $\lim_{c\to+\infty}\psi_c(x) = \psi(x)$ siendo ψ la función definida en el Ejercicio 1.46:

$$\psi(x) = \begin{cases} 0, & x<a, x>b \\ 1/2, & x=a, x=b \\ 1, & a<x<b \end{cases}.$$

Para eso basta considerar todos los posibles casos, veamos dos de ellos, el resto quedan como ejercicio: si $x = a$ es claro que $\Psi_{1c}(a) = 0$ para todo c y $\Psi_{2c}(a) \to 1/2$ cuando $c \to +\infty$ (aquí usamos que $g(s,t) \to \pi$). Si $x = b$, $\Psi_{2c}(b) = 0$ para todo c y $\Psi_{1c}(b) \to 1/2$ cuando $c \to +\infty$.

Como g es uniformemente continua y $g(s,t) \to \pi$ cuando $s \downarrow -\infty$ y $t \uparrow +\infty$ existe C constante tal que $|\psi_c(x)| < C < \infty$. Esto nos permite pasar al límite con $c \to \infty$ dentro de la integral en 3.5 y obtenemos que

$$\int_{-\infty}^{+\infty} \psi_c(x)dF(x) \to \int_{-\infty}^{+\infty} \psi(x)dF(x) \quad \text{cuando } c \to +\infty,$$

nuevamente por el ejercicio 1.46 sabemos que $\int_{-\infty}^{+\infty} \psi(x)dF(x) = F(b) - F(a)$ si a y b son puntos de continuidad de F, esto concluye la prueba de 3.3.

Para demostrar 2). Definimos,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx}\varphi(t)dt,$$

veamos que $f(x)$ es una función continua de x, sea $x_n \to x$, observemos que

$$e^{-itx_n}\varphi(t) \to e^{-itx}\varphi(t) \quad \text{cuando } n \to \infty, \ \forall t,$$

como $\int_{-\infty}^{+\infty} |\varphi(t)|dt < \infty$, y $|e^{-itx_n}| = 1$ podemos usar convergencia dominada y obtenemos que $f(x_n) \to f(x)$. Como f es continua, es integrable en $[a,b]$ para todo $[a,b]$. Si aplicamos el teorema de Fubini de intercambio de integrales,

$$\int_a^b f(x)dx = \int_a^b \frac{1}{2\pi} \left(\int_{-\infty}^{+\infty} e^{-itx}\varphi(t)dt \right) dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(t) \left[\int_a^b e^{-itx}dx \right] dt,$$

si calculamos

$$\int_a^b e^{-itx}dx = \frac{e^{-itx}}{-it}\Big|_a^b = \frac{e^{-ita} - e^{-itb}}{it},$$

por lo tanto

$$\int_a^b f(x)dx = \lim_{c\to+\infty} \frac{1}{2\pi} \int_{-c}^{+c} \frac{e^{-ita} - e^{-itb}}{it}\varphi(t)dt$$

por la parte 1 dicho límite es $F(b) - F(a)$ para todo a y b de continuidad de F. Observemos que hemos probado que $F(x) = \int_{-\infty}^{x} f(y)dy$, como f es continua podemos derivar F y obtenemos que $F' = f$, y como F es una distribución es creciente y por lo tanto $f \geq 0$, es decir f es la densidad de F. □

Observemos que, del Teorema anterior se sigue que la función de distribución caracteriza la distribución, es decir:

Corolario 3.8. *Si F y G son distribuciones tal que*

$$\int_{-\infty}^{+\infty} e^{itx}dF(x) = \int_{-\infty}^{+\infty} e^{itx}dG(x),$$

para todo $t \in \mathbb{R}$ entonces $F = G$.

Definición 3.9. (Función característica en general) Dado $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $X : \Omega \to H$, siendo H un espacio vectorial con un producto interno $\langle \cdot\, , \cdot \rangle$ definimos la función caracterśtica $\varphi_X : H \to \mathbb{C}$ de X como

$$\varphi_X(t) = \mathbb{E}\big(e^{i\langle t,X\rangle}\big).$$

Teorema 3.10. *Una condición necesaria y suficiente para que las componentes del vector* $\xi = (\xi_1, \dots, \xi_d)$ *sean independientes es que la función característica de* ξ *sea el producto de las funciones características de las* ξ_i*, es decir:*

$$\mathbb{E}\Big[\exp\big(i(t_1\xi_1 + \dots + t_d\xi_d)\big)\Big] = \prod_{k=1}^{d} \mathbb{E}\big(\exp(it_k\xi_k)\big) \quad (t_1, \dots, t_k) \in \mathbb{R}^d \tag{3.6}$$

Demostración. Veamos que si se cumple 3.6 para todo $t \in \mathbb{R}^d$ entonces son independientes. Para eso vamos a usar que la función característica de un vector caracteriza la distribución multidimensional. Denotemos F a la distribución de ξ, F_i la distribución de ξ_i y $G(x) = \prod_{k=1}^{d} F_k(x_k)$ siendo $x = (x_1, \dots, x_d)$, queremos probar que $F = G$. Como aplicación directa del teorema de Fubini podemos intercambiar las integrales y obtenemos que:

$$\int_{\mathbb{R}^d} \exp\big(i(t_1x_1 + \dots + t_dx_d)\big)dG(x) = \int_{\mathbb{R}^d} \Big[\exp(it_1x_1) \times \dots \times \exp(it_dx_d)\Big]dG(x) = \prod_{k=1}^{d} \int_{\mathbb{R}} e^{it_kx_k}dF_k(x_k)$$

Por otra parte

$$\int_{\mathbb{R}} e^{it_kx_k}dF_k(x_k) = \mathbb{E}(e^{it_k\xi_k}),$$

si aplicamos 3.6

$$\prod_{k=1}^{d} \mathbb{E}(\exp(it_k\xi_k)) = \mathbb{E}\Big[\exp\big(i(t_1\xi_1 + \dots + t_d\xi_d)\big)\Big] = \int_{\mathbb{R}^d} \exp\big(i(t_1x_1 + \dots + t_dx_d)\big)dF(x),$$

es decir obtuvimos

$$\int_{\mathbb{R}^d} \exp\big(i(t_1x_1 + \dots + t_dx_d)\big)dG(x) = \int_{\mathbb{R}^d} \exp\big(i(t_1x_1 + \dots + t_dx_d)\big)dF(x)$$

como la función característica multidimensional caracteriza la distribución obtenemos que $F = G$ y por lo tanto son independientes. ☐

Vamos a demostrar ahora el teorema que nos permitirá luego demostrar el Teorema Central del Límite.

