Las siguientes son notas del curso de Introducción a la Probabilidad y Estadística que he dictado en 2010, para licenciatura en matemática. Por corrección de erratas y comentarios, se agradece comunicarse a mi correo electrónico: jkalem@cmat.edu.uy

Juan Kalemkerian

Índice general

Capítulo 1

Espacio de probabilidad.

1.1. σ-álgebra de conjuntos.

Definición 1.1. σ -álgebra de subconjuntos de Ω.
Dado un conjunto $\Omega \neq \Phi$, diremos que $\mathcal{A} \subset 2^{\Omega}$ es una σ-álgebra de subconjuntos de Ω si cumple los siguientes axiomas:
i) $\Omega \in \mathcal{A}$.
ii) Si $A \in \mathcal{A}$ entonces $A^c \in \mathcal{A}$.
iii) Si $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$, entonces $\cup_{n=1}^{+\infty} A_n \in \mathcal{A}$.

En todos los teoremas que siguen a continuación se considera dada $\mathcal{A}$ una σ-álgebra de subconjuntos de Ω.

Teorema 1.2.

$$\Phi \in \mathcal{A}.$$

Demostración.
Como $\Omega \in \mathcal{A}$ entonces por ii) $\Phi = \Omega^c \in \mathcal{A}.\checkmark$

Teorema 1.3. $A_1, A_2, ..., A_n \in \mathcal{A}$ *entonces* $\cup_{i=1}^{n} A_i \in \mathcal{A}$.

Demostración.
Basta usar el axioma iii) en el caso en que $A_{n+1} = A_{n+2} = ... = \Phi \in \mathcal{A}$, entonces en este caso se tiene que $\cup_{n=1}^{+\infty} A_n = \cup_{i=1}^{n} A_i \in \mathcal{A}.\checkmark$

Teorema 1.4. *Si* $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$, *entonces* $\cap_{n=1}^{+\infty} A_n \in \mathcal{A}$.

Demostración.
Como $A_n \in \mathcal{A}$ cualquiera sea n, entonces por ii) $A_n^c \in \mathcal{A}$ para todo n. Entonces por iii) $\cup_{n=1}^{+\infty} A_n^c \in \mathcal{A}$, y por lo tanto $\cap_{n=1}^{+\infty} A_n = \left(\cup_{n=1}^{+\infty} A_n^c\right)^c \in \mathcal{A}.\checkmark$

Teorema 1.5. *Si* $A, B \in \mathcal{A}$, *entonces* $A - B \in \mathcal{A}$.

Demostración.
Basta observar que $A - B = A \cap B^c \in \mathcal{A}$ ya que $A, B^c \in \mathcal{A}$, e intersección finita de elementos de $\mathcal{A}$, pertenece a $\mathcal{A}$.✓

Teorema 1.6. *Si $\mathcal{A}_\alpha$ es σ-álgebra de conjuntos sobre Ω para todo $\alpha \in I$, siendo I una familia cualquiera de índices, entonces $\cap_{\alpha\in I}\mathcal{A}_\alpha$ es σ-álgebra de conjuntos sobre Ω.*

Demostración.
Defino $\mathcal{A} = \cap_{\alpha\in I}\mathcal{A}_\alpha$.
i) $\Omega \in \mathcal{A}_\alpha$ para todo $\alpha \in I$, entonces $\Omega \in \mathcal{A}$.
ii) Si $A \in \mathcal{A}$, entonces $A \in \mathcal{A}_\alpha$ para todo $\alpha \in I$, entonces $A^c \in \mathcal{A}_\alpha$ para todo $\alpha \in I$, luego $A^c \in \mathcal{A}$.
iii) Si $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$, entonces $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}_\alpha$ para todo $\alpha \in I$, entonces $\cup_{n=1}^{+\infty}A_n \in \mathcal{A}_\alpha$ para todo $\alpha \in I$, entonces $\cup_{n=1}^{+\infty}A_n \in \mathcal{A}$.✓

Ejemplo 1.7. $\{\Phi, \Omega\}$ es σ-álgebra de conjuntos sobre Ω, cualquiera sea el conjunto Ω.

Ejemplo 1.8. 2^Ω es σ-álgebra de conjuntos sobre Ω, cualquiera sea el conjunto Ω.

Ejemplo 1.9. Si A es tal que $\Phi \subsetneq A \subsetneq \Omega$, entonces $\{\Phi, \Omega, A, A^c\}$ es σ-álgebra de conjuntos sobre Ω, cualquiera sea el conjunto Ω.

Definición 1.10. σ-álgebra generada por una familia de subconjuntos de Ω. Dada $\mathcal{F}$ una familia de subconjuntos de Ω, al conjunto $\cap_{\mathcal{A}\,:\,\mathcal{A}\supset\mathcal{F}}\mathcal{A}$ le llamaremos σ-álgebra engendrada por $\mathcal{F}$ y la notaremos por $\sigma(\mathcal{F})$.

La σ-álgebra generada por una familia de subconjuntos de Ω, siempre existe y además es la "menor" σ-álgebra generada por una familia de subconjuntos de Ω que contiene a $\mathcal{F}$.

Definición 1.11. σ-álgebra de Borel en $\mathbb{R}$. Consideramos $\mathcal{F} = \{A \subset \mathbb{R} : A \text{ es abierto}\}$. Llamaremos σ-álgebra de Borel en $\mathbb{R}$ a $\sigma(\mathcal{F})$.

Teorema 1.12. *Si definimos $\mathcal{I}_1 = \{(a,b) \subset \mathbb{R} : a < b\}$; $\mathcal{I}_2 = \{[a,b) \subset \mathbb{R} : a < b\}$; $\mathcal{I}_3 = \{(a,b] \subset \mathbb{R} : a < b\}$; $\mathcal{I}_4 = \{(a,+\infty) \subset \mathbb{R} : a \in \mathbb{R}\}$; $\mathcal{I}_5 = \{[a,+\infty) \subset \mathbb{R} : a \in \mathbb{R}\}$; $\mathcal{I}_6 = \{(-\infty,a) \subset \mathbb{R} : a \in \mathbb{R}\}$; $\mathcal{I}_7 = \{(-\infty,a] \subset \mathbb{R} : a \in \mathbb{R}\}$. Entonces*

$$\sigma(\mathcal{I}) = \sigma(\mathcal{I}_1) = \sigma(\mathcal{I}_2) = \sigma(\mathcal{I}_3) = \sigma(\mathcal{I}_4) = \sigma(\mathcal{I}_5) = \sigma(\mathcal{I}_6) = \sigma(\mathcal{I}_7).$$

Demostración.
Probaremos a modo de ejemplo que $\sigma(\mathcal{I}_1) = \sigma(\mathcal{I}_2)$, para lo cual basta ver que $\mathcal{I}_1 \subset \sigma(\mathcal{I}_2)$ y que $\mathcal{I}_2 \subset \sigma(\mathcal{I}_1)$.
Efectivamente, $(a,b) = \bigcup_{n:a+1/n<b}[a+1/n, b)$, lo cual prueba que $\sigma(\mathcal{I}_1) \subset \sigma(\mathcal{I}_2)$. Además, $[a,b) = \bigcap_{n=1}^{+\infty}(a-1/n, b)$, lo cual prueba la otra inclusión.
Se deja como ejercicio verificar las demás igualdades. Para trabajar con $\sigma(\mathcal{I})$, tener en cuenta que todo abierto en $\mathbb{R}$ se puede escribir como una unión numerable de

intervalos abiertos. ✓
De manera similar se define la σ-álgebra de Borel en $\mathbb{R}^k$, como la σ-álgebra generada por los abiertos de $\mathbb{R}^k$, o sea como la menor σ-álgebra que contiene a todos los abiertos de $\mathbb{R}^k$. A los conjuntos de esta σ-álgebra, se les llama borelianos.

1.2. Espacio de probabilidad.

Definición 1.13. Espacio de probabilidad.
Dado $\Omega \neq \Phi$, diremos que la terna $(\Omega, \mathcal{A}, P)$ es un espacio de probabilidad sobre Ω si y sólo $\mathcal{A}$ es una σ-álgebra de conjuntos sobre Ω, y P es una función $P : \mathcal{A} \rightarrow [0, 1]$ que cumple los siguientes axiomas:
i) $P(\Omega) = 1$,
ii) si la familia de sucesos $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ son disjuntos dos a dos ($A_i \cap A_j = \Phi$ para todos $i \neq j$), entonces $P\left(\cup_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} P(A_n)$.

En todos los teoremas que siguen se considera dado el espacio de probabilidad $(\Omega, \mathcal{A}, P)$.

Teorema 1.14.
$$P(\Phi) = 0.$$

Demostración.
Consideramos la familia de sucesos disjuntos $A_1 = \Omega$, $A_2 = A_3 = ... = \Phi$, luego aplicamos el axioma ii) y obtenemos

$$P\left(\cup_{n=1}^{+\infty} A_n\right) = P(\Omega) = P(\Omega) + \sum_{n=2}^{+\infty} P(\Phi)$$

por lo tanto $\sum_{n=2}^{+\infty} P(\Phi) = 0$. Si fuera $P(\Phi) \neq 0$, se tendría que la serie sería divergente y no podría ser cierta la igualdad anterior. Entonces $P(\Phi) = 0$.✓

Teorema 1.15. *Si $A_1, A_2, ..., A_n \in \mathcal{A}$ y son disjuntos dos a dos, entonces* $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.

Demostración.
Se aplica el axioma ii) teniendo en cuenta que si se agregan los conjuntos $A_{n+1} = A_{n+2} = ... = \Phi$, se obtiene que

$$P\left(\cup_{n=1}^{+\infty} A_n\right) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{+\infty} P(A_i) = \sum_{i=1}^{n} P(A_i)$$

pero $P\left(\cup_{n=1}^{+\infty} A_n\right) = P(\cup_{i=1}^{n} A_i)$ de donde se deduce el resultado.✓

Teorema 1.16. *Si $A, B \in \mathcal{A}$, entonces* $P(B - A) = P(B) - P(A \cap B)$.

Demostración.
Escribimos la unión disjunta $(B - A) \cup (A \cap B) = B$. Luego, aplicando el axioma ii) obtenemos que $P(B - A) + P(A \cap B) = P(B)$, de donde se deduce el resultado.✓

Corolario 1.17. *Si $A, B \in \mathcal{A}$ son tales que $A \subset B$, entonces*

1. $P(B - A) = P(B) - P(A)$.
2. $P(A) \leq P(B)$.

Demostración.

1. Es inmediato a partir de la propiedad anterior, si se observa que $A \cap B = A$.✓
2. Es inmediato ya que $P(B) - P(A) = P(B - A) \geq 0$.✓

Teorema 1.18. *Si $A, B \in \mathcal{A}$, entonces $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.*

Demostración.
Escribimos $A \cup B = (A - B) \cup (B - A) \cup (A \cap B)$, unión disjunta, entonces

$$P(A \cup B) = P(A - B) + P(B - A) + P(A \cap B) =$$

$$P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B)$$

de donde se deduce el resultado.✓

Teorema 1.19. *Si $A_1, A_2, ..., A_n \in \mathcal{A}$ entonces*

$$P\left(\cup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} P\left(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}\right).$$

Demostración.
Se deja como ejercicio.

Teorema 1.20. *Si $A_1, A_2, ..., A_n \in \mathcal{A}$, entonces $P\left(\cup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$.*

Demostración.
Se deja como ejercicio.

Teorema 1.21. *Propiedad de continuidad de las probabilidades.*

1. *Si la familia de sucesos $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ es tal que: $A_1 \subset A_2 \subset A_3 \subset \ldots$ entonces*

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

2. *Si la familia de sucesos $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ es tal que: $A_1 \supset A_2 \supset A_3 \supset \ldots$ entonces*

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

Demostración.

1. Definimos la familia de sucesos $B_n = A_n - A_{n-1}$ para $n = 1, 2, 3...$ Sobreentenderemos que $A_0 = \Phi$. Como $A_{n-1} \subset A_n$ cualquiera sea n, entonces $P(A_n - A_{n-1}) = P(A_n) - P(A_{n-1})$. Por otro lado $\{B_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$, es una familia disjunta de sucesos, por lo que aplicando el axioma iii) se obtiene que

$$P\left(\cup_{n=1}^{+\infty} B_n\right) = \sum_{n=1}^{+\infty} P(B_n) = \sum_{n=1}^{+\infty} P(A_n - A_{n-1}) = \sum_{n=1}^{+\infty} [P(A_n) - P(A_{n-1})]$$

$$= \text{lím} P(A_n) .\checkmark$$

2. Tomando complementos obtenemos que $A_1^c \subset A_2^c \subset A_3^c \subset ...$, luego aplicando la parte anterior, se obtiene que $P\left(\cup_{n=1}^{+\infty} A_n^c\right) = \text{lím} P(A_n^c)$. O sea que

$$P\left(\left[\cap_{n=1}^{+\infty} A_n\right]^c\right) = 1 - P\left(\cap_{n=1}^{+\infty} A_n\right) =$$

$$\text{lím}\, [1 - P(A_n)] .$$

Entonces

$$P\left(\cap_{n=1}^{+\infty} A_n\right) = \text{lím} P(A_n) .\checkmark$$

Teorema 1.22. *Si la familia de sucesos $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ es tal que $P(A_n) = 1$ para todo n, entonces $P\left(\cap_{n=1}^{+\infty} A_n\right) = 1$.*

Demostración.
Debemos probar que $P\left(\left[\cap_{n=1}^{+\infty} A_n\right]^c\right) = P\left(\cup_{n=1}^{+\infty} A_n^c\right) = 0$. A partir del teorema 1.20 y tomando límite obtenemos

$$P\left(\cup_{n=1}^{+\infty} A_n^c\right) \leq \sum_{n=1}^{+\infty} P(A_n^c) = 0.\checkmark$$

Definición 1.23. Límites superior e inferior de una sucesión de conjuntos. Dados $(\Omega, \mathcal{A}, P)$ espacio de probabilidad y $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$, se definen el límite superior e inferior de la sucesión de sucesos como

$$\text{limsup } A_n := \bigcap_{n=1}^{+\infty} \bigcup_{k=n}^{+\infty} A_k \text{ y liminf } A_n := \bigcup_{n=1}^{+\infty} \bigcap_{k=n}^{+\infty} A_k.$$

respectivamente.

Se deja como ejercicio verificar las siguientes propiedades.

1. limsup $A_n = \{w \in \Omega : w \in A_n$ para infinitos valores de $n\}$ (ocurren infinitos A_n).

2. liminf $A_n =$
$\{w \in \Omega : w \in A_n$ para todo n, salvo a lo sumo para una cantidad finita de índices$\}$ (ocurren A_n para todos los valores de n salvo a lo sumo 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 $P\left(\text{limsup } A_n\right) = \lim P\left(\bigcup_{k=n}^{+\infty} A_k\right)$.

5. Como la sucesión $B_n = \bigcap_{k=n}^{+\infty} A_k$ es creciente, entonces $P\left(\text{liminf } A_n\right) = \lim P\left(\bigcap_{k=n}^{+\infty} A_k\right)$.

6. Si $\{A_n\}_{n\in\mathbb{N}}$ es una sucesión creciente de sucesos, entonces liminf A_n = limsup $A_n = \bigcup_{n=1}^{+\infty} A_n$.

7. Si $\{A_n\}_{n\in\mathbb{N}}$ es una sucesión decreciente de sucesos, entonces liminf A_n = limsup $A_n = \bigcap_{n=1}^{+\infty} A_n$.

Observación 1.24. *La definición de límite superior e inferior de una familia de conjuntos se define de igual modo aunque no estemos en un espacio de probabilidad.*

Teorema 1.25. *Dados $(\Omega, \mathcal{A}, P)$ espacio de probabilidad y una sucesión $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$, entonces se cumple que*

$$P\left(liminf\ A_n\right) \overset{(1)}{\leq} liminfP\left(\ A_n\right) \overset{(2)}{\leq} limsupP\left(A_n\right) \overset{(3)}{\leq} P\left(limsup\ A_n\right).$$

Demostración.
Para la desigualdad (3), vemos que para todo n se tiene que $\bigcup_{k=n}^{+\infty} A_k \supset A_n$, entonces

$$P\left(\text{limsup } A_n\right) = limP\left(\bigcup_{k=n}^{+\infty} A_k\right) \geq limsupP\left(A_n\right).$$

Un razonamiento análogo prueba la desigualdad (1).
La desigualdad (2) es evidente.✓

Ejemplo 1.26. Si Ω es un conjunto infinito numerable, es decir $\Omega = \{w_1, w_2, ..., w_n, ...\}$ entonces si consideramos la sucesión $\{p_n\}_{n\in\mathbb{N}}$ tal que $p_n \geq 0$ para todo $n \in \mathbb{N}$ y $\sum_{n=1}^{+\infty} p_n = 1$, y definimos $P : 2^\Omega \to [0,1]$ tal que para todo $A \in 2^\Omega$, $P(A) = \sum_{n\,:\,x_n\in A} p_n$, entonces se cumple que la terna $\left(\Omega, 2^\Omega, P\right)$ es un espacio de probabilidad. Observamos que según esta definición se tiene que $P\left(\{w_n\}\right) = p_n$ para todo n.

Ejemplo 1.27. Modelo de equiprobabilidad. Si Ω es finito, definimos $P : 2^\Omega \to [0,1]$ tal que $P(A) = \frac{n(A)}{n(\Omega)}$ siendo $n(A)$ la cantidad de elementos que tiene el conjunto A. Observamos que en este caso, se tiene que si $\Omega = \{w_1, w_2, ..., w_n\}$ entonces $P\left(\{w_i\}\right) = 1/n$ para todo $i = 1, 2, 3, ..., n$, lo cual significa que todo elemento de Ω es igualmente probable.

En general, cuando Ω es finito o infinito numerable, si no se aclara nada al respecto se sobreentiende que la σ-álgebra considerada es 2^{Ω}. En numerosas ocasiones se está en presencia de un espacio muestral Ω finito donde cada elemento tiene la misma probabilidad.

Ejemplo 1.28. Se tiran 3 dados y se desea calcular la probabilidad de que salga al menos un 2 en las 3 tiradas.
En este caso, $\Omega = \{(i,j,k) \ : i,j,k \in \{1,2,3,4,5,6\}\}$. Para calcular $n(\Omega)$ observamos que para la terna (i,j,k) tenemos 6 valores posibles de i, por cada valor de i tenemos 6 valores posibles para j por lo que existen $6^2 = 36$ pares (i,j), y por cada uno de estos 36 pares tenemos 6 posibles valores de k, así obtenemos $6^3 = 216$ ternas en Ω.
Por otro lado, para el suceso $A =$ "sale al menos un 2 en las 3 tiradas", podemos realizar la descomposición $A = B \cup C \cup D$ donde $B =$ "sale exactamente dos veces el 2 en las 3 tiradas", $C =$ "sale exactamente un 2 en las 3 tiradas", $D =$ "sale las 3 veces el 2 en las 3 tiradas". Esta unión es disjunta por lo que $P(A) = P(B) + P(C) + P(D)$.
Para calcular $P(B)$ observamos que si el 2 sale en el primer lugar, tenemos 5^2 ternas, pero el 2 puede salir en el segundo o en el tercer lugar, por lo que en total tendremos $n(B) = 3 \times 5^2 = 75$ y entonces $P(B) = 75/216$. Razonando similarmente, obtenemos $P(C) = 3 \times 5/216$ mientras que $P(D) = 1/216$, entonces $P(A) = 91/216$. Hubiera sido más sencillo observar que $A^c =$ "no sale ningún 2 en las 3 tiradas", entonces tenemos $5 \times 5 \times 5$ ternas donde esto ocurre, entonces $P(A^c) = 125/216$ y por lo tanto $P(A) = 1 - 125/216 = 91/216$.

Ejemplo 1.29. Si se tiran 24 veces dos dados, ¿es más ventajoso apostar por la aparición de al menos un doble 6, o no? En este caso, el total de casos posibles son $\underbrace{36 \times 36 \times ... \times 36}_{\text{24 veces}} = 36^{24}$, mientras que si definimos el suceso $A =$ "no aparece ningún doble 6 en las 24 tiradas", tenemos que $n(A) = \underbrace{35 \times 35 \times ... \times 35}_{\text{24 veces}} = 35^{24}$ y por lo tanto $P(A) = (35/36)^{24} = 0{,}508$ por lo que es más conveniente apostar a que no aparece ningún doble 6 en 24 tiradas.

1.3. Apéndice y notas históricas.

Comentario sobre la necesidad de trabajar con sigmas álgebras sobre espacios muestrales no numerables.
Dado un conjunto $\Omega \neq \Phi$, se dice que $\mathcal{A}$ es un álgebra de subconjuntos de Ω si y sólo si cumple los siguientes axiomas:

1. $\Omega \in \mathcal{A}$;
2. Si $A \in \mathcal{A}$ entonces $A^c \in \mathcal{A}$;
3. Si $A, B \in \mathcal{A}$ entonces $A \cup B \in \mathcal{A}$.

En el caso en que $\Omega = (0,1)$, entonces se verifica directamente que el conjunto I formado por uniones finitas de conjuntos de la forma: $(a,b]$; $(0,b]$; $(a,1)$ con $a,b \in (0,1)$ forman un álgebra de subconjuntos de $(0,1)$.
Por otro lado, también se puede verificar directamente que la función $P : I \to [0,1]$ tal que $P(A) =$ longitud de A, cualquiera sea $A \in I$, es una función que cumple ser finitamente aditiva, tal que $P((0,1)) = 1$.
Un teorema importante de teoría de la medida, el teorema de Carathéodory nos dice que si tenemos una terna $(\Omega, I; P)$ donde P es una función $P : I \to [0,1]$ que cumple que $P(\Omega) = 1$ y además es finitamente aditiva (o sea que $P(A \cup B) = P(A) + P(B)$ siempre que $A, B \in I$ sean tales que $A \cap B = \Phi$), entonces existe una única función P^* tal que $(\Omega, \sigma(I), P^*)$ es un espacio de probabilidad, tal que $P^*(A) = P(A)$ para todo $A \in I$. Dicho de otra manera, si tenemos una función de probabilidad finitamente aditiva, definida sobre un álgebra I de subconjuntos de Ω, entonces puede ser extendida de manera única sobre la σ-álgebra generada por I.
Volviendo al ejemplo del espacio $(0,1)$ y el álgebra I, entonces sabemos que $\sigma(I) = \mathcal{B}_{(0,1)}$. Usando estas ideas veremos que existen conjuntos no borelianos. Definimos la relación en $(0,1)$, xRy si y sólo si $x - y \in \mathbb{Q}$. Se verifica en forma inmediata que la misma define una relación de equivalencia en $(0,1)$. Por lo tanto queda el conjunto $(0,1)$ particionado en clases de equivalencia. Elegimos un elemento de cada clase, y con ella formamos un conjunto que llamamos A. O sea que podemos escribir $(0,1) = \cup_{\alpha \in I} A_\alpha$, donde la unión es disjunta, y además $x, y \in A_\alpha$ si y sólo si $x - y \in \mathbb{Q}$. Para cada $\alpha \in I$ elegimos $a_\alpha \in A_\alpha$ de manera arbitraria (esto puede ser realizado gracias al axioma de elección), entonces definimos el conjunto $A = \cup_{\alpha \in I} \{a_\alpha\}$. Veremos a partir del teorema de extensión de Carathéodory que A no es boreliano. Para cada racional $q \in \mathbb{Q} \cap (0,1)$ definimos el conjunto $A_q = \{x + q : x \in A,\ x + q \leq 1\} \cup \{x + q - 1 : x \in A,\ x + q > 1\}$. Observando que los A_q son los trasladados por q del conjunto A, deducimos que si A fuera boreliano, entonces también lo sería A_q para cada $q \in \mathbb{Q} \cap (0,1)$. Observamos además que para todo $q \in \mathbb{Q} \cap (0,1)$ se cumple que $P(A_q) = P(A)$.
Por otro lado, se cumple que $(0,1) = \cup_{q \in \mathbb{Q} \cap (0,1)} A_q$, además la unión es disjunta. Por lo tanto, extendiendo por Carathéodory la función P a la σ-álgebra generada por I que es la σ-álgebra de Borel en $(0,1)$, obtendríamos que

$$1 = P((0,1)) = \sum_{q \in \mathbb{Q} \cap (0,1)} P(A_q) = 0$$

lo cual es absurdo.

Observación 1.30. *Este resultado además de demostrar que existen conjuntos no borelianos, nos permite demostrar también que cuando* $\Omega = (0,1)$, *es imposible definir una función de probabilidad sobre todos los subconjuntos de* $(0,1)$, *de tal modo de que la probabilidad de un intervalo incluido en* $(0,1)$ *sea la longitud del mismo.*
Por lo tanto si queremos trabajar con un espacio de probabilidad donde se elije un punto al azar en el intervalo $(0,1)$, *de tal modo que la probabilidad de un intervalo incluido en* $(0,1)$ *sea la longitud del mismo, no nos quedará más remedio que defi-*

nirlo como la longitud sobre los intervalos, y luego vía el teorema de Carathéodory, extenderlo a la σ-álgebra de Borel sobre $(0, 1)$.

Un poco de historia.
Como fue visto en el ejemplo 1.29, la probabilidad de la aparición de al menos un doble seis cuando se tira 24 veces un par de dados, es 0,492, por lo tanto es levemente desfavorable a apostar a que no sale ningún doble 6. Dada la proximidad de este valor a $1/2$, sin saber realizar este cálculo, difícilmente podríamos prever si era favorable o desfavorable apostar a este evento, por el simple hecho de repetirlo muchas veces y contabilizar su frecuencia. Esta situación se le presentó a Antoine de Gombaud (caballero de Meré), noble francés quien en 1654 interesado en resolver este problema, se lo planteó a Blaise Pascal, quien comenzó a cartearse con Pierre de Fermat, para discutir y llegar a la solución del problema. Si bien los juegos de azar, son tan antiguos como la humanidad, y es natural pensar que los primeros matemáticos babilónicos y griegos ya trabajaron y por lo tanto obtuvieron ciertos resultados probabilísticos, se considera que éste intercambio de correspondencia entre de Fermat y Pascal motivó el inicio de la teoría de la probabilidad, o al menos el comienzo de la construcción de los principios de la misma. Christian Huygens (quien fuera maestro de Leibnitz), enterado de esta correspondencia, en 1657 publicó lo que es conocido como el primer libro de teoría de probabilidades: *De Ratiociniis in Ludo Aleae*, que se trata de un libro de problemas de juegos de azar.
Anterior en el tiempo a esta correspondencia y a Huygens, vale la pena destacar que el matemático italiano Gerolamo Cardano en el siglo XVI ya había resuelto algunos problemas de juegos de azar, e incluso escribió un tratado sobre probabilidad, *"Liber de ludo aleae"*, pero el mismo fue publicado casi un siglo después de su muerte, en 1663.
El primero en dar la definición clásica de probabilidad (casos favorables sobre casos posibles) fue James Bernoulli (1654-1705), en una obra fundamental para el desarrollo de la teoría de la probabilidad: *"Ars Conjectandi"* (El arte de conjeturar), esta obra fue publicada en 1713. En 1812, Pierre Simon de Laplace, en su libro *Théorie analytique des probabilités*, introduce numerosas ideas y técnicas para resolver problemas de azar.
De manera un tanto irregular, numerosos matemáticos aportaron nuevas ideas a la teoría, se plantearon nuevos problemas, y se desarrollaron nuevos conceptos, pero aún quedaba una definición que sea adecuada y satisfactoria a situaciones donde está presente el azar, pero que no tienen que ver con juegos de azar, ni pueden ser repetidos en idénticas condiciones muchas veces. Esta falta de una definición precisa hizo que muchos matemáticos se "desencantaran" y consideraran a la probabilidad no como una teoría matemática, y se alejaron de ella.
Durante los tres siglos en que se buscó una definición adecuada y amplia para la probabilidad, hubieron distintas escuelas, como la clásica, la frecuencista y la subjetivista que tuvieron distintas controversias entre si, ya que todas daban definiciones que no eran totalmente satisfactorias.
La escuela clásica es la que acotaba los problemas probabilísticos a los casos en que Ω es finito con resultados equiprobables, por lo que definían probabilidad como el

número casos favorables sobre el número de casos posibles. Claramente esta definición no es aplicable a muchas situaciones que se dan en la práctica, tanto porque a veces Ω es infinito, como cuando los elementos del mismo no son equiprobables. Otros definieron lo que se llama interpretación frecuencista, que dice que para calcular la probabilidad de un evento se lo debe repetir n veces, y entonces es el límite cuando n tiende a inifinto del número de veces que ocurre el evento dividido el número de repeticiones del experimento (n). Nuevamente es claro que esta interpretación tiene el defecto de que muchas veces el experimento no puede ser repetido en idénticas condiciones, y además, no se pueden hacer infinitos experimentos. Por otro lado, el límite no es el límite usual, hay que definir otro concepto de límite, ya que el azar no permitiría asegurarnos un n tal que a partir del mismo, la probabilidad del suceso diste de la frecuencia observada tan poco como se quiera. Esta escuela está basada en la ley de los grandes números que veremos más adelante.

Por último los subjetivistas, decían que la probabilidad estaba dado por un carácter subjetivo, en el sentido de que la probabilidad de un suceso, es el grado de confianza que se tiene de que el mismo ocurra. De esta manera dos personas distintas pueden tener probabilidades diferentes para un mismo suceso, puesto que sus grados de confianza de que el mismo ocurra son distintos. Incluso una misma persona, en otro momento puede llegar a tener una valoración distinta de la ocurrencia de un suceso y por lo tanto cambiar su grado de confianza. Esta escuela tuvo por precursores a Bruno de Finetti y Leonard Savage.

Hubo que esperar hasta 1933 cuando Andrei Nikolayevich Kolmogorov, en su monografía titulada *"Grundbegriffe der Wahrscheinlichkeitsrechnung"* (Fundamentos de Probabilidad) planteara la definición axiomática de espacio de probabilidad, dándose cuenta a partir de la teoría de la medida y de los trabajos de Borel y Lebesgue, que calcular probabilidades, "es una forma de medir". Se puede decir que a partir de este trabajo, definitivamente y para todos los matemáticos, la probabilidad pasó a ser un tema de matemática, y además concluyó con todas las discusiones sobre la definición de probabilidad, ya que todas ellas quedaron como casos particulares de un espacio de probabilidad.

Si bien un espacio de probabilidad es un caso particular de espacio de medida, tiene conceptos y formas intuitivas de pensar problemas probabilísticos (como la probabilidad condicional y el concepto de independencia, que serán vistos en el próximo capítulo) que la independizan en muchos aspectos de la teoría de la medida.

Capítulo 2

Probabilidad condicional e independencia.

2.1. Probabilidad condicional.

Supongamos que participamos de un juego en el que se tira una moneda sucesivamente dos veces, y nosotros apostamos a que salen ambas caras. La probabilidad que tenemos de ganar la apuesta es $1/4$. Ahora bien, si ya se lanzó la primer moneda y salió cara, ahora nuestra probabilidad de ganar pasó a ser $1/2$. Se observa que en este caso, se agregó información sobre el experimento. En este ejemplo, $\Omega = \{(C, C); (N, C); (C, N); (N, N)\}$ y si le llamamos $A = \{(C, C)\}$ (salen ambas caras) y $B = \{(C, C); (C, N)\}$ (la primera salió cara), como dijimos $P(A) = 1/4$ pero la probabilidad de que ganemos la apuesta sabiendo que el primer lanzamiento salió cara, lo anotaremos como $P(A/B)$ y vale $P(A/B) = 1/2$. Como se ve en este caso, al cambiar la información que tenemos sobre el experimento, observamos que cambió el espacio muestral. Al calcular $P(A/B)$ pensamos el calcular la probabilidad de A, suponiendo que el espacio muestral es B. Si estamos en el modelo de equiprobabilidad, calcularíamos $P(A/B) = \frac{n(A\cap B)}{n(B)}$ ya que ahora nuestros casos posibles son el total de elementos de B, esto es $n(B)$ y los casos favorables son aquellos en los que ocurre el suceso A (de entre los que ocurren B), esto es $n(A \cap B)$, por lo tanto observamos que en el modelo de equiprobabilidad la manera general de calcular la probabilidad condicional sería así:

$$P(A/B) = \frac{n(A \cap B)/n(\Omega)}{n(B)/n(\Omega)} = \frac{P(A \cap B)}{P(B)}.$$

Este cálculo (y otros) motivan la siguiente definición.

Definición 2.1. Si $(\Omega, \mathcal{A}, P)$ es un espacio de probabilidad, dados $A, B \in \mathcal{A}$ donde $P(B) > 0$. Definimos $P(A/B) = \frac{P(A\cap B)}{P(B)}$.

La notación $P(A/B)$, la leemos como la probabilidad de que ocurra A, sabiendo que ocurre B. En todos los teoremas que siguen se considera dado $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad.

Teorema 2.2. *$P(A \cap B) = P(A/B)P(B)$ cualesquiera sean $A, B \in \mathcal{A}$ tal que $P(B) > 0$.*

Demostración.
Evidente a partir de la definición.✓

Teorema 2.3. *$P(A/B) = \frac{P(B/A)P(A)}{P(B)}$ cualesquiera sean $A, B \in \mathcal{A}$ tales que $P(A) > 0$ y $P(B) > 0$.*

Demostración.

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B/A)P(A)}{P(B)}.\checkmark$$

Teorema 2.4. *Si la familia $\{B_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ es tal que*
i) *$B_i \cap B_j = \Phi$ para todos $i \neq j$ (es decir que son sucesos disjuntos dos a dos),* ***ii)*** *$\cup_{n=1}^{+\infty} B_n = \Omega$* ***iii)****$P(B_n) > 0$ para todo $n \in \mathbb{N}$. Entonces cualquiera sea $A \in \mathcal{A}$ se tiene que*

1. ***Fórmula de probabilidades totales.***

$$P(A) = \sum_{n=1}^{+\infty} P(A/B_n)P(B_n).$$

2. ***Fórmula de Bayes.*** *Para A tal que $P(A) > 0$,*

$$P(B_k/A) = \frac{P(A/B_k)P(B_k)}{\sum_{n=1}^{+\infty} P(A/B_n)P(B_n)} \text{ para todo } k \in \mathbb{N}.$$

Demostración.

1. Dado A, de ii) deducimos que $A = \cup_{n=1}^{+\infty}(A \cap B_n)$ unión disjunta, entonces

$$P(A) = \sum_{n=1}^{+\infty} P(A \cap B_n) = \sum_{n=1}^{+\infty} P(A/B_n)P(B_n).\checkmark$$

2. Dado cualquier $k \in \mathbb{N}$, tenemos por aplicación de la propiedad 2 que

$$P(B_k/A) = \frac{P(A/B_k)P(B_k)}{P(A)}$$

y luego usando la fórmula de probabilidades totales se obtiene que

$$P(B_k/A) = \frac{P(A/B_k)P(B_k)}{\sum_{n=1}^{+\infty} P(A/B_n)P(B_n)} \checkmark$$

Observación 2.5. *Este teorema sigue siendo válido si la partición de Ω en unión de los B_n es finita.*

Teorema 2.6. *Si $B \in \mathcal{A}$ es tal que $P(B) > 0$. Definimos $\mathcal{A}_{\mathcal{B}} = \{A \cap B : A \in \mathcal{A}\}$ y $P_B : \mathcal{A}_{\mathcal{B}} \to [0,1]$, tal que $P_B(A) = P(A/B)$. Entonces $(B, \mathcal{A}_{\mathcal{B}}, P_B)$ es un espacio de probabilidad.*

Demostración. Se deja como ejercicio chequear que $\mathcal{A}_{\mathcal{B}}$ es una σ-álgebra de conjuntos sobre B y que P_B define una probabilidad sobre B.

Teorema 2.7. *Si $A, B, C \in \mathcal{A}$ con $P(B) > 0$, entonces*

1. $P(A^c/B) = 1 - P(A/B)$.
2. $P(A \cup C/B) = P(A/B) + P(C/B) - P(A \cap C/B)$.

Demostración. Ambas fórmulas son consecuencias directas de la propiedad anterior.✓

Teorema 2.8. *Si $A_1, A_2, ..., A_n \in \mathcal{A}$ cumplen que $P(A_1 \cap A_2 \cap ... \cap A_{n-1}) > 0$ entonces*

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1) P(A_2/A_1) P(A_3/A_1 \cap A_2) ... P(A_n/A_1 \cap A_2 \cap ... \cap A_{n-1}).$$

Demostración. Se deja como ejercicio.✓

Ejemplo 2.9. Supongamos que se dispone de un bolillero con 44 bolillas numeradas del 1 al 44. Se extraen 5 sucesivamente sin reponerse cada bolilla expraída. Se supone que apostamos a que salen los números 5,13,16,18,33. Deseamos calcular la probabilidad de que acertemos al menos 2 de los 5 extraídos. En este caso, para calcular los casos posibles, se ve que para la primer bolilla hay 44 posibles números, para la segunda 43 (todos menos el que salió en el primer lugar), para la siguiente 42, luego 41 y luego 40, así tenemos $44 \times 43 \times 42 \times 41 \times 40$ casos posibles. Para los favorables, calculamos los del complemento. Observamos que si le llamamos $A =$ "salen al menos dos de los 5 apostados", entonces $A^c = B \cup C$ donde $B =$"no sale ninguno de los 5 apostadosz $C =$"sale exactamente uno de los 5 apostados". La unión es disjunta por lo que $P(A^c) = P(B) + P(C)$. Los casos posibles para B son $39 \times 38 \times 37 \times 36 \times 35$ mientras que para C tenemos que $5 \times 39 \times 38 \times 37 \times 36$ son todas las posibilidades en que acertamos en la primera extracción y no acertamos en las 4 restantes, a esos hay que sumarles los que acertamos en la segunda y erramos en las restantes, etc, etc, como cada uno de esos casos son $5 \times 39 \times 38 \times 37 \times 36$ entonces el total de casos favorables para C son $5 \times 39 \times 38 \times 37 \times 36 \times 5$, de esta forma

$$P(A) = 1 - \frac{39 \times 38 \times 37 \times 36 \times 35 + 5 \times 39 \times 38 \times 37 \times 36 \times 5}{44 \times 43 \times 42 \times 41 \times 40} = 0,0911.$$

Este mismo cálculo podría haberse realizado mediante el uso de la propiedad anterior. Para calcular $P(B)$, llamémosle $A_1 =$"no acierto la primer bolilla extraída",

A_2 ="no acierto la primer bolilla extraída",...,A_5 ="no acierto la quinta bolilla extraída". Entonces $P(A_1) = 39/44$, $P(A_2/A_1) = 38/43$, $P(A_3/A_1 \cap A_2) = 37/42$, $P(A_4/A_1 \cap A_2 \cap A_3) = 36/41$ y $P(A_5/A_1 \cap A_2 \cap A_3 \cap A_4) = 35/40$, así se tiene

$$P(B) = P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = \frac{39 \times 38 \times 37 \times 36 \times 35}{44 \times 43 \times 42 \times 41 \times 40}$$

Para $P(C)$ lo separamos como suma de acertar exactamente la primera, más acertar exactamente la segunda, etc y definimos adecuadamente los conjuntos A_1, A_2, A_3, A_4, A_5 y se procede de manera análoga.