Teorema 3.11. *Dada* $\{X_n\}_{n\in\mathbb{N}}$ *una sucesión de variables aleatorias definidas en* $(\Omega_n, \mathcal{A}_n, P_n)$ *y* X *definida en* $(\Omega, \mathcal{A}, P)$*, son equivalentes*

1 $X_n \xrightarrow{d} X$

2 $\varphi_{X_n}(t) \xrightarrow{n} \varphi_X(t)$ *para todo* $t \in \mathbb{R}$

Demostración. Veamos que 1 implica 2. Vamos a usar el Teorema 2.39 (el directo únicamente). Fijado $t \in \mathbb{R}$ consideremos las funciones $g_1(x) = \operatorname{sen}(tx)$ y $g_2(x) = \cos(tx)$, ambas son continuas y acotadas por lo que $\mathbb{E}(\operatorname{sen}(tX_n)) \to \mathbb{E}(\operatorname{sen}(tX))$ y $\mathbb{E}(\cos(tX_n)) \to_n \mathbb{E}(\cos(tX))$. Por lo tanto $\mathbb{E}(e^{itX_n}) \to_n \mathbb{E}(e^{itX})$.

Veamos que 2 implica 1. Denotemos F_n la función de distribución de X_n y F la función de distribución de X. Para demostrar que $F_n \xrightarrow{d} F$ basta probar que toda subsucesión $\{F_{n_j}\}_j$ de $\{F_n\}_n$ tiene una subsucesión que seguiremos denotando $\{F_{n_j}\}_j$, que converge a F en todo punto de continuidad de F. Veamos esto: supongamos que F_n no converge en distribución a F, entonces existe x_0 un punto de continuidad de F tal que $|F_{n_j}(x_0) - F(x_0)| > \varepsilon$ para infinitos j. Consideremos la sucesión de números $\{F_{n_{j_k}}(x_0)\}_k \subset [0, 1]$

siendo $\{F_{n_{j_k}}\}_k$ la subsucesión de $\{F_{n_j}\}_j$ que converge a F. Tenemos que $F_{n_{j_k}}(x_0) \to F(x_0)$, lo cual contradice que $|F_{n_j}(x_0) - F(x_0)| > \varepsilon$.

En lo que sigue vamos a construir una subsucesión $\{F_{n_{j_k}}\}_k$ de cualquier $\{F_{n_j}\}$ que converge a F (por simplicidad en la notación vamos a denotar $\{F_n\}_{n\in\mathbb{N}}$ en lugar de $\{F_{n_j}\}$). Consideremos una numeración de los números racionales $\mathbb{Q} = \{q_j\}_{j\in\mathbb{N}}$. Como $\{F_n(q_1)\}_n \subset [0,1]$, existe un conjunto de índices $N_1 = \{n_{1,j} : j \in \mathbb{N}\}$ tal que $\{F_{n_{1,j}}(q_1)\}_j$ converge. Nuevamente como $\{F_{n_{1,j}}(q_2)\}_j \subset [0,1]$ existe un conjunto de índices $N_2 = \{n_{2,j} : j \in \mathbb{N}\} \subset N_1$ tal que $\{F_{n_{2,j}}(q_2)\}_j$ converge. Si procedemos así obtenemos conjuntos de índices $N_1 \supseteq N_2 \supseteq N_3 \supseteq \dots$. Definimos $\{F_{n_j}\}_j := \{F_{n_{j,j}}\}_j$. Observemos que, para todo $q_k \in \mathbb{Q}$ existe $\lim_{j\to\infty} F_{n_{j,j}}(q_k)$ que llamamos $g(q_k)$. Definimos la función $G : \mathbb{R} \to [0,1]$ tal que

$$G(x) = \inf\{g(q_k) : q_k > x\} \quad \text{si } x \in \mathbb{R}\setminus\mathbb{Q}; \qquad G(q) = g(q) \quad \text{si } q \in \mathbb{Q}$$

Vamos aprobar que G es una función de distribución y que $F_{n_j}(x_0) \to G(x_0)$ para todo x_0 de continuidad de G.

1) G restricta a $\mathbb{Q}$ es monótona no decreciente: para todo $q < q'$, $F_{n_j}(q) \leq F_{n_j}(q')$ ya que las F_{n_j} son no decrecientes. Tomando límite en j obtenemos que $g(q) \leq g(q')$ y por lo tanto $G(q_i) \leq G(q_j)$ para todo $q_i < q_j$.

2) Es no decreciente en $\mathbb{R}$, sean $x_1 < x_2$ números reales, y $q_{1_i} \to_i x_1$ y $q_{2_i} \to_i x_2$ y $g(q_{1_i}) \to g(x_1)$ $g(q_{2_i}) \to g(x_2)$. Como G es no decreciente en $\mathbb{Q}$, $g(q_{1_i}) \leq g(q_{2_i})$ para $i > i_0$. Por lo tanto $G(x_1) \leq G(x_2)$.

3) Es continua por derecha: sea $x_k \downarrow x$ (decrece a x), llamemos $d = \lim_k G(x_k)$ (dicho límite existe ya que G es no decreciente), como $x_k \downarrow x$ y G es no decreciente (por la parte anterior) es claro que $G(x) \leq d$. Supongamos que $G(x) < d$, como $G(x) = \inf\{g(q_k) : q_k > x\} < d$ existe $y > x$, $y \in \mathbb{Q}$ tal que $g(y) < d$. Para k suficientemente grande $x < x_k < y$ y por lo tanto como G es monótona no decreciente $G(x_k) \leq g(y) < d$ y $\lim_k G(x_k) < d$ lo cual contradice que $d = \lim_k G(x_k)$.

4) $F_{n_j}(x_0) \to_j G(x_0)$ para todo x_0 de continuidad de G. Sea x_0 de continuidad de G, para todo $y > x_0$ $y \in \mathbb{Q}$, como las F_{n_j} son no decrecientes tenemos que

$$\limsup_j F_{n_j}(x_0) \leq \limsup F_{n_j}(y)$$

y por definición de g, $\limsup F_{n_j}(y) = g(y)$ si $y \in \mathbb{Q}$, por lo tanto, tomando ínfimo en y,

$$\limsup_j F_{n_j}(x_0) \leq \inf\{g(y) : y > x_0\} = G(x_0). \tag{3.7}$$

Sea $x_1 < y < x_0$, $y \in \mathbb{Q}$ entonces

$$G(x_1) \leq g(y) = \lim_j F_{n_j}(y) = \liminf_j F_{n_j}(y) \leq \liminf_j F_{n_j}(x_0),$$

Donde en la primer desigualdad hemos usado que G es no dereciente y $G(y) = g(y)$ si $y \in \mathbb{Q}$. En la última hemos usado que las F_{n_j} son no decrecientes. Si ahora tomamos $x_1 \uparrow x_0$ obtenemos que

$$G(x_0) = G(x_0^-) \leq \liminf_j F_{n_j}(x_0) \leq \limsup_j F_{n_j}(x_0) \leq G(x_0).$$

Donde la primera igualdad es porque x_0 es de continuidad de G y en la última usamos (3.7).