Ejemplo 2.10. Se tiene una urna compuesta por 3 bolillas azules, 2 blancas y una roja, y una segunda urna compuesta por 3 blancas y 3 azules. Se extrae una bolilla de la urna uno, se la deposita en la segunda y luego se extrae una bolilla de esta segunda urna. Calculemos las probabilidades de: A ="la segunda bolilla extraída es azul", B ="la primer bolilla extraída es azul, sabiendo que la segunda fue blanca". En este caso, aplicamos la propiedad de probabilidades totales quedando $P(A) =$

$$P(A/1^a \text{ blanca})\, P(1^a \text{ blanca}) + P(A/1^a \text{ azul})\, P(1^a \text{ azul}) + P(A/1^a \text{ roja})\, P(1^a \text{ roja}) =$$

$$\frac{3}{7}\frac{3}{6} + \frac{4}{7}\frac{3}{6} + \frac{3}{7}\frac{1}{6} = 0,571.$$

Para B, usamos el teorema de Bayes quedando $P(B) = P(1^a \text{ azul } / \ 2^a \text{ blanca}) =$

$$\frac{P(\ 2^a \text{ b } / \ 1^a \text{ b})\, P(1^a \text{ b})}{P(\ 2^a \text{ b } / \ 1^a \text{ b})\, P(1^a \text{ b}) + P(\ 2^a \text{ b } / \ 1^a \text{ a})\, P(1^a \text{ a}) + P(\ 2^a \text{ b } / \ 1^a \text{ r})\, P(1^a \text{ roja})} =$$

$$\frac{\frac{4}{7}\frac{3}{6}}{\frac{4}{7}\frac{2}{6} + \frac{3}{7}\frac{3}{6} + \frac{3}{7}\frac{1}{6}} = 0,6.$$

2.2. Independencia.

Definición 2.11. Dado $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, se dice que la familia de sucesos $\{A_\alpha\}_{\alpha \in I}$ donde I es una familia cualquiera de índices, son sucesos independientes si y sólo si, para todo $F \subset I$ finito, se cumple que

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

Observación 2.12. *Si la familia de sucesos se reduce a dos, entonces la definición anterior nos dice que A y B son independientes si y sólo si $P(A \cap B) = P(A)\, P(B)$, lo cual en el caso en que $P(B) > 0$ es equivalente a pedir que se cumpla que $P(A/B) = P(A)$, pero la ventaja que tiene la definición dada es que no requiere que los sucesos tengan probabilidad positiva.*

Observación 2.13. *Si la familia de sucesos se reduce a 3, digamos A, B y C, entonces los mismos son independientes si y sólo si se cumplen las siguientes cuatro condiciones:*

1. $P(A \cap B) = P(A)P(B)$
2. $P(A \cap C) = P(A)P(C)$
3. $P(B \cap C) = P(B)P(C)$
4. $P(A \cap B \cap C) = P(A)P(B)P(C)$

Observación 2.14. *Observamos que en el caso anterior, para pedir que los tres sucesos A, B y C sean independientes, se requiere que sean independientes de a pares, que son las condiciones 1,2 y 3, pero a esto se le debe agregar la condición 4 ya que las condiciones 1,2 y 3 (como se verá en el siguiente ejemplo) no aseguran que A sea independiente del suceso $B \cap C$. Se puede chequear sin dificultad que las 4 condiciones que determinan la independencia de A, B y C aseguran la independencia de A con $B \cup C$ y la de A con $B \cup C^c$ etc.*

Se deja como ejercicio verificar el siguiente ejemplo, donde se muestra que tres sucesos pueden ser independientes tomados de a dos, pero no ser independientes.

Ejemplo 2.15. Se tira un par de dados, uno azul y uno verde. Definimos $A =$"en el dado azul sale el 5", $B =$"en el dado verde sale el 3", $C =$"la suma de los resultados de ambos dados es un número par". Entonces A, B y C son independientes tomados de a pares, pero A, B y C no son independientes.

Teorema 2.16. *Dado $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, si una familia de sucesos $\{A_\alpha\}_{\alpha \in I}$ son independientes, entonces también lo son la familia $\{B_\alpha\}_{\alpha \in I}$, donde para cada $\alpha \in I$, se tiene que, o bien $B_\alpha = A_\alpha$, o bien $B_\alpha = A_\alpha^c$.*

Teorema 2.17. ***Lema de Borel Cantelli.*** *Dados $(\Omega, \mathcal{A}, P)$ espacio de probabilidad y la sucesión $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$, entonces*

1. *Si $\sum_{n=1}^{+\infty} P(A_n) < +\infty$ entonces*
$$P(\mathit{limsup}\ A_n) = 0.$$
2. *Si $\sum_{n=1}^{+\infty} P(A_n) = +\infty$ y además $\{A_n\}_{n \in \mathbb{N}}$ son independientes, entonces*
$$P(\mathit{limsup}\ A_n) = 1.$$

Demostración.

1. $P(\text{limsup } A_n) = \lim P\left(\bigcup_{k=n}^{+\infty} A_k\right) \leq \sum_{k=n}^{+\infty} P(A_n) \to 0$ puesto que la serie es convergente.✓

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

$$P\left(\bigcap_{k=n}^{+\infty} A_k^c\right) \leq P\left(\bigcap_{k=n}^{m} A_k^c\right) = \prod_{k=n}^{m} P\left(A_k^c\right) = \prod_{k=n}^{m} \left[1 - P\left(A_k\right)\right].$$

   Ahora, usando que $1 - x \leq e^{-x}$ para todo $x \geq 0$, se deduce que

$$\prod_{k=n}^{m} \left[1 - P\left(A_k\right)\right] \leq \prod_{k=n}^{m} e^{-P(A_k)} = e^{-\sum_{k=n}^{m} P(A_k)} \underset{m \to +\infty}{\longrightarrow} 0.\checkmark$$

Ejemplo 2.18. Supongamos que se elije al azar un número en el intervalo $(0, 1)$ ¿Cuál es la probabilidad de que aparezcan infinitos 4 en su expansión decimal? ¿Y la probabilidad de que el 44 aprezca infinitas veces?
Para responder a la primer pregunta, definimos los sucesos A_n =“el 4 aparece en el n-ésimo lugar en su expansión decimal”, entonces la sucesión $\{A_n\}_{n\in\mathbb{N}}$ está formada por sucesos independientes, además, $P(A_n) = 1/10$ cualquiera sea n, entonces $\sum_{n=1}^{+\infty} P(A_n) = +\infty$ y por lo tanto la probabilidad de que aparezca el 4 infinitas veces es 1. Para responder la otra pregunta, procedemos de forma similar, definimos B_n =“el 4 aparece en el n-ésimo lugar y en el siguiente en su expansión decimal”, en este caso $P(B_n) = 1/100$ para todo n, pero los B_n no son independientes. De todas formas si consideramos la subsucesión de sucesos $\{B_{2n}\}_{n\in\mathbb{N}}$, ahora si, tenemos una sucesión de sucesos independientes y como $\sum_{n=1}^{+\infty} P(B_{2n}) = +\infty$, tenemos que la probabilidad de que aparezca el 44 inifintas veces en un lugar par seguido de uno impar es 1, pero éste último suceso está incluido en el suceso de que el 44 aparece infintas veces, se entonces la probabilidad de que el 44 aparezca infinitas veces es 1 también.

2.3. Notas históricas.

El ejemplo anterior es conocido como el teorema de los infinitos monos. Emil Borel en su trabajo *“Mécanique Statistique et Irréversibilité”* en 1913 afirmaba que si se pone a un millón de monos durante 10 horas a teclear una máquina de escribir (como una manera de decir que se eligen al azar letras del alfabeto, tantas como pueda teclear durante 10 horas un mono), es extremadamente improbable que sea posible encontrar una secuencia de letras tecleadas que sean el desarrollo de un libro por más pequeño que sea. Ahora, de acuerdo al ejemplo que acabamos de desarrollar, hemos probado que si a un solo mono se le da tiempo infinito, entonces hay probabilidad 1 de que en algún momento escriba la obra completa de Shakespeare, por ejemplo. Sólo basta cambiar el conjunto de los 10 dígitos por los símbolos del alfabeto, y la tirada 44 por la de la obra completa de Shakespeare que es finita.
Thomas Bayes nació en Inglaterra en 1702 y murió en 1761. Se sabe muy poco de su vida, ya que no se dedicó activamente a la matemática, no se vinculó mayormente con

otros matemáticos de su época, y por lo tanto no se destacó tanto mientras estuvo con vida. Sus aportes a la teoría de la probabilidad fueron enormes, ya que fue el primero que definió y trabajó el concepto de probabilidad condicional, en tiempos en que todos los cálculos probabilísticos estaban restringidos a juegos de azar y los cálculos eran realizados según el modelo de equiprobabilidad.
También es esencial su aporte a la definición que utiliza de probabilidad, que fue olvidada hasta el siglo XX, y que fue retomada recién en 1937 por Bruno De Finetti, uno de los primeros precursores de la teoría subjetiva de la probabilidad.
Todos estos aportes fueron publicados en un trabajo titulado *"An Essay Towards Solving a Problem in Doctrine of Chances"* publicado en 1763 (2 años después de su muerte), y el hoy llamado teorema de Bayes, fue publicado en 1764 en las "Philosophical Transactions Vol 53", que es la base de la hoy llamada inferencia bayesiana. Es curioso que Bayes no haya intentado publicar sus trabajos, tanto su teorema como su trabajo publicado en 1763, fueron encontrados por amigos suyos luego de su muerte.

Capítulo 3

Variable Aleatoria.

Definición 3.1. Dado un espacio de probabilidad $\left(\Omega, \mathcal{A}, P\right)$. Diremos que $X : \Omega \to \mathbb{R}^k$ es una variable aleatoria en $\mathbb{R}^k$ si y sólo si, se cumple que para cada A boreliano se cumple que

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

Cuando $k > 1$, también es llamado vector aleatorio.

Observación 3.2. *Dado que la σ-álgebra de Borel está engendrada por los conjuntos abiertos, basta verificar que $X^{-1}(A) \in \mathcal{A}$, para todo A abierto (o para todo A en algún generador de la σ-álgebra de Borel).*

Observación 3.3. *Si Ω es finito o infinito numerable, cualquier función $X : \Omega \to \mathbb{R}^k$ es vector aleatorio, ya que en estos casos, consideramos como σ-álgebra a 2^Ω.*

Observación 3.4. *Toda constante, es vector aleatorio, cualquiera sea $\left(\Omega, \mathcal{A}, P\right)$ espacio de probabilidad, ya que el conjunto $X^{-1}(A)$ es Ω si la constante está en A o vacío si no, en ambos casos $X^{-1}(A) \in \mathcal{A}$.*

En varias ocasiones, es conveniente trabajar con funciones a valores en $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$. Para dichos casos será conveniente extender la σ-álgebra de Borel a $\mathcal{B}_{\overline{\mathbb{R}}}$. Por suerte es posible hacerlo de una forma sencilla.
Si le llamamos $\mathcal{B}$ a la σ-álgebra de Borel en $\mathbb{R}$, definimos $\mathcal{B}_{\overline{\mathbb{R}}} =$

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

Se deja como ejercicio probar que $\mathcal{B}_{\overline{\mathbb{R}}}$ es una σ-álgebra de Borel sobre $\mathbb{R}$.
Frecuentemente para simplificar la notación, se suele escribir el conjunto $X^{-1}(A) = \{w \in \Omega : X(w) \in A\}$ mediante la simple escritura de $\{X \in A\}$. Así, por ejemplo al conjunto $X^{-1}((-\infty, a])$ lo denotaremos por $\{X \leq a\}$.

3.1. Propiedades.

Teorema 3.5. *Dado $X = (X_1, X_2, ..., X_k) : \Omega \to \mathbb{R}^k$. Entonces, X es vector aleatorio si y sólo si $X_1, X_2, ..., X_k$ son variables aleatorias en $\mathbb{R}$.*

Demostración.
Comenzamos observando que cualesquiera sean los conjuntos $A_1, A_2, ..., A_k$, se tiene que

$$X^{-1}(A_1 \times A_2 \times ... \times A_k) = \bigcap_{i=1}^{k} X_i^{-1}(A_i).$$

$\Rightarrow$) Si A es un boreliano en $\mathbb{R}$, entonces

$$X_i^{-1}(A) = X^{-1}\left(\mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R} \times \underbrace{A}_{\text{lugar } i} \times \mathbb{R} \times ... \times \mathbb{R}\right) \in \mathcal{A}.$$

Entonces X_i es variable aleatoria.✓
$\Leftarrow$) Cualesquiera sean $a_1, a_2, ..., a_k \in \mathbb{R}$ se tiene que

$$X^{-1}((-\infty, a_1) \times (-\infty, a_2) \times ... \times (-\infty, a_k)) = \bigcap_{i=1}^{k} X_i^{-1}((-\infty, a_i)) \in \mathcal{A}$$

ya que cada conjunto que intersectamos pertenece a $\mathcal{A}$, entonces X es vector aleatorio en $\mathbb{R}^k$.✓

Teorema 3.6. *Si $X : \Omega \to \mathbb{R}^k$ es vector aleatorio y $g : \mathbb{R}^k \to \mathbb{R}^n$ es continua, entonces $Y = g(X)$ es vector aleatorio en $\mathbb{R}^n$.*

Demostración.
Dado un abierto A en $\mathbb{R}^n$, entonces $g^{-1}(A)$ es abierto por la continuidad de g, por lo que

$$Y^{-1}(A) = (goX)^{-1}(A) = X^{-1}\left[g^{-1}(A)\right] \in \mathcal{A}.✓$$

Teorema 3.7. *Si $X, Y : \Omega \to \mathbb{R}$ son variables aleatorias, entonces también lo son αX, $X + Y$ y XY.*

Demostración.
Es consecuencia inmediata de la propiedad anterior, ya que (X, Y) es vector aleatorio en $\mathbb{R}^2$, y lo componemos con las funciones continuas $g : \mathbb{R}^2 \to \mathbb{R}$ definidas como $g(x, y) = \alpha x$, $g(x, y) = x + y$ y $g(x, y) = xy$ respectivamente. ✓

Teorema 3.8. *Si $X_n : \Omega \to \mathbb{R}$ es variable aleatoria para todo $n \in \mathbb{N}$, entonces también lo son las variables $Y : \Omega \to \mathbb{R} \cup \{+\infty\}$ tal que $Y = sup\{X_1, X_2, ..., X_n, ...\}$ y $Z : \Omega \to \mathbb{R} \cup \{-\infty\}$ tal que $Z = inf\{X_1, X_2, ..., X_n, ...\}$.*

Demostración.
Basta observar que si tenemos una sucesión de números reales $\{x_n\}_{n \in \mathbb{N}}$, entonces, cualesquiera sea $a \in \mathbb{R} \cup \{+\infty\}$ se tiene que

$$sup\{x_1, x_2, ..., x_n, ...\} \leq a \Leftrightarrow x_n \leq a \text{ para todo } n.$$

Entonces

$$Y^{-1}\left((-\infty, a]\right) = \bigcap_{n=1}^{+\infty} X_n^{-1}\left((-\infty, a]\right) \in \mathcal{A}.$$

Entonces Y es variable aleatoria. Por otro lado, como $Z = -\sup\{-X_1, -X_2, ..., -X_n, ...\}$, se deduce de lo recién probado que Z también es variable aleatoria. ✓

Teorema 3.9. *Si $X_n : \Omega \to \mathbb{R}$ es variable aleatoria para todo $n \in \mathbb{N}$, entonces también lo son las variables $limsupX_n : \Omega \to \mathbb{R} \cup \{+\infty\}$ y $liminfX_n : \Omega \to \mathbb{R} \cup \{-\infty\}$.*

Demostración.
Es consecuencia inmediata de la propiedad anterior ya que
limsup $X_n = \inf\limits_{n} \sup\limits_{k \geq n} X_k$, y liminf $X_n = \sup\limits_{n} \inf\limits_{k \geq n} X_k$.✓

3.2. Función de distribución de una variable aleatoria.

Definición 3.10. Función de distribución de una variable aleatoria.
Dados un espacio de probabilidad $\big(\Omega, \mathcal{A}, P\big)$ y $X : \Omega \to \mathbb{R}$ una variable aleatoria, definimos la función $F_X : \mathbb{R} \to \mathbb{R}$ como $F_X(x) = P\left(X \leq x\right)$ para cada $x \in \mathbb{R}$.

Observación 3.11. *Para todo $x \in \mathbb{R}$ se tiene que $\{X \leq x\} = X^{-1}((-\infty, x]) \in \mathcal{A}$, por ser X variable aleatoria.*

En todas las propiedades que siguen se sobreentiende que tenemos un espacio de probabilidad $\big(\Omega, \mathcal{A}, P\big)$ y $X : \Omega \to \mathbb{R}$ una variable aleatoria.

Teorema 3.12. *F_X es monótona creciente.*

Demostración.
Si $a < b$ entonces $\{X \leq a\} \subset \{X \leq b\}$, entonces $P\left(X \leq a\right) \leq P\left(X \leq b\right)$, por lo que $F_X(a) \leq F_X(b)$.✓

Teorema 3.13. $\lim\limits_{x \to +\infty} F_X(x) = 1$.

Demostración.
Como F_X es monótona creciente, basta restringirse a una sucesión particular que tienda a $+\infty$, por ejemplo $\lim\limits_{n \to +\infty} F_X(n)$.

Observamos que $A_n = \{X \leq n\}$ es una sucesión creciente de sucesos, tal que $\bigcup\limits_{n=1}^{+\infty} A_n = \Omega$, entonces por la propiedad de continuidad de las probabilidades se tiene que

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

Teorema 3.14. $\underset{x \to -\infty}{lim} F_X(x) = 0.$

Demostración.
Razonamos análogamente al caso anterior, por lo que basta considerar $\lim\limits_{n\to+\infty} F_X(-n)$.

Consideramos ahora $A_n = \{X \leq -n\}$ decrece a $\bigcap\limits_{n=1}^{+\infty} A_n = \Phi$, por lo que se deduce que $\lim\limits_{n\to+\infty} F_X(-n) = \lim\limits_{n\to+\infty} P\left(A_n\right) = P\left(\bigcap_{n=1}^{+\infty} A_n\right) = P\left(\Phi\right) = 0.$✓

Teorema 3.15. *F_X es continua por derecha.*

Demostración.
Nuevamente, basta ver que $\lim\limits_{n\to+\infty} F_X(a{+}1/n) = F_X(a)$. La sucesión $A_n = \{X \leq a + 1/n\}$ decrece a $\bigcap\limits_{n=1}^{+\infty} A_n = \{X \leq a\}$, de donde se obtiene el resultado.

Teorema 3.16. *Si definimos $F_X(x^-) = P(X < x)$, entonces para cualquier $x \in \mathbb{R}$ se tiene que $F_X(a^-) = \underset{x \to a^-}{lim} F_X(x)$.*

Demostración.
Similar a la anterior, se deja como ejercicio.

Observación 3.17. *Del teorema anterior se deduce que $P(X = x) = F_X(x) - F_X(x^-)$, por lo que la probabilidad de que X tome un valor determinado, viene dado por el "salto" de la función de distribución en dicho x.*

Notas.

1. Dado un espacio de probabilidad sobre un conjunto Ω, $(\Omega, \mathcal{A}, P)$ y tenemos una variable aleatoria en él $X : \Omega \to \mathbb{R}$, la misma nos permite definir naturalmente un espacio de probabilidad donde el espacio muestral sea $\mathbb{R}$. El mismo sería $(\mathbb{R}, \mathcal{B}, F_X)$. Aquí hay un detalle técnico y es el hecho de que F_X debe estar definido en cualquier boreliano de $\mathbb{R}$, pero un teorema de teoría de la medida nos asegura que al ser F_X creciente y positiva, y estar definida en los conjuntos de la forma $(-\infty, x]$ para todo $x \in \mathbb{R}$ que generan la σ-álgebra de Borel, existe una única extensión de F_X a dicha σ-álgebra.

2. Recíprocamente, si tenemos una función $F : \mathbb{R} \to \mathbb{R}$, que cumple las siguientes condiciones: **i)** F es monótona creciente, **ii)** $\lim\limits_{x\to+\infty} F(x) = 1$, **iii)** $\lim\limits_{x\to-\infty} F(x) = 0$, **iv)** F es continua por derecha entonces, un teorema de teoría de la medida nos dice que existe un espacio de probabilidad $\big(\Omega, \mathcal{A}, P\big)$ y una variable aleatoria X definida sobre este espacio tal que $F_X = F$.

3.3. Variables Aleatorias Discretas.

Definición 3.18. Variables aleatorias discretas.
Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Diremos que $X : \Omega \to \mathbb{R}$ es una variable aleatoria discreta si y sólo si existe un subconjunto A_X de $\mathbb{R}$ numerable, tal que $P(X \in A_X) = 1$.

Definición 3.19. Si X es discreta y se considera A_X tal que $P(X = x) > 0$ para todo $x \in A_X$, al conjunto A_X le llamaremos $\text{Rec}(X)$.

Observación 3.20. *$\{X \in A_X\}$ es un suceso ya que al ser A_X numerable, entonces $A_X = \bigcup_{n=1}^{+\infty}\{x_n\}$ por lo que $\{X \in A_X\} = \bigcup_{n=1}^{+\infty}\{X = x_n\} \in \mathcal{A}$ ya que los puntos aislados son borelianos.*

Definición 3.21. Función de probabilidad. Si X es discreta, definimos $p_X : \mathbb{R} \to \mathbb{R}$ tal que $p_X(x) = P(X = x)$ para cada $x \in \mathbb{R}$.

Observación 3.22. *Cuando X es discreta, sólo una cantidad numerable de valores de x son tales que $P(X = x) > 0$ por lo que alcanza definir $p_X(x)$ para los $x \in Rec(X)$.*

Observación 3.23. *Cuando X es discreta, se tiene que $\sum_{x \in Rec(X)} p_X(x) = 1$.*

Observación 3.24. *Cuando X es discreta, entonces*

$$F_X(x) = \sum_{t \in Rec(X) \ : \ t \leq [x]} p_X(t).$$

3.4. Ejemplos de Variables discretas.

Ejemplo 3.25. Variable Bernoulli de parámetro p. Notación: $X \sim$Ber(p).

Si consideramos $(\Omega, \mathcal{A}, P)$ espacio de probabilidad cualquiera, $A \in \mathcal{A}$ tal que $P(A) = p \in (0,1)$ y definimos $X : \Omega \to \mathbb{R}$ tal que $X(w) = \begin{cases} 1 & \text{si} \quad w \in A \\ 0 & \text{si} \quad w \notin A \end{cases}$ diremos que en este caso X distribuye $\text{Ber}(p)$. La función de probabilidad queda en este caso $p_X(x) = \begin{cases} p & si \quad x = 1 \\ 1 - p & si \quad x = 0 \end{cases}$. Se suele decir que si ocurre A es éxito y si no fracaso, entonces p se interpreta como la probabilidad de éxito.

Ejemplo 3.26. Variable Binomial de parámetros n y p. Notación: $X \sim$Bin(n, p).

Si repetimos de manera independiente experimentos de Bernoulli con probabilidad de éxito p en cada prueba y definimos para cada $i = 1, 2, 3, , ..., n$

$$X_i = \begin{cases} 1 & \text{si hay éxito en la } i\text{-ésima prueba} \\ 0 & \text{si no} \end{cases}.$$

Entonces diremos que $X = X_1 + X_2 + ... + X_n$ (cantidad de éxitos en las n pruebas), distribuye Bin(n, p). En este caso es claro que Rec$(X) = \{0, 1, 2, ..., n\}$ y para obtener la función de probabilidad, observamos que si $x \in \{0, 1, 2, ..., n\}$, entonces $P(X = x)$ significa la probabilidad de obtener x éxitos (y por lo tanto $n-x$ fracasos). En primer lugar calculamos la probabilidad de que salga éxito las primeras x veces y fracaso las siguientes $n - x$ veces. Este suceso es $A_1 \cap A_2 \cap ... \cap A_x \cap A^c_{x+1} \cap ... \cap A^c_n$ donde $A_i =$ "sale éxito la vez i-ésima". Como las pruebas son independientes, la probabilidad de esta intersección es igual al producto de las mismas. Siendo p la probabilidad de cada éxito, se deduce que la probabilidad de obtener éxito las primeras x veces y fracaso las restantes es igual a $p^x (1-p)^{n-x}$. Ahora, si consideramos los x éxitos y $n-x$ fracasos en cualquier otro orden, la probabilidad será también $p^x (1-p)^{n-x}$, por lo tanto la probabilidad de obtener x éxitos y $n-x$ fracasos, será $p^x (1-p)^{n-x}$ multiplicado por la cantidad de maneras en que se pueden combinar los x éxitos y $n - x$ fracasos, de todas las maneras posibles. Para obtener dicho número, debemos elegir x lugares de entre los n para ubicar los éxitos (en los restantes lugares van los fracasos), por lo que el total de formas posibles es C^n_x. Entonces se obtuvo que

$$p_X(x) = C^n_x p^x (1-p)^{n-x} \text{ para todo } x \in \{0, 1, 2, ..., n\}.$$

Ejemplo 3.27. Variable Geométrica de parámetro p. Notación: $X \sim$Geo(p).

En este caso se realizan de manera independiente pruebas de Bernoulli hasta obtener el primer éxito. Aquí se define la variable $X =$ "cantidad de fracasos". En este caso, se tiene que Rec$(X) = \{0, 1, 2, ...\}$. Además, si $x \in \{0, 1, 2, ...\}$, el suceso $\{X = x\}$ significa que las primeras x veces hubo fracaso y luego hubo éxito. La probabilidad en este caso es (nuevamente usando que las pruebas son independientes) $(1-p)^x p$, por lo que

$$p_X(x) = (1-p)^x p \text{ para todo } x \in \{0, 1, 2, ...\}.$$

Observación 3.28. *Para el mismo experimento, se puede definir la variable $X =$ "cantidad de pruebas", también llamada con distribución geométrica y para la que se obtiene con el mismo argumento su función de probabilidad como*

$$p_X(x) = (1-p)^{x-1} p \text{ para todo } x \in \{1, 2, 3, ...\}.$$

Ejemplo 3.29. Variable Binomial Negativa de parámetros r, p. Notación: $X \sim$Bin Neg(r, p).

En este caso se realizan de manera independiente pruebas de Bernoulli hasta obtener el r-ésimo éxito. Aquí se define la variable $X =$ "cantidad de fracasos". En este caso, se tiene que Rec$(X) = \{0, 1, 2, ...\}$. Además, si $x \in \{0, 1, 2, ...\}$, el suceso $\{X = x\}$ significa que las primeras $x + r - 1$ veces, hubo $r - 1$ éxitos y x fracasos, y además en la prueba $x + r$ hubo éxito. Entonces la probabilidad del suceso $\{X = x\}$ es la probabilidad de que las primeras $x + r - 1$ veces, hubo $r - 1$ éxitos y x fracasos, que es (razonando como en la binomial) $C^{x+r-1}_{r-1} p^{r-1} (1-p)^x$multiplicado por p. Entonces

$$p_X(x) = C^{x+r-1}_{r-1} p^r (1-p)^x \text{ para todo } x \in \{0, 1, 2, ...\}.$$

Observación 3.30. *Al igual que lo visto para la geométrica, si definimos la variable* $X =$ *"cantidad de pruebas", también se llama binomial negativa, y su función de probabilidad queda*

$$p_X(x) = C_{r-1}^{x-1} p^r (1-p)^{x-r} \text{ para todo } x \in \{r, r+1, r+2, ...\}.$$

Ejemplo 3.31. Variable Hipergeométrica de parámetros N_1, N_2, n**. Notación:** $X \sim$**Hiper**(N_1, N_2, n).

En este caso se considera una población de N elementos, dividida en dos grupos, cuyos totales son N_1 y N_2. $N_1 + N_2 = N$. Se realizan n extracciones sin reposición de objetos de esta población. Le llamaremos éxito cda vez que una extracción sea de entre entre el grupo de los N_1 y fracaso en caso contrario. Definimos en este caso $X =$ "cantidad de éxitos entre las n extracciones". Observamos que $\text{Rec}(X) = \{x \in \mathbb{N} : \max\{0, N_2 - n\} \leq x \leq \min\{n, N_1\}\}$. El total de las formas posibles que hay de extraer n objetos de un total de N, sin reposición y sin importar el orden, es C_n^N. Análogamente, tenemos $C_x^{N_1}$ formas de elegir entre los N_1 elementos x, y por cada una de estas c ombinaciones tenemos $C_{n-x}^{N_2}$ formas de elegir entre los N_2 elementos, los restantes $n - x$, por lo tanto, tendremos $C_x^{N_1}\ C_{n-x}^{N_2}$ casos favorables,entonces

$$p_X(x) = \frac{C_x^{N_1} C_{n-x}^{N_2}}{C_n^N} \text{ para todo } x \in \text{Rec}(X).$$

Ejemplo 3.32. Variable Poisson de parámetro λ**. Notación:** $X \sim$**Poisson**(λ).

Esta variable suele ser útil para modelar diversos fenómenos, por ejemplo aquellos en los cuales se mide la cantidad de sucesos que ocurren en un intervalo de tiempo. $\text{Rec}(X) = \{0, 1, 2, ...\}$. Ejempos de estos fenómenos pueden ser dados por la cantidad de autos que pasan por un determinado puente en un intervalo de tiempo, rompimiento de cromosomas, desintegración de partículas, etc. Bajo ciertas hipótesis sobre el experimento es posible demostrar que existe un valor de $\lambda > 0$ tal que $p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$. Veremos en lo que sigue la deducción de la fórmula que nos da la función de probabilidad, de una variable aleatoria Poisson con parámetro $\lambda > 0$. Para realizar la deducción de la fórmula, será conveniente utilizar la siguiente definición.

Definición 3.33. Dado $\alpha > 0$, si $f : (-\delta, \delta) \to \mathbb{R}$ es tal que $\lim\limits_{h \to 0} \frac{f(h)}{h^\alpha} = 0$ diremos que f es $o(h^\alpha)$.

Observación 3.34. $o(h^\alpha)$ *es una función que representa un infinitésimo de orden mayor que* h^α *cuando* $h \to 0$.

Se deja como ejercicio, verificar las siguientes propiedades concernientes al álgebra de funciones $o(h^\alpha)$.

- $o(h^\alpha) \pm o(h^\alpha) = o(h^\alpha)$.
- Si f es una función acotada, entonces $f(h)o(h^\alpha) = o(h^\alpha)$.

- $o(h^{\alpha}) = o(h^{\beta})$ para cualquier $\beta \leq \alpha$.

Consideramos una familia de variables aleatorias discretas $\{X_t\}_{t>0}$ que toman valores en $\{0, 1, 2, 3, ...\}$. Le llamaremos $p_n(t) = P(X_t = n)$. Supondremos las siguientes hipótesis sobre las variables X_t.

1. H1: Las funciones p_n son derivables en todo punto $0 < p_0(1) < 1$ $p_0(0) = P(X_0 = 0) = 0$ (el proceso arranca en 0).

2. H2: La distribución de $X_{t+h} - X_t$ es igual a la de X_h para todos $t, h > 0$ (el proceso tiene incrementos estacionarios).

3. H3: Las variables $X_{t_2} - X_{t_1}$ y $X_{t_4} - X_{t_3}$ son independientes cualesquiera sean $0 < t_1 < t_2 < t_3 < t_4$ (el proceso tiene incrementos independientes).

4. H4: $P(X_t \geq 2) = o(t)$.

H2 significa que la distribución de $X_{t+h} - X_t$, sólo depende de h (no de t). Si $X_{t+h} - X_t$ cuenta la cantidad de sucesos que se observan en el intervalo $[t, t+h]$, la distribución de esta variable es igual a la de X_h que es la cantidad de sucesos que se observan en el intervalo $[0, h]$.
H3 significa que la cantidad de sucesos que se observan en el intervalo $[t_1, t_2]$ es independiente de la cantidad de sucesos que se observan en $[t_3, t_4]$ siendo estos intervalos disjuntos entre si.
H4 significa que para valores pequeños de t, la probabilidad de observar 2 o más sucesos en un intervalo de longitud t es un infinitésimo de mayor orden que la probabilidad de observar un sólo suceso en el mismo intervalo.

Lema 3.35. *Si se cumplen las condiciones H1, H2, H3 y H4 entonces existe* $\lambda > 0$ *tal que* $p_0(t) = e^{-\lambda t}$.

Demostración.
Para cada $t > 0$, partimos el intervalo $[0, t]$ en n subintervalos $\left(\frac{i-1}{n}t, \frac{it}{n}\right]$ $(i = 1, 2, 3, ..., n)$ de longitud constante e igual a $\frac{t}{n}$. Entonces, decir que en el intervalo $[0, t]$, no se observaron sucesos, es equivalente a decir que en todos los subintervalos $\left(\frac{i-1}{n}t, \frac{it}{n}\right]$ no se observaron sucesos.

$$p_0(t) = P(X_t = 0) = P\left(X_{t/n} = 0; X_{2t/n} - X_{t/n} = 0; ...; X_t - X_{(n-1)t/n} = 0\right) \overset{\text{H3}}{=}$$

$$P\left(X_{t/n} = 0)P(X_{2t/n} - X_{t/n} = 0)...P(X_t - X_{(n-1)t/n} = 0\right) \overset{\text{H2}}{=}$$
$$\left[PX_{t/n} = 0)\right]^n = \left[p_0(t/n)\right]^n.$$

Entonces obtuvimos que $p_0(t) = [p_0(t/n)]^n$ para todo $t > 0$. Entonces, para todo m natural tenemos que $p_0(mt) = [p_0(mt/n)]^n$, pero por otro lado como el intervalo $[0, mt]$ lo podemos partir en m intervalos de igual longitud t, también se cumple que $p_0(mt) = [p_0(t)]^m$. Entonces $[p_0(t)]^m = [p_0(mt/n)]^n$, por lo que $[p_0(t)]^{m/n} = p_0(mt/n)$

para todos $t > 0$, m y n naturales. Hacemos $t = 1$ y obtenemos $[p_0(1)]^{m/n} = p_0(m/n)$ para todos m y n naturales. Tomando límites, se deduce que $[p_0(1)]^t = p_0(t)$ para todo $t > 0$. Asumiendo que $0 < p_0(1) < 1$, existe $\lambda > 0$, tal que $p_0(1) = e^{-\lambda}$ y entonces $p_0(t) = e^{-\lambda t}$ para todo $t > 0$.✓

Teorema 3.36. *Bajo las hipótesis H1, H2, H3 y H4, se cumple que*

$$p_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \text{ para todo } t > 0 \text{ y } n = 0, 1, 2, 3, ...$$

Demostración.
Sabemos que $p_0(t) = e^{-\lambda t} = 1 - \lambda t + o(t)$. Como además por H4 $P(X_t \geq 2) = o(t)$, se deduce que

$$p_1(t) = P(X_t = 1) = 1 - p_0(t) - P(X_t \geq 2) = \lambda t + o(t).$$

Entonces para cada $h > 0$ tenemos que

$$p_n(t+h) = P(X_{t+h} = n) =$$

$$P(X_t = n; X_{t+h} - X_t = 0) + P(X_t = n-1; X_{t+h} - X_t = 1) +$$

$$\sum_{i=2}^{n} P(X_t = n-i; X_{t+h} - X_t = i).$$

Ahora, observamos que

$$\sum_{i=2}^{n} P(X_t = n-i; X_{t+h} - X_t = i) \leq P(X_{t+h} - X_t \geq 2) = 1 - p_0(h) - p_1(h) = o(h).$$

Entonces

$$p_n(t+h) = P(X_t = n; X_{t+h} - X_t = 0) + P(X_t = n-1; X_{t+h} - X_t = 1) + o(h) \overset{\text{H3}}{=}$$

$$P(X_t = n)P(X_{t+h} - X_t = 0) + P(X_t = n-1)P(X_{t+h} - X_t = 1) + o(h) \overset{\text{H2}}{=}$$

$$p_n(t)p_h(0) + p_{n-1}(t)\,p_1(t) + o(h) =$$

$$p_n(t)(1 - \lambda h + o(h)) + p_{n-1}(t)(\lambda h + o(h)) + o(h).$$

Y como $p_{n-1}(t)$ y $p_n(t)$ son probabilidades, son acotadas, por lo que multiplicadas por $o(h)$ dan $o(h)$ y por lo tanto podemos asegurar que

$$p_n(t+h) = p_n(t)(1 - \lambda h) + p_{n-1}(t)\,\lambda h + o(h).$$

Si restamos a ambos términos $p_n(t)$ y dividimos entre h obtenemos

$$\frac{p_n(t+h) - p_n(t)}{h} = p_{n-1}(t)\,\lambda - p_n(t)\,\lambda + \frac{o(h)}{h}$$

si ahora tomamos límite cuando $h \to 0$ obtenemos la relación

$$p_n'(t) = p_{n-1}(t)\,\lambda - p_n(t)\,\lambda.$$

Observemos que conociendo la función $p_{n-1}(t)$, tenemos una ecuación diferencial lineal de primer orden con condición inicial $p_n(0) = 0$. Como conocemos $p_0(t) = e^{-\lambda t}$, podemos hallar $p_1(t)$, luego $p_2(t)$ y así sucesivamente. Se deja como ejercicio verificar por inducción que la solución es $p_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$.✓

Observación 3.37. *La ecuación $p_n'(t) = p_{n-1}(t)\,\lambda - p_n(t)\,\lambda$ en el caso $n = 0$ queda $p_0'(t) = -p_0(t)\,\lambda$ que junto con la condición inicial $p_0(0) = 0$ da por solución $p_0(t) = e^{-\lambda t}$. Por lo tanto si en H1 no pedimos que $0 < p_0(1) < 1$ y a cambio pedimos que $p_1(t) = \lambda t + o(t)$, obtenemos una demostración del resultado, sin necesidad del lema previo.*

3.5. Variables aleatorias absolutamente continuas.

Definición 3.38. Variables aleatorias absolutamente continuas.
Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Diremos que $X : \Omega \to \mathbb{R}$ es una variable aleatoria absolutamente continua si y sólo si existe una función $f_X : \mathbb{R} \to \mathbb{R}$ tal que $f_X(x) \geq 0$ para todo $x \in \mathbb{R}$ y $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$. A la función f_X se le denomina densidad de X.

Teorema 3.39. *Si X es absolutamente continua y A es un boreliano cualquiera, entonces*

$$P(X \in A) = \int_A f_X.$$

La demostración del teorema surge de la teoría de la medida, pero es evidente si consideramos como conjunto A a un intervalo $(a, b]$ cualquiera, ya que sabemos que

$$P(X \in (a, b]) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X - \int_{-\infty}^{a} f_X = \int_{a}^{b} f_X.$$

Como los conjuntos de la forma $(a, b]$ generan la σ-álgebra de Borel, por un argumento de teoría de medida se extiende la igualdad para todo A boreliano.

Observación 3.40. *Cuando decimos $\int_A f_X$, nos estamos refiriendo a la integral de Lebesgue, ya que la integral de Riemann está definida únicamente sobre intervalos, de todas formas la integral de Lebesgue coincide con la de Riemann sobre intervalos.*

Observación 3.41. *Si X es absolutamente continua, entonces*

$$\int_{-\infty}^{+\infty} f_X = 1.$$

Observación 3.42. *Si X es absolutamente continua entonces*

$$P(X = a) = 0 \text{ cualquiera sea } a.$$

Observación 3.43. *Si X es absolutamente continua entonces F_X es continua ya que $F_X(x^-) = F_X(x) - P(X = x) = F_X(x)$.*

Observación 3.44. *Si x es punto de continuidad de f_X, entonces F_X es derivable en x y además $F'_X(x) = f_X(x)$.*

Observación 3.45. *Dada una función de densidad, si cambiamos la definición de la misma en un conjunto de puntos de medida nula, no cambia la función de distribución, ya que la integral sobre este conjunto valdrá cero.*

Observación 3.46. *Si $f : \mathbb{R} \to \mathbb{R}$ es tal que $f(x) \geq 0$ para todo $x \in \mathbb{R}$ y cumple $\int_{-\infty}^{+\infty} f(x)dx = 1$, entonces existe un espacio de probabilidad $\big(\Omega, \mathcal{A}, P\big)$ y una variable aleatoria X absolutamente continua tal que $f_X = f$. Lo anterior se debe a que definiendo $F : \mathbb{R} \to \mathbb{R}$ tal que $F(x) = \int_{-\infty}^{x} f(t)dt$, entonces, F es monótona creciente, continua en todo punto, con límites 1 y 0 a $+\infty$ y $-\infty$ respectivamente. Luego aplicamos el teorema de existencia de un espacio de probabilidad para estos casos.*

3.6. Ejemplos de variables absolutamente continuas.

Ejemplo 3.47. Variable uniforme en el intervalo $[a,b]$. **Notación:** $X \sim U\,[a,b]$.