5) Los límites de G a $-\infty$ y $+\infty$ son 0 y 1 respectivamente. Observemos que como $\varphi_{X_{n_j}}(s) \to \varphi_X(s)$ para todo s y $|\varphi_{X_{n_j}}(s)| \leq 1$ podemos aplicar convergencia dominada y obtenemos que

$$\int_0^t \varphi_{X_{n_j}}(s)ds \to_j \int_0^t \varphi_X(s)ds \quad \text{para todo } t.$$

Por otra parte,

$$\int_0^t \varphi_{X_{n_j}}(s)ds = \int_0^t \left(\int_{-\infty}^{+\infty} e^{isu}dF_{n_j}(u)\right) ds = \int_{-\infty}^{+\infty} \left(\int_0^t e^{isu}ds\right) dF_{n_j}(u) = \int_{-\infty}^{+\infty} \left(\frac{e^{iut}-1}{iu}\right) dF_{n_j}(u).$$

Análogamente

$$\int_0^t \varphi_X(s)ds = \int_{-\infty}^{+\infty} \left(\frac{e^{iut}-1}{iu}\right) dF(u).$$

Como $F_{n_j}(u) \to G(u)$ para todo u de continuidad y $\frac{e^{iut}-1}{iu}$ es continua y acotada (la parte real e imaginaria son acotadas), se puede probar que:

$$\int_{-\infty}^{+\infty} \left(\frac{e^{iut}-1}{iu}\right) dF_{n_j}(u) \to_j \int_{-\infty}^{+\infty} \left(\frac{e^{iut}-1}{iu}\right) dG(u).$$

Este último paso no es inmediato ya que no sabemos que G sea una distribución aún (es lo que queremos probar). No obstante dicha convergencia se prueba usando las propiedades ya demostradas de G.

Obtuvimos finalmente que

$$\int_0^t \varphi_X(s)ds = \int_{-\infty}^{+\infty} \left(\frac{e^{iut}-1}{iu}\right) dF(u) = \int_{-\infty}^{+\infty} \left(\frac{e^{iut}-1}{iu}\right) dG(u),$$

para todo t, por lo tanto

$$\frac{1}{t}\int_0^t \varphi_X(s)ds = \frac{1}{t}\int_{-\infty}^{+\infty} \left(\frac{e^{iut}-1}{iu}\right) dG(u).$$

Observemos que como φ_X es continua, por el Teorema del valor medio para integrales

$$\frac{1}{t}\int_0^t \varphi_X(s)ds = \varphi_X(\theta) \quad \text{para algún } 0 \leq \theta \leq t$$

usando nuevamente que $\varphi_X(t)$ es continua (en $t=0$),

$$\lim_{t\to 0} \frac{1}{t}\int_0^t \varphi_X(s)ds = \lim_{t\to 0} \varphi_X(\theta) = \varphi_X(0) = 1. \tag{3.8}$$

Usando que $\lim_{w\to 0}(e^w - 1)/w = 1$ obtenemos, por el teorema de convergencia dominada,

$$\lim_{t\to 0} \frac{1}{t}\int_{-\infty}^{+\infty} \left(\frac{e^{iut}-1}{iu}\right) dG(u) = \int_{-\infty}^{+\infty} dG$$

Usando la definición de integral de Rieman-Stieltjes, es facil ver que

$$\int_{-\infty}^{+\infty} dG = \lim_{t\to+\infty} G(t) - \lim_{t\to-\infty} G(t)$$

Como $0 \leq G(x) \leq 1$ para todo x, se sigue de (3.8) que $G(+\infty) = 1$ y $G(-\infty) = 0$. Esto concluye la prueba de que G es una función de distribución, tenemos entonces que $F_{n_j} \xrightarrow{d} G$ cuando $j \to \infty$.

Como G es una distribución podemos encontrar una variable Y tal que G sea su distribución, por lo tanto, usando que 1 implica 2, $\varphi_{X_{n_j}}(t) \to \varphi_Y(t)$ para todo t, pero por hipótesis $\varphi_{X_{n_j}}(t) \to \varphi_X(t)$, es decir $\varphi_X(t) = \varphi_Y(t)$ para todo t, y por unicidad (es decir Corolario 3.8) $G = F$. $\square$

3.2. Teorema Central del Límite

3.2.1. Variables iid

Teorema 3.12. ***(P. Levy)*** *Sean $X_1, \dots, X_n$ iid con distribución F tal que $\mathbb{E}(X_1) = \mu$ y $Var(X_1) = \sigma^2$. Si denotamos $S_n = \sum_{i=1}^n X_i$ entonces*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1).$$

Demostración. Observemos que basta probarlo para el caso $\mathbb{E}(X_1) = 0$ y $\mathrm{Var}(X_1) = 1$: si definimos $Y_i = (X_i - \mu)/\sigma$

$$\frac{Y_1 + \cdots + Y_n}{\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Por el Teorema 3.11 (usando que 2 implica 1) y la Proposición 3.4 (con $\mu = 0$ y $\sigma = 1$) basta probar que

$$\varphi_{S_n/\sqrt{n}}(t) \to e^{-\frac{1}{2}t^2}.$$

Si usamos las propiedades 3. y 4. de la Proposición 3.3 obtenemos que

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right) = \prod_{i=1}^n \varphi_{X_i}\left(\frac{t}{\sqrt{n}}\right),$$

como las variables son idénticamente distribuidas,

$$\prod_{i=1}^n \varphi_{X_i}\left(\frac{t}{\sqrt{n}}\right) = \left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

Si usamos la expresión (3.2) con $n = 2$ obtenemos que

$$\varphi_X(t) = \varphi_X(0) + t\varphi_X'(0) + \frac{t^2}{2}\varphi_X^{(2)} - \frac{t^2}{2}\varepsilon_2(t) = 1 - \frac{t^2}{2} - \frac{t^2}{2}\varepsilon_2(t)$$

donde $\varepsilon_2(t) \to 0$, cuando $t \to 0$, finalmente

$$\varphi_{S_n/\sqrt{n}}(t) = \left(1 - \frac{t^2}{2n} - \frac{t^2}{2n}\varepsilon_2\Big(\frac{t}{\sqrt{n}}\Big)\right)^n = \exp\left(n\log\left[1 - \frac{t^2}{2n} - \frac{t^2}{2n}\varepsilon_2\Big(\frac{t}{\sqrt{n}}\Big)\right]\right),$$

usando el equivalente $\log(1+u) \sim u$ cuando $u \to 0$ obtenemos que

$$\exp\left(n\log\left[1 - \frac{t^2}{2n} - \frac{t^2}{2n}\varepsilon_2\Big(\frac{t}{\sqrt{n}}\Big)\right]\right) \sim \exp\left[-\frac{t^2}{2} - \frac{t^2}{2}\varepsilon_2\Big(\frac{t}{\sqrt{n}}\Big)\right] \to_n e^{-\frac{t^2}{2}}.$$

□

Teorema 3.13. ***Desigualdad de Berry y Esseen*** *Si $X_1, \dots, X_n$ son variables i.i.d con $\mathbb{E}(X_1) = 0$, $Var(X_1) = \sigma^2 < \infty$ y $\mathbb{E}(|X_1|^3) < \infty$. Denotemos $F_n = P(S_n \leq x)$ siendo $S_n = \frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}}$, entonces*

$$\sup_{x \in \mathbb{R}} \big|F_n(x) - \Phi(x)\big| \leq C\frac{\mathbb{E}(|X_1|^3)}{\sigma^3\sqrt{n}},$$

donde $\frac{1}{\sqrt{2\pi}} \leq C < 0{,}8$.