Cuando X es tal que $f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } \; x \in (a,b) \\ 0 & \text{si } \; x \notin (a,b) \end{cases}$ se dice que X tiene distribución uniforme en el intervalo $[a,b]$. En este caso $F_X(x) = \begin{cases} 0 & \text{si } \; x \leq a \\ \frac{x-a}{b-a} & \text{si } \; a \leq x \leq b \\ 1 & \text{si } \; x \geq b \end{cases}$ y se observa que si elegimos c, d, e, f tales que $a < c < d < b$, $a < e < f < b$, con $d - c = f - e$, entonces

$$P\,(c < X < d) = F_X(d) - F_X(c) = \frac{d-c}{b-a} = \frac{f-e}{b-a} = P\,(e < X < f)$$

por lo que intervalos incluidos en $[a,b]$ de igual longitud tienen igual probabilidad.

Ejemplo 3.48. Variable Exponencial de parámetro $\lambda > 0$. **Notación:** $X \sim$**Exp**(λ).

Cuando X es tal que $f_X(x) = \begin{cases} 0 & \text{si } \; x < 0 \\ \lambda e^{-\lambda x} & \text{si } \; x \geq 0 \end{cases}$ se dice que X tiene distribución exponencial de parámetro λ. En este caso $F_X(x) = \begin{cases} 0 & \text{si } \; x < 0 \\ 1 - e^{-\lambda x} & \text{si } \; x \geq 0 \end{cases}$.

Ejemplo 3.49. Variable Normal de parámetros μ **y** $\sigma^2 > 0$. **Notación:** $X \sim N\,(\mu, \sigma^2)$.

Cuando X es tal que $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}\left(\frac{x-\mu}{\sigma}\right)^2}$ se dice que X tiene distribución normal con media μ y varianza σ^2. Veremos que ésta función es una densidad. Dado que es positiva, basta ver que integra uno. Observamos que haciendo el cambio de variable $t = \frac{x-\mu}{\sigma}$, obtenemos que $\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{\frac{-1}{2\sigma^2}(x-\mu)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{-1}{2}t^2} dt$, por lo que bastará con probar que es equivalente a probar que $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{-1}{2}t^2} dt = 1$. Calculemos $\iint\limits_{\mathbb{R}^2} e^{\frac{-1}{2}\left(x^2+y^2\right)} dxdy$. Dado que la integral es convergente, es igual a $\lim\limits_{n\to+\infty} \iint\limits_{D_n} e^{\frac{-1}{2}\left(x^2+y^2\right)} dxdy$ siendo $D_n = \left\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq n^2\right\}$.
Pasando a coordenadas polares, obtenemos que

$$\iint\limits_{D_n} e^{\frac{-1}{2}\left(x^2+y^2\right)} dxdy = \int_0^{2\pi} d\varphi \int_0^n e^{-r^2/2} r dr = 2\pi \left(1 - e^{-n^2/2}\right) \to 2\pi.$$

Por lo tanto, tenemos que

$$2\pi = \iint\limits_{\mathbb{R}^2} e^{\frac{-1}{2}\left(x^2+y^2\right)} dxdy = \int_{-\infty}^{+\infty} e^{-x^2/2} dx \int_{-\infty}^{+\infty} e^{-y^2/2} dy = \left(\int_{-\infty}^{+\infty} e^{-x^2/2} dx\right)^2$$

entonces,

$$\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}.$$

3.7. Variables aleatorias mixtas.

Existen variables aleatorias que no son discretas ni absolutamente continuas. A este tipo de variables se les suele llamar mixtas. Para construir un ejemplo de una variable de este tipo, basta considerar una función de $\mathbb{R}$ en $\mathbb{R}$, con límites 0 y 1 a menos y más infinito respectivamente, creciente y continua por derecha, tal que tenga un sólo punto de discontinuidad, con un salto menor estricto que 1. Un ejemplo concreto de esta situación se puede obtener en el siguiente ejemplo.

Ejemplo 3.50. Dada $X \sim U(0,1)$, definimos $Y = \max\{X, 1/2\}$.

$$F_Y(y) = P(Y \leq y) = P(\max\{X, 1/2\} \leq y) = P(X \leq y, 1/2 \leq y) =$$

$$\left\{ \begin{array}{cc} P(X \leq y) & \text{si } 1/2 \leq y \\ P(\Phi) & \text{si } 1/2 > y \end{array} \right. = \left\{ \begin{array}{cc} F_X(y) & \text{si } 1/2 \leq y \\ 0 & \text{si } 1/2 > y \end{array} \right. = \left\{ \begin{array}{ccc} 0 & \text{si} & y < 1/2 \\ y & \text{si} & 1/2 \leq y \leq 1 \\ 1 & \text{si} & y > 1 \end{array} \right.$$

Por lo tanto, observando que $P(Y = 1/2) = F_Y(1/2) - F_Y(1/2^-) = 1/2$ (lo cual nos asegura que Y no es absolutamente continua) y que $P(Y = y) = 0$ para todo $y \neq 1/2$ se deduce que Y tampoco puede ser discreta.

Capítulo 4

Distribución conjunta.

Definición 4.1. Dadas $X_1, X_2, ..., X_k$ variables aleatorias sobre un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, se define la distribución del vector aleatorio $(X_1, X_2, ..., X_k)$ (o también la distribución conjunta de las variables $X_1, X_2, ..., X_k$) como la función

$$F_{X_1,X_2,...,X_k} : \mathbb{R}^k \to \mathbb{R} \text{ tal que}$$

$$F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) := P(X_1 \leq x_1, X_2 \leq x_2, ..., X_k \leq x_k).$$

Como siempre, el suceso $\{X_1 \leq x_1, X_2 \leq x_2, ..., X_k \leq x_k\}$ es la abreviación de

$$\{w \in \Omega : X_1(w) \leq x_1;\ X_2(w) \leq x_2; ....; X_k(w) \leq x_k\} = \bigcap_{i=1}^{k} X_i^{-1}((-\infty, x_i]).$$

Veremos en lo que sigue diversas propiedades de las distribuciones conjuntas.

4.1. Propiedades.

Teorema 4.2. *Fijado i, mirando $F_{X_1,X_2,...,X_k} : \mathbb{R} \to \mathbb{R}$ como función únicamente de la variable x_i (dejando las demás fijas), entonces $F_{X_1,X_2,...,X_k}$ es continua por derecha y monótona creciente.*

Teorema 4.3. $\underset{x_1,x_2,...,x_k \to +\infty}{lim} F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = 1.$

Teorema 4.4. $\underset{\text{algún } x_i \to -\infty}{lim} F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = 0.$

Teorema 4.5. $\underset{x_1 \to +\infty}{lim} F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = F_{X_2,...,X_k}(x_2, ..., x_k).$

Observación 4.6. *Usando esta propiedad, $k-1$ veces, obtenemos la distribución de cada variable X_i haciendo tender todas las demás a $+\infty$.*

Teorema 4.7.

$$\underset{x_1,x_2,...,x_{i-1},x_{i+1},...,x_k \to +\infty}{lim} F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = F_{X_i}(x_i) \text{ para todo } i = 1, 2, 3, ..., k.$$

Las demostraciones de estas propiedades se realizan de manera similar al caso univariado, haremos como ejemplo el teorema 1.3.
Dado que $F_{X_1,X_2,...,X_k}$ es monótona creciente como función de cada variable, basta hallar el límite sobre alguna sucesión en particular en cada variable. Por ello, definimos los conjuntos $A_n = \bigcap_{i=1}^{k} X_i^{-1}\left((-\infty, n]\right)$. Observamos que la sucesión de conjuntos $\{A_n\}_{n\in\mathbb{N}}$ crece a Ω, luego por la propiedad de continuidad de las probabilidades se deduce que

$$\lim_{n\to+\infty} F_{X_1,X_2,...,X_k}(n,n,...,n) = \lim_{n\to+\infty} P\left(A_n\right) = P\left(\bigcup_{n=1}^{+\infty} A_n\right) = P\left(\Omega\right) = 1. \checkmark$$

Teorema 4.8. *Si para cada $p \in \mathbb{R}^k$, $i = 1,2,3,...,k$ y $h_1, h_2, ..., h_k \in \mathbb{R}^+$ definimos el operador $\Delta_{h_i}^{(i)} F_X\left(p\right) = F_X\left(p + h_i e_i\right) - F_X(p)$, (donde $e_1, e_2, ..., e_k$ son los vectores de la base canónica de $\mathbb{R}^k$) entonces*

$$\Delta_{h_k}^{(k)} \Delta_{h_{k-1}}^{(k-1)} ... \Delta_{h_1}^{(1)} F_X\left(p\right) \geq 0.$$

Observamos que en el caso bivariado, tenemos que

$$P\left(a < X \leq b;\ c < Y \leq d\right) = F_{X,Y}\left(b,d\right) - F_{X,Y}\left(b,c\right) - F_{X,Y}\left(a,d\right) + F_{X,Y}\left(a,c\right).$$

Demostración.
Se deja como ejercicio. Sugerencia, probar por inducción que

$$\Delta_{h_k}^{(k)} \Delta_{h_{k-1}}^{(k-1)} ... \Delta_{h_1}^{(1)} F_X\left(p\right) =$$

$$\sum_{\varepsilon_1,\varepsilon_2,...,\varepsilon_k \in \{0,1\}} (-1)^{k-\sum_{i=1}^{k}\varepsilon_i} F_X\left(p_1 + \varepsilon_1 h_1, p_2 + \varepsilon_2 h_2, ..., p_k + \varepsilon_k h_k\right) =$$

$$P\left(p_1 < X_1 \leq p_1 + h_1,\ p_1 < X_2 \leq p_2 + h_2, ...,\ p_k < X_k \leq p_k + h_k\right) \geq 0. \checkmark$$

Como en el caso univariado, podríamos preguntarnos cuándo una función $F : \mathbb{R}^k \to \mathbb{R}$ es la función de distribución de un vector $(X_1, X_2, ..., X_k)$ en cierto espacio de probabilidad. Nuevamente, definiríamos la terna $\left(\mathbb{R}^k, \mathcal{B}, P^*\right)$ definiendo $P^*\left(A\right)$ de tal modo que
$P^*\left((-\infty, x_1] \times (-\infty, x_2] \times ... \times (-\infty, x_k]\right) = F\left(x_1, x_2, ..., x_k\right)$. Para ello necesitamos nuevamente del teorema de extensión de medidas. Esto es posible cuando F cumple las siguientes propiedades: **i)** F es continua por derecha y monótona creciente como función de cada una de sus variables, **ii)** $\lim_{x_1,x_2,...,x_k\to+\infty} F\left(x_1, x_2, ..., x_k\right) = 1$, **iii)** $\lim_{\text{algún } x_i\to-\infty} F\left(x_1, x_2, ..., x_k\right) = 0$, **iv)** $\Delta_{h_k}^{(k)} \Delta_{h_{k-1}}^{(k-1)} ... \Delta_{h_1}^{(1)} F\left(p\right) \geq 0$ para todo $p \in \mathbb{R}^k$ y $h_1, h_2, ..., h_k \in \mathbb{R}^+$.

Observación 4.9. *En el caso en que $k = 1$, se tiene que la condición iv) se cumple automáticamente ya que queda $F(b) - F(a)$ para $a < b$ condición que se satisface al ser F monótona creciente.*

Teorema 4.10. *Si $F : \mathbb{R}^k \to \mathbb{R}$ cumple las propiedades i) ii) iii) y iv) entonces, existe un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y un vector aleatorio $(X_1, X_2, ..., X_k)$ tales que $F_{X_1,X_2,...,X_k} = F$.*

4.2. Vectores aleatorios discretos.

Definición 4.11. Vectores aleatorios discretos.
Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, se dice que el vector aleatorio $(X_1, X_2, ..., X_k) : \Omega \to \mathbb{R}^k$ es discreto si y sólo si existe $A \subset \mathbb{R}^k$ numerable tal que $P\left((X_1, X_2, ..., X_k) \in A\right) = 1$.

Veremos ahora que un vector aleatorio es discreto si y sólo si todas sus variables componentes son discretas.

Teorema 4.12. *Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, entonces el vector aleatorio $(X_1, X_2, ..., X_k)$ es discreto si y sólo si X_i es discreta para todo $i = 1, 2, 3, ..., k$.*

Demostración.
$\Rightarrow$) Existe $A \subset \mathbb{R}^k$ numerable tal que $P\left((X_1, X_2, ..., X_k) \in A\right) = 1$. Entonces definimos $A_1 := \pi_1(A)\,, A_2 := \pi_2(A)\,, ..., A_k := \pi_k(A)$ como las proyecciones sobre cada una de las componentes, es decir $\pi_i : \mathbb{R}^k \to \mathbb{R}$ tal que $\pi_i(x_1, x_2, ..., x_k) = x_i$ para cada $i = 1, 2, 3, ..., k$.
Observando que, para todo $i = 1, 2, 3, ..., k$, se tiene que $\{(X_1, X_2, ..., X_k) \in A\} \subset \{X_i \in A_i\}$, entonces

$$1 = P\left((X_1, X_2, ..., X_k) \in A\right) \leq P\left(X_i \in A_i\right),$$

entonces X_i es discreta.
$\Leftarrow$) Como todas las X_i son discretas, entonces existen conjuntos $A_1, A_2, ..., A_k \subset \mathbb{R}$ numerables tales que $P\left(X_i \in A_i\right) = 1$ para todo $i = 1, 2, 3, ..., k$. Entonces definimos $A = A_1 \times A_2 \times ... \times A_k$ es numerable (por ser producto cartesiano finito de conjuntos numerables) y además, como intersección finita de conjuntos de probabilidad 1 tiene probabilidad 1, nos queda

$$P\left((X_1, X_2, ...X_k) \in A\right) = P\left(\bigcap_{i=1}^{k} \{X_i \in A_i\}\right) = 1.$$

Entonces $(X_1, X_2, ...X_k)$ es discreto.✓
De manera análoga a las variables discretas, y dado que un vector discreto toma valores en un conjunto numerable con probabilidad 1, tiene sentido definir a función de probabilidad conjunta, como la probabilidad de tomar cada uno de los valores de su recorrido.

Definición 4.13. Si $X = (X_1, X_2, ..., X_k)$ es discreto, entonces le llamamos recorrido de X al conjunto $\text{Rec}(X) = \{x = (x_1, x_2, ..., x_k) \in \mathbb{R}^k$ tales que $P(X = x) > 0\}$.

Definición 4.14. Función de probabilidad conjunta. Si $X = (X_1, X_2, ..., X_k)$ es discreto, definimos para cada $x \in \mathbb{R}^k$,

$$p_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = P(X_1 = x_1, X_2 = x_2, ..., X_k = x_k).$$

a la que le llamaremos función de probabilidad conjunta de las variables $X_1, X_2, ..., X_k$.

Observación 4.15. *Si A es boreliano en $\mathbb{R}^k$, entonces*

$$P(X \in A) = \sum_{x \in A \cap Rec(X)} p_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k).$$

Observación 4.16.

$$\sum_{x \in Rec(X)} p_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = 1.$$

Ejemplo 4.17. Vector multinomial.
Supongamos un experimento donde se repiten de forma independiente n pruebas, donde en cada una de ellas hay k resultados posibles, digamos $E_1, E_2, ..., E_k$. La probabilidad en cada prueba de que se observe el resultado E_i es p_i, para $i = 1, 2, 3, ..., k$, donde $p_1 + p_2 + ... + p_k = 1$. Se definen para este experimento las variables $X_1, X_2, ..., X_k$, como X_i = "cantidad de pruebas entre las n en que se obtuvo el resultado E_i" para $i = 1, 2, 3, ..., k$. Se dice en estos casos que el vector $(X_1, X_2, ..., X_k)$ tiene distribución multinomial con parámetros $n, p_1, p_2, ..., p_k$.
Notación. $(X_1, X_2, ..., X_k) \sim \text{Mult}(n, p_1, p_2, ..., p_k)$.
Vamos a deducir su función de probabilidad puntual.
Fijemos $x_1, x_2, ..., x_k \in \{0, 1, 2, ..., n\}$ tales que $x_1 + x_2 + ... + x_k = n$. El suceso $\{X_1 = x_1, X_2 = x_2, ..., X_k = x_k\}$ significa que de entre las n pruebas, x_1 veces se obtuvo E_1 como resultado, x_2 veces se obtuvo E_2,..., x_k veces se obtuvo E_k. La probabilidad de que las primeras x_1 veces se obtenga E_1, las siguientes x_2 veces se obtenga E_2, y así sucesivamente hasta que las últimas x_k veces se obtenga E_k, es, debido a la independencia de cada prueba, igual a $p_1^{x_1} p_2^{x_2} ... p_k^{x_k}$. Si intercambiamos de lugar el orden donde salen las x_1 veces E_1, x_2 veces E_2, ... x_k veces E_k, la probabilidad será también $p_1^{x_1} p_2^{x_2} ... p_k^{x_k}$ ya que x_1 veces aparecerá el factor p_1, x_2 veces p_2, ..., x_k veces p_k. Por lo tanto la probabilidad de $\{X_1 = x_1, X_2 = x_2, ..., X_k = x_k\}$ será $p_1^{x_1} p_2^{x_2} ... p_k^{x_k}$ multiplicado por la cantidad de formas de elegir x_1 lugares para ubicar las veces en que sale E_1, x_2 lugares para ubicar las veces en que sale E_2,..., x_k lugares para ubicar las veces en que sale E_k. Para obtener este número, debemos primero elegir x_1 lugares entre los n para ubicar los E_1, esto se puede realizar de $C^n_{x_1}$ formas, luego nos quedan $n - x_1$ lugares, disponibles, de los cuales debemos elegir x_2 para ubicar los E_2, lo cual se puede realizar de $C^{n-x_1}_{x_2}$ formas, luego quedan $n - x_1 - x_2$ lugares disponibles, de los cuales debemos elegir x_3 para ubicar los E_3, lo que se puede realizar de $C^{n-x_1-x_2}_{x_3}$ formas, y así seguimos sucesivamente.
Al final, el número de todas las combinaciones posibles es $C^n_{x_1} C^{n-x_1}_{x_2} C^{n-x_1-x_2}_{x_3} .... C^{x_k}_{x_k} = \frac{n!}{x_1! x_2! ... x_k!}$. Así obtuvimos que para todos $x_1, x_2, ..., x_k \in \{0, 1, 2, ..., n\}$ tales que $x_1 +$

$x_2 + ... + x_k = n$,

$$P\left(X_1 = x_1, X_2 = x_2, ..., X_k = x_k\right) = \frac{n!}{x_1!x_2!...x_k!} p_1^{x_1} p_2^{x_2} ... p_k^{x_k}.$$

Observación 4.18. *Si* $(X_1, X_2, ..., X_k)$ $\sim$*Mult*$(n, p_1, p_2, ..., p_k)$, *entonces* X_i $\sim$*Bin*(n, p_i) *para* $i = 1, 2, 3, ..., k$.

4.3. Vectores aleatorios absolutamente continuos.

Definición 4.19. Vectores aleatorios absolutamente continuos.
Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, se dice que el vector aleatorio $(X_1, X_2, ..., X_k) : \Omega \to \mathbb{R}^k$ es absolutamente continuo, si y sólo si existe $f_{X_1,X_2,...,X_k} : \mathbb{R}^k \to \mathbb{R}$ tal que:
i) $f_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) \geq 0$ para todo $(x_1, x_2, ..., x_k) \in \mathbb{R}^k$,
ii) $F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} ... \int_{-\infty}^{x_k} f_{X_1,X_2,...,X_k}(u_1, u_2, ..., u_k) du_1 du_2 ... du_k$.

A la función $f_{X_1,X_2,...,X_k}$ se la denomina densidad del vector $(X_1, X_2, ..., X_k)$, o también densidad conjunta de las variables $X_1, X_2, ..., X_k$.
En $\mathbb{R}^2$, se tiene que para todo $(x, y) \in \mathbb{R}^2$ (por aplicación del teorema de Fubini),

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f_{X,Y}(u, v) dv \right) du = \int_{-\infty}^{y} \left(\int_{-\infty}^{x} f_{X,Y}(u, v) du \right) dv.$$

4.3.1. Propiedades.

Teorema 4.20. *Dado un espacio de probabilidad* $(\Omega, \mathcal{A}, P)$. *Si el vector aleatorio* $(X_1, X_2, ..., X_k) : \Omega \to \mathbb{R}^k$ *es absolutamente continuo con densidad* $f_{X_1,X_2,...,X_k}$, *entonces, para todo boreliano* $A \subset \mathbb{R}^k$ *se cumple que*

$$P\left((X_1, X_2, ..., X_k) \in A\right) = \int \cdots \int_A f_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) dx_1 dx_2 ... dx_k.$$

Demostración.
Nuevamente el resultado se sigue del teorema de existencia y unicidad de extensión de medidas, ya que la propiedad es válida para todo boreliano de la forma

$$A = (-\infty, x_1] \times (-\infty, x_2] \times ... \times (-\infty, x_k]$$

y dado que los mismos generan la σ-álgebra de Borel en $\mathbb{R}^k$ se concluye la demostración.✓

Observación 4.21. *El significado de* $\int \ldots \int_A f$ *es el de la integral de Lebesgue, que en el caso en que el boreliano* A *es un producto cartesiano de intervalos o una unión disjunta de productos cartesianos de intervalos, entonces dicha integral coincide con la de Riemann.*

Observación 4.22. *Si el boreliano A tiene medida de Lebesgue nula, entonces* $P(X \in A) = 0$.