3.2.2. Arreglos triangulares

Supongamos que para cada $n \geq 1$ tenemos una sucesión $X_{n1}, \dots, X_{nn}$ de variables independientes, tales que

$$\mathbb{E}(X_{nk}) = 0, \qquad \mathbb{E}(X_{nk}^2) = \sigma_{nk}^2 > 0, \qquad \sum_{k=1}^{n} \sigma_{nk}^2 = 1,$$

denotamos

$$S_n = X_{n1} + \cdots + X_{nn} \qquad F_{nk}(x) = \mathbb{P}(X_{nk} \leq x)$$

$$\Phi(x) = \mathbb{P}\big(N(0,1) \leq x\big) \qquad \Phi_{nk}(x) = \mathbb{P}\big(N(0, \sigma_{nk}^2) \leq x\big) = \Phi(x/\sigma_{nk})$$

El objetivo de esta sección es probar:

Teorema 3.14. ***(Lindeberg.)*** *Una condición necesaria y suficiente para que*

$$S_n \xrightarrow{d} N(0,1)$$

es

$$(\Lambda) \qquad \forall \varepsilon > 0 \qquad \sum_{k=1}^{n} \int_{|x|>\varepsilon} |x||F_{nk}(x) - \Phi_{nk}(x)|dx \to 0 \quad \text{cuando } n \to \infty \tag{3.9}$$

Para eso vamos a usar dos lemas, el primero se deja como ejercicio:

Lema 3.15. *Vale la desigualdad*

$$\left| e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!} \right| \leq \frac{1}{n!}|x|^n$$

Lema 3.16. *Si $\mathbb{E}(|X|^n) < \infty$, entonces $x^n(F(x) - F(-x) - 1) \to 0$ siendo F la función de distribución de X*

Demostración. Observemos que $x^n(1 - F(x) + F(-x)) \leq x^n \mathbb{P}(|X| \geq x)$, veamos que $x^n \mathbb{P}(|X| \geq x) \to 0$. Como

$$\mathbb{E}|X|^n = \sum_{k=1}^{\infty} \int_{k-1}^{k} |x|^n dF(x) < \infty,$$

por lo tanto

$$\sum_{k \geq x+1} \int_{k-1}^{k} |x|^n dF(x) \to 0 \text{ cuando } n \to +\infty,$$

por otra parte

$$\sum_{k \geq x+1} \int_{k-1}^{k} |x|^n dF(x) \geq x^n \mathbb{P}(|X| \geq x).$$

□

Demostración del Teorema
Veremos únicamente que la condición (Λ) es suficiente, denotamos

$$f_{nk}(t) = \mathbb{E}(e^{itX_{nk}}), \qquad f_n(t) = \mathbb{E}(e^{itS_n}),$$

$$\varphi_{nk}(t) = \int_{-\infty}^{+\infty} e^{itx} d\Phi_{nk}(x), \qquad \varphi(t) = \int_{-\infty}^{+\infty} e^{itx} d\Phi(x).$$

Tenemos que probar que, para todo t, $f_n(t) \to \varphi(t)$ cuando $n \to \infty$. Observemos que, como las X_{ik} son independientes

$$f_n(t) = \prod_{k=1}^{n} f_{nk}(t),$$

por otra parte como $\sum_{k=1}^{n} \sigma_{nk}^2 = 1$, si $Z \sim N(0,1)$, $Z \stackrel{d}{=} Z_1 + \cdots + Z_n$ con $Z_i \sim N(0, \sigma_{in}^2)$ para todo $i = 1, \ldots, n$, e independientes, por lo tanto

$$\varphi(t) = \prod_{k=1}^{n} \varphi_{nk}(t),$$

es decir

$$|f_n(t) - \varphi(t)| = \left| \prod_{k=1}^{n} f_{nk}(t) - \prod_{k=1}^{n} \varphi_{nk}(t) \right|.$$

Observemos que, para todo $k = 1, \ldots, n$, $|f_{nk}(t)| \leq 1$ y $|\varphi_{nk}(t)| \leq 1$, vamos a usar la siguiente desigualdad para números reales: para todo x_i, y_i, $|x_i| \leq 1$ y $|y_i| \leq 1$, se tiene

$$|x_1 x_2 \ldots x_n - y_1 y_2 \ldots y_n| \leq \sum_{k=1}^{n} |x_k - y_k|,$$

dicha desigualdad se prueba por inducción. Si aplicamos esta desigualdad obtenemos que

$$\begin{aligned} |f_n(t) - \varphi(t)| &\leq \sum_{k=1}^{n} |f_{nk}(t) - \varphi_{nk}(t)| && (3.10)\\ &= \sum_{k=1}^{n} \left| \int_{-\infty}^{+\infty} e^{itx} dF_{nk}(x) - \int_{-\infty}^{+\infty} e^{itx} d\Phi_{nk}(x) \right| && (3.11)\\ &= \sum_{k=1}^{n} \left| \int_{-\infty}^{+\infty} e^{itx} d(F_{nk} - \Phi_{nk})(x) \right| && (3.12) \end{aligned}$$

donde en la última igualdad hemos usado la Proposición 1.37. Observemos que

$$0 = \int_{-\infty}^{+\infty} x dF_{nk}(x) = \int_{-\infty}^{+\infty} x d\Phi_{nk}(x) \quad y \quad \sigma_{nk}^2 = \int_{-\infty}^{+\infty} x^2 dF_{nk} = \int_{-\infty}^{+\infty} x^2 d\Phi_{nk},$$

por lo tanto, para todo $k = 1, \ldots, n$

$$\int_{-\infty}^{+\infty} e^{itx} d(F_{nk} - \Phi_{nk})(x) = \int_{-\infty}^{+\infty} \Big(e^{itx} - itx + \frac{1}{2} t^2 x^2\Big) d(F_{nk} - \Phi_{nk})(x),$$

si aplicamos la fórmula de integración por partes 1.42, obtenemos

$$\begin{aligned} \int_{-\infty}^{+\infty} \Big(e^{itx} - itx + \frac{1}{2} t^2 x^2\Big) d(F_{nk} - \Phi_{nk})(x) &= \Big(e^{itx} - itx + \frac{1}{2} t^2 x^2\Big) (F_{nk}(x) - \Phi_{nk}(x)) \Big|_{-\infty}^{+\infty} \\ &\quad - \int_{-\infty}^{+\infty} (ite^{itx} - it + t^2 x) \big(F_{nk}(x) - \Phi_{nk}(x)\big) dx \quad (3.13) \end{aligned}$$

Veamos que $\Big(e^{itx} - itx + \frac{1}{2}t^2x^2\Big)(F_{nk}(x) - \Phi_{nk}(x))\Big|_{-\infty}^{+\infty} = 0$,

$$\Big(e^{itx} - itx + \frac{1}{2}t^2x^2\Big)(F_{nk}(x) - \Phi_{nk}(x))\Big|_{-\infty}^{+\infty} = \frac{t^2}{2}\lim_{a\to+\infty} x^2\left[1 + \left(\frac{2e^{itx}}{t^2x^2} - \frac{i}{tx}\right)\right](F_{nk}(x) - \Phi_{nk}(x))\Big|_{-a}^{a},$$

el límite anterior es

$$\lim_{a\to+\infty} a^2\left[1 + \left(\frac{2e^{ita}}{t^2a^2} - \frac{i}{ta}\right)\right](F_{nk}(a) - \Phi_{nk}(a)) - a^2\left[1 + \left(\frac{2e^{-ita}}{t^2a^2} + \frac{i}{ta}\right)\right](F_{nk}(-a) - \Phi_{nk}(-a)),$$

este límite es igual a:

$$\lim_{a\to+\infty} a^2\Big[(F_{nk}(a) - F_{nk}(-a) - 1) - (\Phi_{nk}(a) - \Phi_{nk}(-a) - 1)\Big] = 0,$$

ya que, por el Lema Lema 3.16, $\lim_{a\to+\infty} a^2(F_{nk}(a) - F_{nk}(-a) - 1) = 0 = \lim_{a\to+\infty} a^2(\Phi_{nk}(a) - \Phi_{nk}(-a) - 1)$. Volviendo a (3.13)

$$\begin{aligned}\int_{-\infty}^{+\infty}\Big(e^{itx} - itx + \frac{1}{2}t^2x^2\Big)d(F_{nk} - \Phi_{nk})(x) &= -\int_{-\infty}^{+\infty}(ite^{itx} - it + t^2x)\big(F_{nk}(x) - \Phi_{nk}(x)\big)dx\\ &= -\,it\int_{-\infty}^{+\infty}(e^{itx} - 1 - itx)\big(F_{nk}(x) - \Phi_{nk}(x)\big)dx,\end{aligned}$$

de donde

$$\left|\int_{-\infty}^{+\infty} e^{itx}d(F_{nk} - \Phi_{nk})(x)\right| = \left|t\int_{-\infty}^{+\infty}(e^{itx} - 1 - itx)\big(F_{nk}(x) - \Phi_{nk}(x)\big)dx\right| \tag{3.14}$$

$$\leq |t|\int_{-\infty}^{+\infty}\left|e^{itx} - 1 - itx\right|\left|F_{nk}(x) - \Phi_{nk}(x)\right|dx \tag{3.15}$$

$$\begin{aligned}\int_{-\infty}^{+\infty}\left|e^{itx} - 1 - itx\right|\left|F_{nk}(x) - \Phi_{nk}(x)\right|dx = &\int_{|x|\leq\varepsilon}\left|e^{itx} - 1 - itx\right|\left|F_{nk}(x) - \Phi_{nk}(x)\right|dx +\\ &\int_{|x|>\varepsilon}\left|e^{itx} - 1 - itx\right|\left|F_{nk}(x) - \Phi_{nk}(x)\right|dx =: I_1 + I_2\end{aligned}$$

Para acotar I_1 usamos que $\left|e^{itx} - 1 - itx\right| \leq \frac{1}{2}t^2|x|^2$, que se sigue del Lema 3.15 y acotamos $|x|^2 \leq \varepsilon|x|$ si $|x| \leq \varepsilon$,

$$\int_{|x|\leq\varepsilon}\left|e^{itx} - 1 - itx\right|\left|F_{nk}(x) - \Phi_{nk}(x)\right|dx \leq \frac{1}{2}|t|^2\varepsilon\int_{|x|\leq\varepsilon}|x|\left|F_{nk}(x) - \Phi_{nk}(x)\right|dx,$$

por otra parte, integrando por partes,

$$\int_{|x|\leq\varepsilon}|x|\left|F_{nk}(x) - \Phi_{nk}(x)\right|dx \leq \int_{\mathbb{R}}|x|\left|F_{nk}(x) - \Phi_{nk}(x)\right|dx \leq 2\sigma_{nk}^2.$$

Para acotar I_2 usamos nuevamente el Lema 3.15 acotando $|e^{itx} - 1| \leq t|x|$, y $|itx| \leq |t||x|$, y obtenemos

$$\int_{|x|>\varepsilon}\left|e^{itx} - 1 - itx\right|\left|F_{nk}(x) - \Phi_{nk}(x)\right|dx \leq 2|t|^2\int_{|x|>\varepsilon}|x|\left|F_{nk}(x) - \Phi_{nk}(x)\right|dx$$

Volviendo a (3.14)

$$\left|\int_{-\infty}^{+\infty} e^{itx} d(F_{nk}-\Phi_{nk})(x)\right| \leq \frac{1}{2}|t|^3\varepsilon\sigma_{nk}^2 + 2|t|^2 \int_{|x|>\varepsilon} |x|\,|F_{nk}(x)-\Phi_{nk}(x)|\,dx,$$

si sumamos en k y usamos que $\sum_{k=1}^{n}\sigma_{nk}^2 = 1$ obtenemos que

$$\sum_{k=1}^{n}\left|\int_{-\infty}^{+\infty} e^{itx} d(F_{nk}-\Phi_{nk})(x)\right| \leq \frac{1}{2}|t|^3\varepsilon + 2|t|^2 \sum_{k=1}^{n}\int_{|x|>\varepsilon} |x|\,|F_{nk}(x)-\Phi_{nk}(x)|\,dx,$$

tomando límite en $n\to\infty$ y usando (Λ) se sigue la prueba.

Ejercicio 3.17. Probar que la condición (3.16) implica que $\max_{k=1,\dots,n}\frac{\sigma_{nk}^2}{V_n^2}\to 0$.

Teorema 3.18. ***(Lindeberg.)*** *Si $X_{n1},\dots,X_{nn}$ son variables aleatorias independientes tal que $\mathbb{E}(X_{nk}) = m_{nk}$ y $Var(X_{nk}) = \sigma_{nk}^2 < \infty$. Denotemos $S_n = \sum_{k=1}^{n} X_{nk}$ y $V_n^2 = \sum_{k=1}^{n}\sigma_{nk}^2$. Si se cumple la condición de Lindeberg: para todo $\varepsilon > 0$*

$$\lim_{n\to+\infty}\frac{1}{V_n^2}\sum_{k=1}^{n}\mathbb{E}\left((X_{nk}-m_{nk})^2\mathbb{I}_{\{|X_{nk}-m_{nk}|\geq\varepsilon V_n\}}\right) = 0 \tag{3.16}$$

entonces

$$\frac{1}{V_n}\sum_{k=1}^{n}(X_{nk}-m_{nk}) \xrightarrow{d} N(0,1).$$

Demostración. Veamos que a condición (3.16) implica la condición (3.9). Vamos a suponer sin pérdida de generalidad que $V_n = 1$ y $m_{nk} = 0$. Si aplicamos la fórmula (1.6) para el método de integración por cambio de variable para integrales de Rieman-Stieltjes, haciendo el cambio $u = x/\sigma_{nk}$ obtenemos

$$\int_{|x|>\varepsilon} x^2 d\Phi_{nk}(x) \leq \sigma_{nk}^2 \int_{|x|>\varepsilon/\sigma_{nk}} x^2 d\Phi(x) \leq \sigma_{nk}^2 \int_{|x|>\varepsilon/\sqrt{\max_k \sigma_{nk}}} x^2 d\Phi(x)$$