Teorema 4.23. *Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Si el vector aleatorio $(X_1, X_2, ..., X_k) : \Omega \to \mathbb{R}^k$ es absolutamente continuo con densidad $f_{X_1,X_2,...,X_k}$, entonces,*

$$\frac{\delta^k F_{X_1,X_2,...,X_k}}{\delta x_1 \delta x_2 ... \delta x_k}(x_1, x_2, ..., x_k) = f_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k)$$

igualdad válida para todos los $(x_1, x_2, ..., x_k) \in \mathbb{R}^k$ salvo en un conjunto de medida nula.

Demostración.
Basta derivar sucesivamente a la función

$$F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} ... \int_{-\infty}^{x_k} f_{X_1,X_2,...,X_k}(u_1, u_2, ..., u_k) du_1 du_2 ... du_k$$

respecto a $x_1, x_2, ..., x_k$ en todo punto de continuidad de $f_{X_1,X_2,...,X_k}$, el conjunto de puntos donde se puede realizar esta operación es el de puntos de continuidad de $f_{X_1,X_2,...,X_k}$ que son todos salvo un conjunto de medida nula.✓
En lo que sigue, responderemos a la siguiente pregunta: ¿$(X_1, X_2, ..., X_k)$ es absolutamente continuo, es equivalente a decir que cada X_i es absolutamente continua para $i = 1, 2, 3, ..., k$?

Teorema 4.24. *Dado el vector aleatorio $(X_1, X_2, ..., X_k) : \Omega \to \mathbb{R}^k$ definido sobre un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.*
Si el vector aleatorio $(X_1, X_2, ..., X_k)$ es absolutamente continuo, entonces X_i es absolutamente continua para todo $i = 1, 2, 3, ..., k$.
Además la densidad de X_i es

$$f_{X_i}(u_i) = \int \cdots \int_{\mathbb{R}^{k-1}} f_{X_1,X_2,...,X_k}(u_1, u_2, ..., u_k) du_1 du_2 ... du_{i-1} du_{i+1} ... du_k.$$

Demostración.
Sabemos que $\lim\limits_{x_1,x_2,...,x_{i-1},x_{i+1},...,x_k \to +\infty} F_{X_1,X_2,...X_k}(x_1, x_2, ..., x_k) = F_{X_i}(x_i)$ para todo $i = 1, 2, 3, ..., k$, entonces

$$\lim_{x_1,x_2,...,x_{i-1},x_{i+1},...,x_k \to +\infty} F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) =$$

$$\lim_{x_1,x_2,...,x_{i-1},x_{i+1},...,x_k \to +\infty} \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} ... \int_{-\infty}^{x_k} f_{X_1,X_2,...,X_k}(u_1, u_2, ..., u_k) du_1 du_2 .. du_k =$$

$$\text{(aplicando Fubini)} \int_{-\infty}^{x_i} \left(\int \cdots \int_{\mathbb{R}^{k-1}} f_{X_1,X_2,...,X_k}(u_1, u_2, ..., u_k) du_1 du_2 ... du_{i-1} du_{i+1} ... du_k \right) du_i$$

Entonces

$$F_{X_i}(x_i) = \int_{-\infty}^{x_i} \left(\int \dots \int_{\mathbb{R}^{k-1}} f_{X_1,X_2,\dots,X_k}(u_1, u_2, ..., u_k) du_1 du_2 ... du_{i-1} du_{i+1} ... du_k \right) du_i$$

de donde se deduce el resultado.✓

Observación 4.25. *En el caso particular en dimensión 2, el teorema anterior nos dice que si* (X, Y) *es absolutamente continuo con densidad* $f_{X,Y}$*, entonces* X *e* Y *son absolutamente continuas con densidades*

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy, \ y \ f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx.$$

El recíproco del teorema anterior no tiene por qué cumplirse, para ello consideremos el siguiente ejemplo.
Definimos (X, Y) vector en $\mathbb{R}^2$, tal que (X, Y) toma valores en la diagonal del cuadrado $[0, 1] \times [0, 1]$ con distribución uniforme. Es decir, si definimos el conjunto $D = \left\{(x, y) \in [0, 1]^2 : y = x\right\}$, entonces para todo $I \subset D$ intervalo, se cumple que $P((X, Y) \in A) = \text{long}(I)/\sqrt{2}$. Observamos en este caso que el vector (X, Y) no es absolutamente continuo, ya que toma valores en un segmento con probabilidad uno. Como un segmento tiene medida nula, toda integral doble sobre dicho conjunto vale 0. Entonces, si (X, Y) admitiera densidad, se tendría que $1 = P((X, Y) \in A) = \iint_A f_{X,Y}(x, y) = 0$. Se deja como ejercicio, hallar la distribución conjunta de (X, Y) y deducir que tanto X como Y tienen distribución uniforme en $[0, 1]$ y por lo tanto X e Y son absolutamente continuas.
Nuevamente, para que una función $f : \mathbb{R}^k \to \mathbb{R}$ sea la función de densidad de un vector $(X_1, X_2, ..., X_k)$ en algún espacio de probabilidad, se debe cumplir que:
i) $f(x) \geq 0$ para todo $x \in \mathbb{R}^k$ (alcanza que sea para todo x salvo en un conjunto de medida nula) y
ii) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} f_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) dx_1 dx_2 ... dx_k = 1$,
ya que a partir de estas dos condiciones, definiendo

$$F(x_1, x_2, ..., x_k) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} ... \int_{-\infty}^{x_k} f_{X_1,X_2,...,X_k}(u_1, u_2, ..., u_k) du_1 du_2 ... du_k$$

se deducen de manera inmediata las 4 condiciones que requiere la función F para ser la distribución de cierto vector aleatorio en cierto espacio de probabilidad.

Ejemplo 4.26. Vector normal multivariado.
Dados un vector $(\mu_1, \mu_2, ..., \mu_k) \in \mathbb{R}^k$ y una matriz $\sum$ de dimensiones $k \times k$, simétrica y definida positiva, se dice que el vector $(X_1, X_2, ..., X_k)$ tiene distribución normal multivariada con parámetros $(\mu, \sum)$ si su densidad viene dada por la fórmula

$$f_X(x) = f_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = \frac{1}{\left(\sqrt{2\pi}\right)^k \sqrt{\det\left(\sum\right)}} e^{\frac{-1}{2}(x-\mu) \sum^{-1} (x-\mu)^T}.$$

Observación 4.27. *En el caso particular en que $k = 1$ queda la distribución normal de parámetros (μ, σ^2).*

Para verificar que ésta función integra 1, basta realizar en la misma el cambio de variable $t = (x - \mu)A^{-1}$ siendo A una matriz tal que $A^2 = \sum$ (una raíz cuadrada de $\sum$) y luego observar que

$$\int \int ... \int_{\mathbb{R}^k} \frac{1}{\left(\sqrt{2\pi}\right)^k} e^{\frac{-1}{2} tt^T} dt_1 dt_2 ... dt_k =$$

$$\frac{1}{\left(\sqrt{2\pi}\right)^k} \int \int ... \int_{\mathbb{R}^k} e^{\frac{-1}{2}\left(t_1^2 + t_2^2 + ... + t_k^2\right)} dt_1 dt_2 ... dt_k =$$

$$\frac{1}{\left(\sqrt{2\pi}\right)^k} \int_{-\infty}^{+\infty} e^{\frac{-1}{2} t_1^2} dt_1 \int_{-\infty}^{+\infty} e^{\frac{-1}{2} t_2^2} dt_2 ... \int_{-\infty}^{+\infty} e^{\frac{-1}{2} t_k^2} dt_k = 1$$

ya que quedó un producto de k integrales donde cada función integrando es la densidad normal $(0, 1)$ que integra 1.
Se puede probar que cuando $X = (X_1, X_2, ..., X_k)$ es normal multivariado, entonces la distribución de cada X_i es $N(\mu_i, \sigma_i^2)$ para $i = 1, 2, 3..., k$.
El caso particular en que $k = 2$, se llama también normal bivariada, y en este caso si $\mu = (\mu_1, \mu_2)$ y $\sum = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}$, obtenemos la fórmula

$$f_{X,Y}(x, y) = \frac{e^{\frac{-1}{2\left(\sigma_1^2 \sigma_2^2 - \sigma_{1,2}^2\right)}\left(x^2\sigma_2^2 + y^2\sigma_1^2 + \sigma_1^2\mu_2^2 + \sigma_2^2\mu_1^2 - 2xy\sigma_{1,2} + 2x\mu_2\sigma_{1,2} + 2y\mu_1\sigma_{1,2} - 2x\sigma_2^2\mu_1 - 2y\sigma_1^2\mu_2 - 2\mu_1\mu_2\sigma_{1,2}\right)}}{2\pi\sqrt{\left(\sigma_1^2\sigma_2^2 - \sigma_{1,2}^2\right)}}.$$

4.4. Independencia de variables aleatorias.

Definición 4.28. Dado $\left(\Omega, \mathcal{A}, P\right)$ espacio de probabilidad, se dice que las variables aleatorias $X_1, X_2, ..., X_k$ son independientes si y sólo si para todos $A_1, A_2, ..., A_k$ borelianos, se cumple que

$$P\left(X_1 \in A_1, X_2 \in A_2, ..., X_k \in A_k\right) = P\left(X_1 \in A_1\right) P\left(X_2 \in A_2\right) ... P\left(X_k \in A_k\right).$$

Observación 4.29. *Se observa que sólo ésta igualdad ya implica que las variables tomadas de a dos o de a tres, etc son indpendientes, ya que por ejemplo para ver que X_1 y X_2 son independientes, basta considerar $A_3 = A_4 = ... = A_k = \Omega$ con lo que obtenemos $P\left(X_1 \in A_1, X_2 \in A_2\right) = P\left(X_1 \in A_1\right) P\left(X_2 \in A_2\right)$.*

Teorema 4.30. *Dado $\left(\Omega, \mathcal{A}, P\right)$ espacio de probabilidad, entonces las variables aleatorias $X_1, X_2, ..., X_k$ son independientes si y sólo si se cumple que*

$$F_{X_1, X_2, ..., X_k}\left(x_1, x_2, ..., x_k\right) = F_{X_1}\left(x_1\right) F_{X_2}\left(x_2\right) ... F_{X_k}\left(x_k\right) \textit{ para todo } \left(x_1, x_2, ..., x_k\right) \in \mathbb{R}^k.$$

Demostración.
$\Rightarrow$) Basta considerar los borelianos $A_1 = (-\infty, x_1]$, $A_2 = (-\infty, x_2]$, ..., $A_k = (-\infty, x_k]$, entonces

$$P(X_1 \in A_1, X_2 \in A_2, ..., X_k \in A_k) = F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k)$$

mientras que

$$P(X_1 \in A_1) P(X_2 \in A_2) ... P(X_k \in A_k) = F_{X_1}(x_1) F_{X_2}(x_2) ... F_{X_k}(x_k)$$

y como las variables son independientes, se obtiene la igualdad buscada.
$\Leftarrow$) La igualdad $F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = F_{X_1}(x_1) F_{X_2}(x_2) ... F_{X_k}(x_k)$ para todo $(x_1, x_2, ..., x_k) \in \mathbb{R}^k$ implica que se cumple que $P(X_1 \in A_1, X_2 \in A_2, ..., X_k \in A_k) = P(X_1 \in A_1) P(X_2 \in A_2) ... P(X_k \in A_k)$ para los borelianos en $\mathbb{R}^k$ de la forma $A_1 \times A_2 \times ... \times A_k = (-\infty, x_1] \times (-\infty, x_2] \times ... \times (-\infty, x_k]$. Luego, como esta familia de borelianos (al variar $x_1, x_2, ..., x_k$) generan la σ-álgebra de Borel en $\mathbb{R}^k$, por extensión, se deduce que la propiedad es válida para todos $A_1, A_2, ..., A_k$ borelianos. ✓
Dado que en el caso discreto determinar la distribución conjunta es equivalente a determinar la función de probabilidad conjunta, y en el caso absolutamente continuo, determinar la función de distribución es equivalente a determinar la densidad conjunta (salvo conjuntos de medida nula), se tienen los siguientes corolarios.

Corolario 4.31. *En el caso discreto, se tiene que las variables aleatorias $X_1, X_2, ..., X_k$ son independientes si y sólo si se cumple que*

$$p_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = p_{X_1}(x_1) p_{X_2}(x_2) ... p_{X_k}(x_k)$$

para todo $(x_1, x_2, ..., x_k) \in \mathbb{R}^k$.

Demostración.
$\Rightarrow$) Cualesquiera sean los reales $x_1, x_2, ..., x_k$ basta considerar los borelianos $A_1 = \{x_1\}, A_2 = \{x_2\}, ..., A_k = \{x_k\}$ y usar la definición de independencia.
$\Leftarrow$) Dados los reales $x_1, x_2, ..., x_k$, se tiene que $F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) =$

$$\sum_{t_1 \in \mathrm{Rec}(X_1)\,:\,t_1 \le x_1} ... \sum_{t_k \in \mathrm{Rec}(X_k)\,:\,t_k \le x_k} p_{X_1,X_2,...,X_k}(t_1, t_2, ..., t_k) =$$

$$\sum_{t_1 \in \mathrm{Rec}(X_1)\,:\,t_1 \le x_1} ... \sum_{t_k \in \mathrm{Rec}(X_k)\,:\,t_k \le x_k} p_{X_1}(x_1) p_{X_2}(x_2) ... p_{X_k}(x_k) =$$

$$\sum_{t_1 \in \mathrm{Rec}(X_1)\,:\,t_1 \le x_1} p_{X_1}(x_1) \sum_{t_2 \in \mathrm{Rec}(X_2)\,:\,t_2 \le x_2} p_{X_2}(x_2) ... \sum_{t_k \in \mathrm{Rec}(X_k)\,:\,t_k \le x_k} p_{X_k}(x_k) =$$

$$F_{X_1}(x_1) F_{X_2}(x_2) ... F_{X_k}(x_k). ✓$$

Corolario 4.32. *En el caso absolutamente continuo, Si* $(X_1, X_2, ..., X_k)$ *es vector absolutamente continuo, se tiene que las variables aleatorias* $X_1, X_2, ..., X_k$ *son independientes si y sólo si se cumple que*

$$f_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = f_{X_1}(x_1) f_{X_2}(x_2) ... f_{X_k}(x_k)$$

$$\textit{para todo } (x_1, x_2, ..., x_k) \in \mathbb{R}^k \textit{punto de continuidad de} f_{X_1,X_2,...,X_k}.$$

Demostración.
$\Rightarrow$) $F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = F_{X_1}(x_1)F_{X_2}(x_2)...F_{X_k}(x_k)$, para todo $(x_1, x_2, ..., x_k) \in \mathbb{R}^k$ punto de continuidad de $f_{X_1,X_2,...,X_k}$, si derivamos sucesivamente de ambos lados de la igualdad, primero respecto de x_1 luego respecto de x_2... y por último respecto de x_k, del lado izquierdo queda $f_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k)$ y del derecho queda $f_{X_1}(x_1)f_{X_2}(x_2)...f_{X_k}(x_k)$, por lo tanto la igualdad se obtiene en todo punto de $\mathbb{R}^k$, salvo en un conjunto de medida nula.
$\Leftarrow$)

$$F_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} ... \int_{-\infty}^{x_k} f_{X_1}(u_1)f_{X_2}(u_2)...f_{X_k}(u_k)du_1du_2...du_k =$$

$$\int_{-\infty}^{x_1} f_{X_1}(u_1)du_1 \int_{-\infty}^{x_2} f_{X_2}(u_2)du_2... \int_{-\infty}^{x_k} f_{X_k}(u_k)du_k =$$

$$F_{X_1}(x_1)F_{X_2}(x_2)...F_{X_k}(x_k). \checkmark$$

Definición 4.33. Dado $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad, se dice que la familia de variables aleatorias $\{X_t\}_{t\in I}$ donde I es una familia arbitraria de índices si y sólo si para todo $F \subset I$ finito, se cumple que $\{X_t\}_{t\in F}$ son independientes.

Ejemplo 4.34. Si el vector $X = (X_1, X_2, ..., X_k)$ es normal multivariado, con parámetros $(\mu, \sum)$, donde la matriz $\sum$ es diagonal, es decir cuando $\sigma_{i,j} = 0$ para todos $i \neq j$, observamos que

$$(x-\mu) \sum^{-1} (x-\mu)^T = \sum_{i=1}^{k} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2$$

por lo que la densidad conjunta queda

$$f_{X_1,X_2,...,X_k}(x_1, x_2, ..., x_k) = \frac{1}{\sqrt{2\pi\sigma_1^2\sigma_2^2...\sigma_k^2}} e^{\sum_{i=1}^{k}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2} =$$

$$\prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{\frac{-1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2}$$

por lo que se deduce que $X_1, X_2, ..., X_k$ son independientes cuyas distribuciones son $X_i \sim N\left(\mu_i, \sigma_i^2\right)$ para $i = 1, 2, 3, ..., k$. Más adelante se verá el significado de los parámetros $(\mu, \sum)$.

Teorema 4.35. ***Convolución de dos variables aleatorias.***
Dadas dos variables aleatorias independientes $X, Y : \Omega \to \mathbb{R}$ definidas sobre un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Consideremos la variable $Z = X + Y$.
Entonces:

(i) *Si X e Y son discretas, entonces Z es discreta y además*

$$p_Z(z) = \sum_{x \in Rec(X)\ z-x \in Rec(Y)} p_X(x)p_Y(z-x).$$

(ii) *Si (X, Y) es absolutamente continuo, entonces Z es absolutamente continua y además*

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx.$$

Demostración.

(i)

$$p_Z(z) = P(Z = z) = P(X + Y = z) = \sum_{x \in \mathrm{Rec}(X)} P(X + Y = z;\ X = x) =$$

$$\sum_{x \in \mathrm{Rec}(X)} P(Y = z - x;\ X = x) = \sum_{x \in \mathrm{Rec}(X),\ z-x \in \mathrm{Rec}(Y)} P(Y = z - x)\, P(X = x) =$$

$$\sum_{x \in \mathrm{Rec}(X)\ z-x \in \mathrm{Rec}(Y)} p_X(x)p_Y(z-x).$$

(ii) Si le llamamos $A = \{(x, y) \in \mathbb{R}^2 : x + y \leq z\}$, entonces

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = \iint_A f_{X,Y}(x, y)dxdy =$$

$$\iint_A f_X(x)f_Y(y)dxdy = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-x} f_X(x)f_Y(y)dy \right) dx =$$

$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-x} f_Y(y)dy \right) f_X(x)dx$$

ahora realizando en la integral en y el cambio de variable $t = y + x$ y nos queda

$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z} f_Y(t-x)dt \right) f_X(x)dx = \int_{-\infty}^{z} \left(\int_{-\infty}^{+\infty} f_X(x)f_Y(t-x)dx \right) dt.$$

Por lo tanto Z es absolutamente continua con densidad

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx. \checkmark$$

Ejemplo 4.36. Si $X \sim N(\mu_1, a^2)$, $Y \sim N(\mu_2, b^2)$ son independientes, entonces $Z = X + Y \sim N(\mu_1 + \mu_2, a^2 + b^2)$.
Basta probarlo para el caso $\mu_1 = \mu_2 = 0$, ya que si $X \sim N(\mu, \sigma^2)$,entonces $X = \mu + T$ donde $T \sim N(0, \sigma^2)$.
Aplicamos entonces la fórmula de la convolución y obtenemos que

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)dx = f_Z(z) = \frac{1}{2\pi ab} \int_{-\infty}^{+\infty} e^{\frac{-x^2}{2a^2}} e^{\frac{-(z-x)^2}{2b^2}} dx =$$

$$\frac{1}{2\pi ab} e^{\frac{-z^2}{2(a^2+b^2)}} \int_{-\infty}^{+\infty} e^{\frac{-1}{2a^2b^2}\left(x\sqrt{a^2+b^2} - \frac{za^2}{\sqrt{a^2+b^2}}\right)^2} dx.$$

Luego de hacer el cambio de variable $t = \frac{1}{ab}\left(x\sqrt{a^2+b^2} - \frac{za^2}{\sqrt{a^2+b^2}}\right)$, obtenemos que la última integral es igual a

$$\frac{1}{2\pi\sqrt{a^2+b^2}} e^{\frac{-z^2}{2(a^2+b^2)}} \int_{-\infty}^{+\infty} e^{\frac{-t^2}{2}} dx = \frac{1}{\sqrt{2\pi\left(a^2+b^2\right)}} e^{\frac{-z^2}{2(a^2+b^2)}}$$

que es la función de densidad correspondiente a una variable con distribución $N(0, a^2 + b^2)$.
Observamos que de esta propiedad, se deduce que toda combinación lineal de variables normales independientes es normal.