Por el ejercicio (1.46) $\sqrt{\max_k \sigma_{nk}} \to 0$ cuando $n\to\infty$, por lo tanto, si sumamos en k y usamos que $V_n = 1$,

$$\sum_{k=1}^{n}\int_{|x|>\varepsilon} x^2 d\Phi_{nk}(x) \leq \int_{|x|>\varepsilon/\sqrt{\max_k \sigma_{nk}}} x^2 d\Phi(x)\sum_{k=1}^{n}\sigma_{nk}^2 = \int_{|x|>\varepsilon/\sqrt{\max_k \sigma_{nk}}} x^2 d\Phi(x) \to 0,$$

de esto y la condición (3.16) se sigue que, para todo $\varepsilon > 0$,

$$\sum_{k=1}^{n}\int_{|x|>\varepsilon} x^2 d(F_{nk}(x)-\Phi_{nk}(x)) \to 0, \quad \text{cuando } n\to\infty. \tag{3.17}$$

Fijemos $\varepsilon > 0$ y sea h una función par, derivable con derivada continua que satisface que $|h(x)| \leq x^2$, $h'(x)signo(x) \geq 0$, $h(x) = x^2$ para $|x| > 2\varepsilon$, $h(x) = 0$ si $|x| \leq \varepsilon$. De (3.17) obtenemos que

$$\sum_{k=1}^{n}\int_{|x|>\varepsilon} h(x) d(F_{nk}(x)-\Phi_{nk}(x)) \to 0, \quad \text{cuando } n\to\infty.$$

Observemos que si aplicamos el método de integración por partes (usando que, $h(\varepsilon) = 0$, $x^2(1-F_{nk}(x))\to 0$ y $x^2(1-\Phi_{nk}(x))\to 0$ cuando $x\to\infty$ obtenemos que

$$\int_{x>\varepsilon} h'(x)\big((1-F_{nk}(x)) + (1-\Phi_{nk}(x)\big)dx = \int_{x>\varepsilon} h(x) d(F_{nk}(x)+\Phi_{nk}(x))$$

y que

$$\int_{x\leq -\varepsilon} h'(x)\big(F_{nk}(x)+\Phi_{nk}(x)\big)dx = -\int_{x\leq -\varepsilon} h(x)d(F_{nk}(x)+\Phi_{nk}(x))$$

Finalmente, como $h'(x)=x/2$ si $|x|>2\varepsilon$, acotando para $x\geq 2\varepsilon$, $|F_{nk}(x)-\Phi_{nk}(x)|\leq 1-F_{nk}(x)+1-\Phi_{nk}(x)$, y para $x\leq -2\varepsilon$ $|F_{nk}(x)-\Phi(x)|\leq F_{nk}(x)+\Phi(x)$ y sumando en k, obtenemos que

$$\sum_{k=1}^{n}\int_{|x|\geq 2\varepsilon}|x||F_{nk}(x)-\Phi_{nk}(x)|dx\to 0, \quad n\to+\infty,$$

como ε es arbitrario se sigue (3.9). □

Observemos que en el Teorema anterior pedimos que haya independencia entre las $X_{n1},\dots,X_{nn}$ pero no necesariamente se requiere entre X_{ji} y X_{kl} con $k\neq j$. Teniendo esto en cuenta probar que el Teorema 3.12 es un caso particular del Teorema 3.18. Es decir, se cumple la condiccion de Lindeberg, si las variables son iid con varianza finita. Una condición suficiente para que se verifique la condición de Lindeberg es la condición de Lyapunov, que en general es más fácil de verificar.

Definición 3.19. Sean $X_1,\dots,X_n$ son variables independientes con segundo momento finito, sea $m_k=\mathbb{E}(X_k)$, $\sigma_k^2=Var(X_k)$, $S_n=X_1+\cdots+X_n$ y $V_n^2=\sum_{k=1}^n\sigma_k^2$. Las variables X_i verifican la condición de Lyapunov si existe $\delta>0$ tal que

$$\lim_{n\to+\infty}\frac{1}{V_n^{2+\delta}}\sum_{k=1}^{n}\mathbb{E}|X_k-m_k|^{2+\delta}=0. \tag{3.18}$$

Ejercicio 3.20. La condición (3.18) implica la Condición (3.16).

Ejercicio 3.21. *Método delta.* Probar que si Y_n es una sucesión de variables aleatorias tal que $\sqrt{n}(Y_n-\mu)\to N(0,\sigma^2)$, con $\mu\in\mathbb{R}$ y f es una función derivable en μ con derivada acotada en un entorno de μ entonces $\sqrt{n}(f(Y_n)-f(\mu))\to N(0,\sigma^2 f'(\mu)^2)$. Sugerencia: hacer un desarrollo de Taylor de f en un entorno de μ y usar el Lema de Slutsky.

Capítulo 4

Sucesiones estacionarias y teoría ergódica

En este capítulo veremos que es posible debilitar un poco la hipótesis de independencia en la ley fuerte de los grandes números y en el teorema central del límite usando técnicas de teoría ergódica.

4.1. Sucesiones estacionarias (en sentido estricto) de variables aleatorias

Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y una sucesión $X = \{X_n\}_{n \in \mathbb{N}}$ de variables aleatorias, dado $k > 0$ se define el operador *shift* entre sucesiones θ_k, de la siguiente manera

$$\theta_k X = \{X_n\}_{n>k} = \{X_{k+1}, X_{k+2}, \dots\}.$$

Definición 4.1. Una sucesión de variables aleatorias X es estacionaria (en sentido estricto) si la distribución de probabilidad de X es igual a la de $\theta_k X$ para todo $k \geq 1$:

$$P\big((X_1, X_2, \dots) \in B\big) = P\big((X_{k+1}, X_{k+2}, \dots) \in B\big) \quad \forall B \in \mathcal{B}(\mathbb{R}^\infty)$$

Como ejemplo trivial de este tipo de sucesiones tenemos las sucesiones de variables i.i.d. En este caso sabemos por la ley fuerte de Kolmogorov que si $\mathbb{E}|X_1| < \infty$,

$$\frac{X_1 + \dots + X_n}{n} \to \mathbb{E}(X_1),$$

veremos que dicho resultado se puede generalizar para el caso en que la sucesión X es estacionaria. Para eso vamos a introducir y estudiar algunos conceptos de teoría ergódica.

Definición 4.2. Una función $T : \Omega \to \Omega$ es medible si para todo $A \in A$

$$T^{-1}(A) = \{\omega : T(\omega) \in A\} \in \mathcal{A}.$$

Una función medible preserva la medida si para todo $A \in \mathcal{A}$, $P(A) = P(T^{-1}(A))$.

Ejercicio 4.3. Probar que si T preserva la medida $\mathbb{E}(X) = E(X(T))$

Proposición 4.4. *Sea T una función que preserva la medida y X_1 una variable aleatoria, si definimos la variable X_k tal que $X_k(\omega) = X_1(T^{k-1}(\omega))$ entonces la sucesión $X = \{X_1, X_2, \dots\}$ es estacionaria.*

Demostración. Sean

$$\begin{aligned} A =&\{\omega : X(\omega) \in B\} = \Big\{\omega : \big(X_1(\omega), X_1(T(\omega)), X_1(T^2(\omega)), \dots\big) \in B\Big\} \\ A_1 =&\{\omega : \theta_1 X(\omega) \in B\} = \Big\{\omega : \big(X_1(T(\omega)), X_1(T^2(\omega)), X_1(T^3(\omega)) \dots\big) \in B\Big\} \end{aligned}$$

Es claro que $\omega \in A_1$ si y sólo si $T(\omega) \in A$, o lo que es lo mismo $A_1 = T^{-1}(A)$. Como $P(T^{-1}(A)) = P(A)$ ya que T preserva la medida, obtenemos que $P(A_1) = P(A)$. De igual manera se llega a que para todo $k > 1$, $P(A_k) = P(A)$ siendo $A_k = \{\omega : \theta_k X \in B\}$. □

El recíproco de la proposición anterior es cierta en el siguiente sentido:

Proposición 4.5. *Dada una sucesión estacionaria $X = \{X_n\}_{n\in\mathbb{N}}$ definida en $(\Omega, \mathcal{A}, P)$ podemos construir un nuevo espacio de probabilidad $(\tilde\Omega, \tilde{\mathcal{A}}, \tilde P)$, una variable aleatoria $\tilde X_1$ y una transformación $\tilde T$ que preserva la medida tal que la distribución de $\tilde X = \{\tilde X_1, \tilde X_1(\tilde T), \tilde X_1(\tilde T^2), \dots\}$ es igual a la de X.*

Demostración. Tomemos $\tilde\Omega = \mathbb{R}^\infty$, $\tilde{\mathcal{A}} = \mathcal{B}(\mathbb{R}^\infty)$ y $\tilde P = P_X$, donde $P_X(B) = P(\omega : X(\omega) \in B)$, $B \in \mathcal{B}(\mathbb{R}^\infty)$. La función $\tilde T$ es el shift $\tilde T(x_1, x_2, \dots) = (x_2, x_3, \dots)$. Si $\tilde\omega = (x_1, x_2, \dots)$ definimos $\tilde X_1(\tilde\omega) = x_1$, y $\tilde X_n(\tilde\omega) = \tilde X_1(\tilde T^{n-1}(\tilde\omega)) = x_n$ para $n > 1$. Veamos que $\tilde T$ preserva la medida. Sea $A = \{\tilde\omega : (x_1, x_2, \dots, x_k) \in B\}$, con $B \in \mathcal{B}(\mathbb{R}^k)$, entonces

$$\tilde T^{-1}(A) = \big\{\tilde\omega : (x_2, x_3, \dots, x_{k+1}) \in B\big\}.$$

Como X es estacionario

$$\tilde P(A) = P\{\omega : (X_1, \dots, X_k) \in B\} = P\{\omega : (X_2, \dots, X_{k+1}) \in B\} = \tilde P(\tilde T^{-1}(A)).$$

Que X y $\tilde X$ tienen la misma distribución se sigue de que

$$\tilde P\big\{\tilde\omega : (\tilde X_1, \dots, \tilde X_k) \in B\big\} = \tilde P(B) = P\big\{\omega : (X_1, \dots, X_k) \in B\big\}.$$

□

Teorema 4.6. *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, sea T una función que preserva la medida, dado $A \in \mathcal{A}$ existe $D \subset A$ tal que $P(D) = P(A)$ y $\forall\omega \in D$, $T^n(\omega) \in A$ para infinitos n.*

Demostración. Sea $C = \{\omega \in A : T^n(\omega) \notin A,\ \forall n \geq 1\}$. Como $C \cap T^{-n}(C) = \emptyset$ para todo $n \geq 1$, obtenemos que $\emptyset = T^{-m}\big(C \cap T^{-n}(C)\big) = T^{-m}(C) \cap T^{-(m+n)}(C)$. Por lo tanto la sucesión $\{T^{-n}(C)\}_{n\geq 0}$ está formada por conjuntos disjuntos, por lo tanto $\sum_{n=0}^\infty P(T^{-n}(C)) \leq P(\omega) = 1$ por otra parte como T preserva la medida $\sum_{n=0}^\infty P(C) = \sum_{n=0}^\infty P(T^{-n}(C))$ es decir $\sum_{n=0}^\infty P(C) \leq 1$ y por lo tanto $P(C) = 0$. Si ahora en lugar de T, tomamos T^k (que tambien preserva la medida) obtenemos que $P(C_k) = 0$ siendo $C_k = \{\omega \in A : (T^k)^n \notin A\}$. Tomamos $D = \cup_k C_k$. □

Dejamos como ejercicio el siguiente corolario

Corolario 4.7. *Sea X una variable aleatoria no negativa, existe $D \subset \{\omega : X(\omega) > 0\}$ con $P(D) = P(\{\omega : X(\omega) > 0\})$ tal que para todo $\omega \in D$*

$$\sum_{k=0}^\infty X(T^k(\omega)) = \infty$$

4.2. Ergodicidad y mixing

Definición 4.8. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, y T una transformación que preserva la medida

1. Un conjunto $A \in \mathcal{A}$ es invariante si $T^{-1}(A) = A$.
2. T es ergódica si todo conjunto invariante tiene medida 0 o 1.
3. Una variable aleatoria X es invariante si $X(\omega) = X(T(\omega))$ para todo $\omega \in \Omega$.

Definición 4.9. Una función T que preserva la medida es mixing si para todo $A, B \in \mathcal{A}$

$$\lim_{n\to\infty} P\big(A \cap T^{-n}(B)\big) = P(A)P(B).$$

En el contexto de sucesiones estacionarias en sentido estricto, se define el concepto de α-mixing de la siguiente manera:

Definición 4.10. Sea $X = X_1, X_2, \dots,$ una sucesión estacionaria en sentido estricto, denotemos $\mathcal{A}_n^\infty = \sigma(X_n, X_{n+1}, \dots)$, $\mathcal{A}_1^n = \sigma(X_1, \dots, X_n)$ y

$$\alpha_n = \sup_{\substack{A \in \mathcal{A}_1 \\ B \in \mathcal{A}_n^\infty}} |P(A \cap B) - P(A)P(B)|,$$

decimos que X es α-mixing si $\alpha_n \to_n 0$

Ejercicio 4.11. Toda transformación mixing es ergódica.