Ejemplo 4.37. Si $X \sim$Bin(n, p), $Y \sim$Bin(m, p) son independientes, entonces $Z = X + Y \sim$Bin$(n + m, p)$.
En este caso,

$$p_Z(z) = P(Z = z) = P(X + Y = z) =$$

$$\sum_{x=0}^{n+m} P(Y = z - x) P(X = x) = \sum_{x\,:\,x\leq n,\ z-x\leq m} C_{z-x}^m p^{z-x} (1-p)^{m-z+x} C_x^n p^x (1-p)^{n-x} =$$

$$\sum_{x\,:\,x\leq n,\ z-x\leq m} C_{z-x}^m C_x^n p^z (1-p)^{n+m-z} = p^z (1-p)^{n+m-z} \sum_{x\,:\,x\leq n,\ z-x\leq m} C_{z-x}^m C_x^n$$

Ahora, teniendo en cuenta el coeficiente que multiplica al término t^z cuando desarrollamos $(1+t)^n (1+t)^m = (1+t)^{n+m}$, obtenemos la igualdad

$$\sum_{x\,:\,x\leq n,\ z-x\leq m} C_{z-x}^m C_x^n = C_x^{n+m}$$

Por lo tanto

$$p_Z(z) = C_x^{n+m} p^z (1-p)^{n+m-z}.$$

4.5. Método del Jacobiano.

Frecuentemente, conocemos la distribución de un vector aleatorio X y debemos trabajar con una función del mismo, digamos $Y = g(X)$. Si el vector X es absolutamente continuo y la función g es diferenciable deseamos saber si Y es también absolutamente continuo, y si lo es, obtener una fórmula que nos permita hallar la densidad de Y. El siguiente teorema apunta en esa dirección.

Teorema 4.38. *Dados $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad, $X = (X_1, X_2, ..., X_k) : \Omega \to \mathbb{R}^k$ vector aleatorio y $g : U \to V$ donde U, V son abiertos de $\mathbb{R}^k$ tales que $P(X \in U) = 1$, g es biyectiva y diferenciable con $detJ_g(x) \neq 0$ para todo $x \in U$.*
Si X es absolutamente continuo entonces $Y = g(X)$ es absolutamente continuo con densidad conjunta dada por

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \frac{1}{|detJ_g\left(g^{-1}(y)\right)|} \mathbf{1}_V(y).$$

Demostración.
Basta ver que para todo boreliano B en $\mathbb{R}^k$, se puede expresar $P(Y \in B)$ como una integral sobre el conjunto B de cierta función, la cual será necesariamente (salvo conjuntos de medida nula) la densidad del vector Y.

$$P(Y \in B) = P(g(X) \in B) = P\left(X \in g^{-1}(B)\right) = \int \cdots \int_{g^{-1}(B) \cap U} f_X(x) dx_1 dx_2 ... dx_k.$$

Ahora, realizando el cambio de variable $y = g(x)$ en la integral nos queda

$$\int \cdots \int_{B \cap V} f_X(g^{-1}(y)) \frac{1}{|\det J_g\left(g^{-1}(y)\right)|} dy_1 dy_2 ... dy_k =$$

$$\int \cdots \int_{B} f_X\left(g^{-1}(y)\right) \frac{1}{|\det J_g\left(g^{-1}(y)\right)|} \mathbf{1}_V(y) dy_1 dy_2 ... dy_k. \checkmark$$

En el caso particular en que $k = 1$ tenemos el siguiente corolario.

Corolario 4.39. *Dados $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad, $X : \Omega \to \mathbb{R}$ variable aleatoria y $g : U \to V$ donde U, V son abiertos de $\mathbb{R}$ tales que $P(X \in U) = 1$, g es biyectiva y derivable, con $g'(x) \neq 0$ para todo $x \in U$.*
Si X es absolutamente continua entonces $Y = g(X)$ es absolutamente continua con densidad dada por

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \frac{1}{|g'\left(g^{-1}(y)\right)|} \mathbf{1}_V(y).$$

Ejemplo 4.40. Como aplicación veremos que si $X, Z \sim N(0, 1)$ independientes, y definimos $Y = |Z|$ entonces probaremos que $X^2 + Y^2 \sim$Exp$(\lambda = 1/2)$.

En primer lugar observamos que, para $y > 0$, se tiene que $F_Y(y) = P(|Z| \leq y) = P(-y \leq Z \leq y) = F_Z(y) - F_Z(-y) = 2F_Z(y) - 1$, por lo tanto $f_Y(y) = 2f_Z(y)\mathbf{1}_{\{y>0\}} = \frac{2}{\sqrt{2\pi}}e^{\frac{-y^2}{2}}\mathbf{1}_{\{y>0\}}$. También vemos que $P\left((X,Y) \in \mathbb{R} \times \mathbb{R}^+\right) = 1$.
Consideramos la función $g : \mathbb{R} \times \mathbb{R}^+ \to V$ siendo $V = \{(u,v) \in \mathbb{R}^2 : v > u^2\}$ tal que $g(x,y) = (x, x^2 + y^2)$. Esta función es invertible y su inversa es $g^{-1}(w,t) = \left(w, \sqrt{t - w^2}\right)$. $\det J_g(x,y) = 2y$.
Dado que X e Y son independientes, se tiene que su densidad conjunta es $f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{\pi}e^{\frac{-1}{2}\left(x^2+y^2\right)}\mathbf{1}_{\{y>0\}}$.
La densidad conjunta de $(W,T) = g(X,Y) = (X, X^2 + Y^2)$ será entonces

$$f_{W,T}(w,t) = f_{X,Y}(g^{-1}(w,t))\frac{1}{|\det J_g(g^{-1}(w,t))|}\mathbf{1}_V(w,t) = \frac{1}{\pi}e^{\frac{-t}{2}}\frac{1}{2\sqrt{t-w^2}}\mathbf{1}_V(w,t).$$

Hallamos la densidad de $T = X^2 + Y^2$ a partir de la densidad conjunta como

$$f_T(t) = \int_{-\infty}^{+\infty} f_{W,T}(w,t)du \overset{\text{si } t>0}{=} \int_{-\sqrt{t}}^{\sqrt{t}} \frac{1}{\pi}e^{\frac{-v}{2}}\frac{1}{2\sqrt{t-w^2}}dw$$

luego, realizando el cambio de variable $u = \sqrt{t}sen\varphi$ obtenemos $f_T(t) = \frac{1}{2}e^{-t/2}$ y, dado que para $t < 0$, se tiene $f_T(t) = 0$, se deduce que

$$f_T(t) = \frac{1}{2}e^{-t/2}\mathbf{1}_{\{t>0\}}$$

por lo que $V = X^2 + Y^2 \sim \text{Exp}(\lambda = 1/2)$.
Ejercicio.
Si X e Y son independientes con distribución exponencial de parámetro $\lambda = 1$. Hallar la densidad conjunta del vector $(X+Y, X-Y)$.

Capítulo 5

Integral de Riemann-Stieltjes.

Dadas funciones $g,\ F:[a,b]\to\mathbb{R}$ que cumplan ciertos requisitos, definiremos la expresión $\int_a^b g(x)dF(x)$ de tal manera que cuando consideremos el caso particular en que $F(x)=x$ nos quede la definición clásica de integral de Riemann. Definimos una partición del intervalo $[a,b]$ como el conjunto finito $P=\{a=x_0,x_1,....,x_n=b\}$ donde $x_{i-1}<x_i$ para todo $i=1,2,...,n$. Junto con la partición, elegimos para cada $i=1,2,...,n$, puntos intermedios $c_i\in[x_{i-1},x_i]$. Es decir que dar la partición P equivale a dar los puntos de subdivisión x_i y los puntos intermedios c_i.

Definición 5.1. Dadas $g,\ F:[a,b]\to\mathbb{R}$ y P partición (con sus correspondientes puntos intermedios c_i), definimos la suma parcial de Riemann-Stieltjes como

$$S(P,g,F)=\sum_{i=1}^{n} g(c_i)\left(F(x_i)-F(x_{i-1})\right).$$

Observamos que cuando $F(x)=x$, si le pedimos a g que sea integrable Riemann, dichas sumas "se acercarán" indefinidamente al valor $\int_a^b g(x)dx$ conforme "afinemos suficientemente" la partición, en esa dirección apuntaremos.

Definición 5.2. Dada P partición en $[a,b]$ definimos $\|P\|=$máx$\{x_i-x_{i-1},\ i=1,2,...,n\}$ y le llamaremos norma de la partición.

Definición 5.3. Dadas $g,\ F:[a,b]\to\mathbb{R}$, diremos que $\lim\limits_{\|P\|\to 0} S(P,g,F)=I$ si y sólo si dado $\varepsilon>0$, existe $\delta>0$ tal que para toda P partición de $[a,b]$ (con sus correspondientes puntos intermedios c_i) con $\|P\|<\delta$, se cumple que $|S(P,g,F)-I|<\varepsilon$.

Definición 5.4. Integral de Riemann-Stieltjes.
Dadas $g,\ F:[a,b]\to\mathbb{R}$, si existe y es finito $\lim\limits_{\|P\|\to 0} S(P,g,F)=I$, diremos que la integral de Riemann-Stieltjes de g respecto de F en el intervalo $[a,b]$ existe y vale I.

$$\text{Notación: } \int_a^b gdF=\int_a^b g(x)dF(x).$$

Observación 5.5. *En el caso particular en que $F(x) = x$, la definición coincide con la definición de función integrable Riemann en $[a, b]$.*

Se deja como ejercicio verificar el enunciado de los ejemplos que siguen.

Ejemplo 5.6. Si $F(x) = k$ constante, entonces cualquiera sea $g : [a, b] \to \mathbb{R}$ existe $\int_a^b gdF$ y además $\int_a^b gdF = 0$.

Ejemplo 5.7. Si $g : [a, b] \to \mathbb{R}$ es continua, $F(x) = \mathbf{1}_{[c,b]} = \begin{cases} 1 & \text{si } x \in [c, b] \\ 0 & \text{si no} \end{cases}$ con $c \in (a, b)$ existe $\int_a^b gdF$ y además $\int_a^b gdF = g(c)$.

Ejemplo 5.8. Si $g(x) = F(x) = \mathbf{1}_{[a,c]} = \begin{cases} 0 & \text{si } x \in [a, c] \\ 1 & \text{si no} \end{cases}$ con $c \in (a, b)$ entonces no existe $\int_a^b gdF$.

Ejemplo 5.9. Si $g(x) = k$ constante, entonces existe $\int_a^b gdF$ para cualquier F y vale $\int_a^b kdF(x) = k\,(F(b) - F(a))$.

Veremos en lo que sigue un par de caracterizaciones para la existencia de $\int_a^b gdF$.

Teorema 5.10. *Los siguientes enunciados son equivalentes.*

(a) Existe $\lim\limits_{\|P\| \to 0} S\,(P, g, F)$ y vale I (finito).

(b) Condición de Cauchy.
Dado $\varepsilon > 0$, existe $\delta > 0$ tal que si P y Q son dos particiones de $[a, b]$ tales que $\|P\| < \delta$ y $\|Q\| < \delta$, se cumple que $|S\,(P, g, F) - S\,(Q, g, F)| < \varepsilon$.

(c) Para toda sucesión$\{P_n\}$ de particiones en $[a, b]$ tales que $\|P_n\| \to 0$ se cumple que $\lim\limits_{n\to+\infty} S\,(P_n, g, F) = I$.

Demostración.
$(a) \Rightarrow (b)$ Dado $\varepsilon > 0$, existe $\delta > 0$ tal que para toda P partición de $[a, b]$ (con sus correspondientes puntos intermedios c_i) tal que $\|P\| < \delta$, se cumple que $|S\,(P, g, F) - I| < \varepsilon/2$. Entonces si tomamos P y Q dos particiones de $[a, b]$ tales que $\|P\| < \delta$ y $\|Q\| < \delta$, se cumplirá que

$$|S\,(P, g, F) - S\,(Q, g, F)| \leq |S\,(P, g, F) - I| + |S\,(Q, g, F) - I| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

$(b) \Rightarrow (c)$ Fijamos $\{P_n\}$ sucesión de particiones en $[a, b]$ tales que $\|P_n\| \to 0$. Dado $\varepsilon > 0$, tomamos el $\delta > 0$ de la condición de Cauchy, y por lo tanto existirá un n_0 tal que $\|P_n\| < \delta$ para todo $n \geq n_0$. Entonces si consideramos $n, m \geq n_0$, obtendremos que $|\ S\,(P_n, g, F) - S\,(P_m, g, F)| < \varepsilon$ por lo que la sucesión $\{S\,(P_n, g, F)\}$ es de Cauchy, entonces existirá $I \in R$ tal que $\lim\limits_{n\to+\infty} S\,(P_n, g, F) = I$.
Observamos que el valor de I depende de la elección de la sucesión de particiones, faltaría probar que el límite es el mismo cualquiera sea la sucesión de particiones.

Consideremos entonces $\{P'_n\}$ otra sucesión de particiones en $[a,b]$ tales que $\|P'_n\| \to 0$ y sea I' tal que $\lim_{n\to+\infty} S(P'_n, g, F) = I'$. Consideramos entonces la siguiente sucesión de particiones: $P_1, P'_1, P_2, P'_2, ..., P_n, P'_n, ...$ entonces es claro que esta nueva sucesión, llamémosle $\{Q_n\}$, cumple que $\|Q_n\| \to 0$ y por lo tanto existe I'' tal que $\lim_{n\to+\infty} S(Q_n, g, F) = I''$. Pero $\{S(P_n, g, F)\}$ y $\{S(P'_n, g, F)\}$son subsucesiones de $\{S(Q_n, g, F)\}$ y por lo tanto $I = I' = I''$.
$(c) \Rightarrow (a)$ Supongamos por absurdo que (a) no es cierto, entonces existe $\varepsilon > 0$ tal que para todo $\delta > 0$, existe una partición P_δ, tal que $|S(P_\delta, g, F) - I| \geq \varepsilon$. Tomando $\delta = 1/n$, encontramos una sucesión de particiones $\{P_n\}$ tal que para todo n, $|S(P_n, g, F) - I| \geq \varepsilon$ entonces $\lim_{n\to+\infty} S(P_n, g, F) \neq I$. ✓

Teorema 5.11. *Si $g : [a,b] \to \mathbb{R}$ es continua y $F : [a,b] \to \mathbb{R}$ es monótona, entonces existe $\int_a^b gdF$.*

Demostración.
Probaremos que se cumple la condición de Cauchy. Fijamos $\varepsilon > 0$. Como g es uniformemente continua en $[a,b]$ existe $\delta > 0$ tal que si $|x-y| < \delta$ entonces $|g(x) - g(y)| < \frac{\varepsilon}{F(b)-F(a)}$.Tomamos una partición $P = \{a, x_1, x_2, ..., x_{n-1}, b\}$ con puntos intermedios

$c_i \in [x_{i-1}, x_i]$ $i = 1, 2, ..., n$ y una partición $Q = \{a, y_1, y_2, ..., y_{m-1}, b\}$ con puntos intermedios $d_i \in [y_{i-1}, y_i]$ $i = 1, 2, ..., m$. $S(P, g, F) = \sum_{i=1}^n g(c_i)(F(x_i) - F(x_{i-1}))$, $S(Q, g, F) = \sum_{i=1}^m g(d_i)(F(y_i) - F(y_{i-1}))$.
Unimos los puntos que forman la partición P con la de Q, a la que le llamamos $\{a, z_1, z_2, ..., z_{k-1}, b\}$ ($k \leq n+m-1$ pues algunos puntos de P pueden coincidir con algunos de Q). Podemos escribir entonces

$$S(P, g, F) = \sum_{i=1}^k g(c'_i)(F(z_i) - F(z_{i-1})) \text{ y } S(Q, g, F) = \sum_{i=1}^k g(d'_i)(F(z_i) - F(z_{i-1}))$$

donde los c'_i son los mismos que los c_i (más explícitamente, cuando $[z_{j-1}, z_j] \subset [c_{i-1}, c_i]$ entonces $c'_j = c_i$). Análogamente, d'_i son los mismos que los d_i. Observamos que$|c'_i - d'_i| < \delta$ si le pedimos a las particiones P y Q, $\|P\| < \delta/2$ y $\|Q\| < \delta/2$. Entonces

$$|S(P, g, F) - S(Q, g, F)| = \left| \sum_{i=1}^k (g(c'_i) - g(d'_i))(F(z_i) - F(z_{i-1})) \right| \leq$$

$$\sum_{i=1}^k |g(c'_i) - g(d'_i)| \, |F(z_i) - F(z_{i-1})| =$$

$$\sum_{i=1}^k |g(c'_i) - g(d'_i)| (F(z_i) - F(z_{i-1})) \leq \sum_{i=1}^k \frac{\varepsilon}{F(b) - F(a)} (F(z_i) - F(z_{i-1})) = \varepsilon.✓$$

Nota. Con la misma idea, se puede probar que si F es monótona creciente y g es acotada y tiene una cantidad finita de discontinuidades, pero F y g no tienen discontinuidades en común, entonces existe $\int_a^b gdF$.