Para demostrar el resultado principal de éste capítulo vamos a necesitar el siguiente lema:

Lema 4.12. *(**Teorema Ergódico Maximal.**) Sea T una transformación que preserva la medida y X una variable aleatoria tal que $\mathbb{E}|X| < \infty$. Sean*

$$\begin{aligned} S_k(\omega) =& X(\omega) + X(T(\omega)) + \cdots + X(T^{k-1}(\omega)) \\ M_k(\omega) =& \max\{0, S_1(\omega), \dots, S_k(\omega)\}. \end{aligned}$$

entonces

$$\mathbb{E}(X\mathbb{I}_{\{M_n>0\}}) \geq 0.$$

Demostración. Si $n \geq k$, como $M_n(T(\omega)) \geq S_k(T(\omega))$ se tiene que $X(\omega)+M_n(T(\omega)) \geq X(\omega)+S_k(T(\omega)) = S_{k+1}(\omega)$, como $X(\omega) \geq S_1(\omega) - M_n(T(\omega))$ obtenemos que

$$X(\omega) \geq \max\{S_1(\omega), \dots, S_n(\omega)\} - M_n(T(\omega)).$$

Por lo tanto

$$X(\omega)\mathbb{I}_{\{M_n>0\}} \geq \max\{S_1(\omega), \dots, S_n(\omega)\}\mathbb{I}_{\{M_n>0\}} - M_n(T(\omega))\mathbb{I}_{\{M_n>0\}},$$

de donde tomando esperanza:

$$\mathbb{E}\big(X(\omega)\mathbb{I}_{\{M_n>0\}}\big) \geq \mathbb{E}\Big(\big(\max\{S_1(\omega), \dots, S_n(\omega)\} - M_n(T(\omega))\big)\mathbb{I}_{\{M_n>0\}}\Big),$$

Observemos que en $\{M_n > 0\}$, $\max\{S_1(\omega), \dots, S_n(\omega)\} = M_n(\omega)$, es decir $\mathbb{E}(X(\omega)\mathbb{I}_{\{M_n>0\}}) \geq \mathbb{E}\big(M_n - M_n(T)\big)$, como T preserva la medida, por 4.3, $\mathbb{E}(M_n) = \mathbb{E}(M_n(T))$. □

Teorema 4.13. ***(Birkhoff y Khinchin.)*** *Sea T una transformación ergódica y X una variable aleatoria tal que $\mathbb{E}|X| < \infty$, entonces*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} X(T^k(\omega)) = \mathbb{E}(X).$$

Demostración. Vamos a suponer, sin perdida de generalidad, que $\mathbb{E}(X) = 0$. Denotemos $\overline{\eta} = \limsup S_n/n$ y $\underline{\eta} = \liminf S_n/n$. Veremos que,

$$0 \leq \underline{\eta} \leq \overline{\eta} \leq 0 \text{ en un conjunto de probabilidad 1.}$$

Observemos que como $\overline{\eta}(\omega) = \overline{\eta}(T(\omega))$ la variable $\overline{\eta}$ es invariante y por lo tanto para todo $\varepsilon > 0$ el conjunto $A_\varepsilon = \{\overline{\eta} > \varepsilon\}$ es invariante, por lo tanto $P(A_\varepsilon)$ es 0 o 1. Denotemos

$$\begin{aligned} X^*(\omega) &= (X - \varepsilon)\mathbb{I}_{A_\varepsilon} \\ S_k^* &= X + \cdots + X(T^{k-1}) \\ M_k^* &= \max\{0, S_1^*, \ldots, S_k^*\} \end{aligned}$$

Por el lema anterior $\mathbb{E}\big(X^*\mathbb{I}_{\{M_n^*>0\}}\big) \geq 0$ para todo $n \geq 0$.

$$\{M_n^* > 0\} = \left\{\max_{1\leq k\leq n} S_k^*\right\} \uparrow \left\{\sup_{k\geq 1} S_k^* > 0\right\} = \left\{\sup_{k\geq 1} \frac{S_k^*}{k} > 0\right\}.$$

Observemos que para todo $k > 0$ se tiene que $\left\{\frac{S_k^*}{k} > 0\right\} = \left\{\frac{S_k}{k} > \varepsilon\right\} \cap A_\varepsilon$, de donde

$$\left\{\sup_{k\geq 1} \frac{S_k^*}{k} > 0\right\} = \left\{\sup_{k\geq 1} \frac{S_k}{k} > \varepsilon\right\} \cap A_\varepsilon.$$

Recordemos que por definición $\overline{\eta} = \limsup(S_n/n) = \inf_{k\geq 1}\sup_{n\geq k}(S_n/n) \leq \sup_{k\geq 1} S_k/k$, de esto se deduce que $\left\{\sup_{k\geq 1}\frac{S_k}{k} > \varepsilon\right\} \cap A_\varepsilon = A_\varepsilon$. Como $\mathbb{E}|X^*| \leq \mathbb{E}|X| + \varepsilon$ podemos aplicar convergencia dominada en $\mathbb{E}(X^*\mathbb{I}_{\{M_n^*\}})$ y obtenemos que

$$0 \leq \mathbb{E}(X^*\mathbb{I}_{\{M_n^*\}}) \to \mathbb{E}(X^*\mathbb{I}_{A_\varepsilon}).$$

Por otra parte $\mathbb{E}(X^*\mathbb{I}_{A_\varepsilon}) = \mathbb{E}((X-\varepsilon)\mathbb{I}_{A_\varepsilon}) = \mathbb{E}(X\mathbb{I}_{A_\varepsilon}) - \varepsilon P(A_\varepsilon)$, observemos que como $P(A_\varepsilon) = 0$ o 1 y $\mathbb{E}(X) = 0$ obtenemos que para todo $\varepsilon > 0$ $P(A_\varepsilon) = 0$, es decir $P(\overline{\eta} \leq 0) = 1$. De igual manera, si consideramos $-X$ en lugar de X obtenemos que

$$\limsup\left(\frac{S_n}{n}\right) = -\liminf\frac{S_n}{n} = -\underline{\eta},$$

y razonando como antes llegamos a que $P(-\underline{\eta} \leq 0) = 1$ es decir $P(\underline{\eta} \geq 0) = 1$. □

Veamos para terminar el enunciado del TCL para procesos estacionarios en sentido estricto, con una condición de mixing.

Teorema 4.14. ***(Bradley 1985.)*** *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión estacionaria en sentido estricto, de variables aleatorias centradas. Supongamos que $\mathbb{E}(X_n^2) = \sigma_n < \infty$ para todo n, $\sigma_n \to \infty$, y existe $\delta > 0$ tal que*

$$\mathbb{E}(|X_1|^{2+\delta}) < \infty \quad y \quad \sum_{n=1}^{\infty} \alpha_n^{\delta/(2+\delta)} < \infty,$$

entonces $\sum_{k=2}^{\infty} |\mathbb{E}(X_1 X_k)| < \infty$, *si*

$$\sigma^2 = \mathbb{E}(X_1^2) + 2 \sum_{k=2}^{\infty} \mathbb{E}(X_1 X_k) > 0$$

entonces

$$n^{-1/2} \sum_{n=1}^{\infty} X_n \xrightarrow{d} N(0, \sigma^2)$$

Bibliografía

[1] P. Billingsley. *Convergence of Probability Measures. 2nd ed.* John Wiley & Sons., 1999.

[2] G. Folland. *Real Analysis: Modern techniques and their applications.* Wiley.

[3] P. Huber and E. Ronchetti. Wiley series in probability and statistics, 2009.

[4] E. Lehmann. *Elements of Large-Sample Theory.* Springer Text in Statistics, Springer., 1999.

[5] A.N. Shiryaev. *Probability.* Springer-Verlag, 1984.