Teorema 5.12. *Si $g:[a,b]\to\mathbb{R}$ es continua y $F:[a,b]\to\mathbb{R}$ es monótona y derivable tal que $F'(x)=f(x)$ para todo $x\in[a,b]$, siendo f integrable Riemann en $[a,b]$, entonces*

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

Demostración.
Dada una partición P de $[a,b]$, existen $d_i\in[x_{i-1},x_i]$ $i=1,2,...,n$ tales que $F(x_i)-F(x_{i-1})=f(d_i)(x_i-x_{i-1})$, ahora si elegimos como puntos intermedios de la partición a los d_i, obtenemos

$$S(P,g,F)=\sum_{i=1}^{n}g(d_i)\left(F(x_i)-F(x_{i-1})\right)=\sum_{i=1}^{n}g(d_i)f(d_i)(x_i-x_{i-1}).$$

Tomando límite cuando $\|P\|\to 0$ se obtiene el resultado ya que la última sumatoria tiende a la integral de Riemann de $g(x)f(x)$ en $[a,b]$ (producto de funciones integrables Riemann es integrable Riemann). ✓

5.1. Propiedades.

Proposición 5.13. *Si $g,h,F:[a,b]\to\mathbb{R}$ son tales que existen las integrales $\int_a^b gdF$ y $\int_a^b hdF$ entonces también existe $\int_a^b(\alpha g+\beta h)\,dF$ cualesquiera sean $\alpha,\beta\in\mathbb{R}$ y además*

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

Demostración.
Cualquiera sea P partición de $[a,b]$, se tiene que

$$S(P,\alpha g+\beta h,F)=\sum_{i=1}^{n}\left(\alpha g(c_i)+\beta h(c_i)\right)\left(F(x_i)-F(x_{i-1})\right)=$$

$$\alpha\sum_{i=1}^{n}g(c_i)\left(F(x_i)-F(x_{i-1})\right)+\beta\sum_{i=1}^{n}h(c_i)\left(F(x_i)-F(x_{i-1})\right)=\alpha S(P,g,F)+\beta S(P,h,F)$$

por lo que tomando límite cuando $\|P\|\to 0$ se obtiene el resultado. ✓

Proposición 5.14. *Si $h,F,G:[a,b]\to\mathbb{R}$ son tales que existen las integrales $\int_a^b hdF$ y $\int_a^b hdG$ entonces también existe $\int_a^b hd(\alpha F+\beta G)$ cualesquiera sean $\alpha,\beta\in\mathbb{R}$ y además*

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

Demostración.
Cualquiera sea P partición de $[a,b]$, se tiene que

$$S(P,h,\alpha F+\beta G)=\sum_{i=1}^{n}h(c_i)\left[\alpha\left(F(x_i)-F(x_{i-1})\right)+\beta\left(G(x_i)-G(x_{i-1})\right)\right]=$$

$$\alpha\sum_{i=1}^{n}h(c_i)\left[\left(F(x_i)-F(x_{i-1})\right)\right]+\beta\sum_{i=1}^{n}h(c_i)\left[\left(G(x_i)-G(x_{i-1})\right)\right]=\alpha S(P,h,F)+\beta S(P,h,G)$$

por lo que tomando límite cuando $\|P\|\to 0$ se obtiene el resultado. ✓

Proposición 5.15. *Si g, $F:[a,b]\to\mathbb{R}$ son tales que existe $\int_a^b gdF$ entonces cualquiera sea $c\in(a,b)$, se cumple que existen $\int_a^c gdF$ y $\int_c^b gdF$ y además*

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

Demostración.
Primero probaremos que existe $\int_a^c gdF$ usando la condición de Cauchy. Como $\int_a^b gdF$ existe, fijado $\varepsilon>0$, existe $\delta>0$ tal que si P y Q son dos particiones de $[a,b]$, donde $\|P\|<\delta$ y $\|Q\|<\delta$ se cumple que $|S(P,g,F)-S(Q,g,F)|<\varepsilon$. Consideremos entonces $\widetilde{P}$ y $\widetilde{Q}$ dos particiones de $[a,c]$ tales que $\left\|\widetilde{P}\right\|<\delta$ y $\left\|\widetilde{Q}\right\|<\delta$. Completamos $\widetilde{P}$ y $\widetilde{Q}$ a P y Q particiones de $[a,b]$, agregando los mismos puntos de modo que $\|P\|<\delta$ y $\|Q\|<\delta$. Entonces $\left|S\left(\widetilde{P},g,F\right)-S\left(\widetilde{Q},g,F\right)\right|=|S(P,g,F)-S(Q,g,F)|<\varepsilon$. Por lo tanto existe $\int_a^c gdF$. Análogamente se prueba que existe $\int_c^b gdF$. Sabemos ahora que las tres integrales existen. Consideramos entonces la sucesión de particiones $\{P_n\}$ tales que $\|P_n\|\to 0$ y tales que $c\in P_n$ para todo n. Podemos escribir entonces $P_n=P_n^{(1)}\cup P_n^{(2)}$, donde $P_n^{(1)}$ es partición de $[a,c]$ con $\left\|P_n^{(1)}\right\|\to 0$ y $P_n^{(2)}$ es partición de $[a,c]$ con $\left\|P_n^{(2)}\right\|\to 0$. Entonces, se tiene que

$$S(P_n,g,F)=S\left(P_n^{(1)},g,F\right)+S\left(P_n^{(2)},g,F\right)$$

y tomando límite cuando $n\to+\infty$ se obtiene

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

Proposición 5.16. *Si g, $F:[a,b]\to\mathbb{R}$ son tales que $g\geq 0$, F es monótona creciente y existe $\int_a^b g(x)dF(x)$, entonces*

$$\int_a^b gdF\geq 0.$$

Demostración.
Cualquiera sea P partición de $[a,b]$, se tiene que

$$S(P,g,F)=\sum_{i=1}^{n} g(c_i)\left(F(x_i)-F(x_{i-1})\right)\geq 0$$

puesto que cada sumando es no negativo, entonces $\int_a^b gdF\geq 0$. ✓

Proposición 5.17. *Si $g,h,\ F:[a,b]\to\mathbb{R}$ son tales que $g\geq h$, F es monótona creciente y existen $\int_a^b gdF$, $\int_a^b hdF$, entonces*

$$\int_a^b gdF\geq\int_a^b hdF.$$

Demostración.
$g-h\geq 0$, entonces por la propiedad anterior $0\leq\int_a^b(g-h)dF=\int_a^b gdF-\int_a^b hdF$ por lo que se deduce que $\int_a^b gdF\geq\int_a^b hdF$. ✓

Proposición 5.18. *Si $g,\ F:[a,b]\to\mathbb{R}$ son tales que $\alpha\leq g(x)\leq\beta$ para todo $x\in[a,b]$, F es monótona creciente y existe $\int_a^b gdF$ entonces*

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

Demostración.
Es un corolario inmediato de la propiedad anterior. ✓

Proposición 5.19. *Si $g:[a,b]\to\mathbb{R}$ es continua y $F:[a,b]\to\mathbb{R}$ es monótona creciente, entonces*

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

Demostración.
Cualquiera sea P partición de $[a,b]$, se tiene que

$$|S(P,g,F)|=\left|\sum_{i=1}^{n} g(c_i)\left(F(x_i)-F(x_{i-1})\right)\right|\leq\sum_{i=1}^{n}|g(c_i)|\left(F(x_i)-F(x_{i-1})\right)=S(P,|g|,F).$$

Tomando límite cuando $\|P\|\to 0$ se obtiene el resultado. ✓

Proposición 5.20. ***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 $\int_a^b gdF=g(c)\left(F(b)-F(a)\right)$.

Demostración.
La existencia de la integral se debe a que g es continua y F es monótona. Como g es continua, por el toerema de Weierstrass tiene mínimo y máximo que les llamamos m y M respectivamente. Entonces por la propiedad anterior, se tiene que $m\leq\frac{\int_a^b gdF}{F(b)-F(a)}\leq M$ y como g es continua, resulta que existe $c\in[a,b]$ tal que $\frac{\int_a^b gdF}{F(b)-F(a)}=g(c)$. ✓

5.2. Métodos de integración.

Teorema 5.21. ***Fórmula de integración por partes.***
Si g, $F : [a, b] \to \mathbb{R}$ son tales que existe $\int_a^b gdF$, entonces también existe $\int_a^b Fdg$ y además

$$\int_a^b Fdg = gF\,\Big|_a^b - \int_a^b gdF.$$

Demostración.
Recordamos la fórmula de Abel:

$$\sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n-1} A_i \left(b_i - b_{i+1}\right) + A_n b_n \text{ siendo } A_n = \sum_{i=1}^{n} a_i.$$

Tomamos una partición cualquiera $P = \{a, x_1, x_2, ..., x_{n-1}, b\}$ con correspondientes puntos intermedios $c_1, c_2, ..., c_n$. Si aplicamos dicha fórmula para $S\left(P, F, g\right) = \sum_{i=1}^{n} F(c_i)\left(g\left(x_i\right) - g\left(x_{i-1}\right)\right)$ tomando $a_i = g\left(x_i\right) - g\left(x_{i-1}\right)$ y $b_i = F(c_i)$, obtenemos

$$S\left(P, F, g\right) = \sum_{i=1}^{n-1} \left(g(x_i) - g(a)\right)\left(F(c_i) - F(c_{i+1})\right) + F(c_n)\left(g(b) - g(a)\right) =$$

$$\sum_{i=1}^{n-1} g(x_i)\left(F(c_i) - F(c_{i+1})\right) - \left(F(c_1) - F(c_n)\right)g(a) + F(c_n)\left(g(b) - g(a)\right) =$$

$$\sum_{i=1}^{n-1} g(x_i)\left(F(c_i) - F(c_{i+1})\right) - F(c_1)g(a) + F(c_n)g(b) =$$

$$\sum_{i=1}^{n-1} g(x_i)\left(F(c_i) - F(c_{i+1})\right) + \left(F(a) - F(c_1)\right)g(a) + \left(F(c_n) - F(b)\right)g(b) + F(b)g(b) - F(a)g(a) =$$

$$S(\widetilde{P}, g, F) + g(b)F(b) - g(a)F(a)$$

siendo $\widetilde{P}$ la partición formada por los puntos $a, c_1, c_2, ..., c_n, b$ y los puntos intermedios son $a, x_1, x_2, ..., x_{n-1}, b$. Observamos además que $\left\|\widetilde{P}\right\| \leq 2\left\|P\right\|$ por lo que tomando límite cuando $\left\|P\right\| \to 0$ en la igualdad

$$S(P, F, g) = S(\widetilde{P}, g, F) + g(b)F(b) - g(a)F(a)$$

obtenemos que existe $\int_a^b Fdg$ y la fórmula de partes. ✓

Proposición 5.22. ***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 $\int_c^d gohd\left(Foh\right)$ y además

$$\int_c^d g(h(t))dF(h(t)) = \int_a^b g(x)dF(x).$$

Demostración.
Supondremos que h es creciente, el caso decreciente es análogo. Si $P = \{c, t_1, t_2, ..., t_{n-1}, d\}$ es una partición de $[c, d]$ con puntos intermedios $c_i \in [t_{i-1}, t_i]$ $i = 1, 2, ..., n$ entonces

$$S(P, goh, Foh) = \sum_{i=1}^{n} g(h(c_i))[F(h(x_i)) - F(h(x_{i-1}))] = S\left(\widetilde{P}, g, F\right)$$

siendo $\widetilde{P} = \{a, h(t_1), h(t_2), ..., h(t_{n-1}), b\}$ con puntos intermedios $h(c_i)$ (esto se puede hacer ya que h es creciente y biyectiva). Además como h es continua, si $\|P\| \to 0$ entonces $\|h(P)\| = \left\|\widetilde{P}\right\| \to 0$, lo cual se deduce ya que h es uniformemente continua (dado $\varepsilon > 0$ existe $\delta > 0$ tal que si $|x - y| < \delta$ entonces $|h(x) - h(y)| < \varepsilon$). Por lo tanto tomando límite cuando $\|P\| \to 0$ se deduce que $\int_c^d gohd(Foh)$ existe y la fórmula buscada. ✓

5.3. Extensión a funciones complejas e integrales impropias.

Definición 5.23. Integrales con integrando complejo. Dadas $g : [a, b] \to \mathbb{C}$, $g = g_1 + ig_2$ y $F : [a, b] \to \mathbb{R}$, diremos que existe $\int_a^b gdF$ si y sólo si existen $\int_a^b g_1dF$ e $\int_a^b g_2dF$ y en ese caso,

$$\int_a^b gdF = \int_a^b g_1dF + \int_a^b g_2dF.$$

Definición 5.24. Integrales impropias.
Si g, $F : \mathbb{R} \to \mathbb{R}$ son tales que $\int_a^b gdF$ existe cualesquiera sean a y b, definimos

$$\int_{-\infty}^{+\infty} gdF = \lim_{a\to-\infty\ b\to+\infty} \int_a^b gdF.$$

en caso de que exista el límite.

Definición 5.25. Dadas $g : \mathbb{R} \to \mathbb{C}$ $(g = g_1 + g_2)$ $F : \mathbb{R} \to \mathbb{R}$, diremos que existe $\int_{-\infty}^{+\infty} gdF$ si y sólo si existen $\int_{-\infty}^{+\infty} g_1dF$ y $\int_{-\infty}^{+\infty} g_2dF$ y además

$$\int_{-\infty}^{+\infty} gdF = \int_{-\infty}^{+\infty} g_1dF + i\int_{-\infty}^{+\infty} g_2dF.$$

5.4. Aplicaciones a la teoría de la probabilidad.

Proposición 5.26. *Si F_X es función de distribución de una variable aleatoria X, entonces*

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

Demostración.
Basta observar que $\int_a^b dF_X(x) = F_X(a) - F_X(b)$ de donde se deduce el resultado. ✓
Nota. Se puede probar que $\int_A dF_X(x) = P(X \in A)$ cualquiera sea A boreliano en $\mathbb{R}$ (donde nuevamente el significado de esta integral es el de Lebesgue).

Proposición 5.27. *Si X es discreta cuyo recorrido es $A = \{x_1, x_2, ...\}$ y $g : [a,b] \to \mathbb{R}$ es continua, entonces*

$$\int_a^b g(x)dF_X(x) = \sum_{x\in(a,b]\cap A} g(x)p_X(x).$$

Demostración.
$F_X(x) = \sum_{i\,:\,x_i\leq x} p_X(x_i) = \sum_i p_X(x_i)\mathbf{1}_{[x_i,+\infty)}(x)$. Definimos para cada n, $A_n = \{x_1, x_2, ..., x_n\}$ y $F_n(x) = \sum_{i=1}^n p_X(x_i)\mathbf{1}_{[x_i,+\infty)}(x)$. Dado $\varepsilon > 0$, existe n_0 tal que para cada $n \geq n_0$ se cumple que $P(X \in A_n) \geq 1 - \varepsilon/n$. Por lo tanto para cada $x \in \mathbb{R}$ se tiene que $0 \leq F_X(x) - F_n(x) \leq \varepsilon/n$ (para $n \geq n_0$) . Como g es continua, entonces $|g(x)| \leq k$ para todo $x \in [a,b]$ y por lo tanto

$$\left|\int_a^b g(x)d\left(F_X(x) - F_n(x)\right)\right| \leq \int_a^b |g(x)|\, d\left(F_X(x) - F_n(x)\right) \leq 2k\varepsilon/n \underset{n\to+\infty}{\to} 0$$

$$\int_a^b g(x)dF_n(x) = \int_a^b g(x)d\left(\sum_{i=1}^n p_X(x_i)\mathbf{1}_{[x_i,+\infty)}(x)\right) = \sum_{i=1}^n p_X(x_i)\int_a^b g(x)d\mathbf{1}_{[x_i,+\infty)}(x) =$$
$$\sum_{i\,:\,x_i\in(a,b]\cap A_n} g(x_i)p_X(x_i).$$

$$\int_a^b g(x)dF_X(x) = \int_a^b g(x)dF_n(x) + \int_a^b g(x)d\left(F_X(x) - F_n(x)\right) =$$
$$\sum_{i\,:\,x_i\in(a,b]\cap A_n} g(x_i)p_X(x_i) + \int_a^b g(x)d\left(F_X(x) - F_n(x)\right)$$

Tomando límite cuando $n \to +\infty$ se obtiene el resultado. ✓

Proposición 5.28. *Si X es absolutamente continua con densidad f_X y $g : [a,b] \to \mathbb{R}$ es continua, entonces*

$$\int_a^b g(x)dF_X(x) = \int_a^b g(x)f_X(x)dx.$$

Demostración.
Es corolario inmediato del teorema 1.11. ✓

5.5. Integrales de Riemann-Stieltjes múltiples.

Si (X,Y) es un vector aleatorio y $F_{X,Y}:\mathbb{R}^2\to\mathbb{R}$ su función de distribución. Supongamos que $g:[a,b]\times[c,d]\to\mathbb{R}$, definiremos $\iint_{[a,b]\times[c,d]} g(x,y)dF_{X,Y}(x,y)$. Si $F_{X,Y}:\mathbb{R}^2\to\mathbb{R}$ es una función de distribución conjunta y $g:[a,b]\times[c,d]\to\mathbb{R}$. Dada $P_X=\{a=x_0,x_1,x_2,...,x_{n-1},x_n=b\}$ es una partición de $[a,b]$ con puntos intermedios $c_i\in[x_{i-1},x_i]$ $i=1,2,,...,n$, $P_Y=\{c=y_0,y_1,y_2,...,y_{m-1},y_m=d\}$ es una partición de $[c,d]$ con puntos intermedios $c_i'\in[y_{i-1},y_i]$ $i=1,2,,...,m$, definimos las sumas parciales de Riemann-Stieltjes, sobre $P_X\times P_Y$ como $S(P_X\times P_Y,g,F_{X,Y})=\sum_{i=1}^n\sum_{j=1}^m g(c_i,c_j')p_{ij}$ siendo

$$p_{ij}=P\left((X,Y)\in(x_{i-1},x_i]\times(y_{j-1},y_j]\right)=$$

$$F_{X,Y}(x_i,y_j)-F_{X,Y}(x_{i-1},y_j)-F_{X,Y}(x_i,y_{j-1})+F_{X,Y}(x_{i-1},y_{j-1}).$$

Definimos la norma de la partición como $\|P\|=$máx$\{\|P_X\|,\|P_Y\|\}$. Como en el caso univariado diremos que $\lim_{\|P\|\to 0} S(P,g,F_{X,Y})=I$ si y sólo si dado $\varepsilon>0$, existe $\delta>0$ tal que para toda P partición de $[a,b]\times[c,d]$ (con sus correspondientes puntos intermedios c_i y c_i') con $\|P\|<\delta$, se cumple que $|S(P,g,F_{X,Y})-I|<\varepsilon$.

Definición 5.29. Integral doble de Riemann-Stieltjes.
Dadas $g:[a,b]\times[c,d]\to\mathbb{R}$, $F_{X,Y}:\mathbb{R}^2\to\mathbb{R}$ función de distribución de un vector aleatorio (X,Y)si existe $\lim_{\|P\|\to 0} S(P,g,F)=I$, diremos que la integral de Riemann-Stieltjes de g respecto de $F_{X,Y}$ en $[a,b]\times[c,d]$ existe y vale I.

Notación:

$$\int_a^b gdF=\iint_{[a,b]\times[c,d]} g(x,y)dF_{X,Y}(x,y)$$

Es válido el mismo teorema de las tres equivalencias para la existencia de la integral, probadas en el caso univariado, con demostraciones análogas. De manera análoga se prueban también el siguiente teorema y las propiedades que siguen.
Teorema.
Si F es distribución, y $g:[a,b]\times[c,d]\to\mathbb{R}$ es continua, entonces existe $\iint_{[a,b]\times[c,d]} gdF$.
Propiedades.
Las siguientes propiedades, pueden ser demostradas de manera similar al caso univariado.

1. Si $g,h:[a,b]\times[c,d]\to\mathbb{R}$ $F=F_{X,Y}$ son tales que existen las integrales $\iint_{[a,b]\times[c,d]} gdF$ y $\iint_{[a,b]\times[c,d]} hdF$ entonces también existe $\iint_{[a,b]\times[c,d]}(\alpha g+\beta h)\,dF$ cualesquiera sean $\alpha,\beta\in\mathbb{R}$ y además

$$\iint_{[a,b]\times[c,d]}(\alpha g+\beta h)\,dF=\alpha\iint_{[a,b]\times[c,d]} gdF+\beta\iint_{[a,b]\times[c,d]} hdF.$$

2. Si F, G son distribuciones, $h : [a,b]\times[c,d] \to \mathbb{R}$, son tales que existen las integrales $\iint_{[a,b]\times[c,d]} hdF$ y $\iint_{[a,b]\times[c,d]} hdG$ entonces también existe $\iint_{[a,b]\times[c,d]} hd\left(\alpha F + \beta G\right)$ cualesquiera sean $\alpha, \beta \in R$ y además
$$\iint_{[a,b]\times[c,d]} hd\left(\alpha F + \beta G\right) = \alpha \iint_{[a,b]\times[c,d]} hdF + \beta \iint_{[a,b]\times[c,d]} hdG.$$
3. Si F es distribución, $g : [a,b]\times[c,d] \to \mathbb{R}$ son tales que $g \geq 0$, y existe $\iint_{[a,b]\times[c,d]} gdF$, entonces $\iint_{[a,b]\times[c,d]} gdF \geq 0$.
4. Si F es distribución, $g, h : [a,b]\times[c,d] \to \mathbb{R}$ son tales que $g \geq h$, y existen $\iint_{[a,b]\times[c,d]} gdF$ y $\iint_{[a,b]\times[c,d]} hdF$ entonces $\iint_{[a,b]\times[c,d]} gdF \geq \iint_{[a,b]\times[c,d]} hdF$.

5.5.1. Aplicaciones a la teoría de la probabilidad.

1. Si $F_{X,Y}$ es la función de distribución de una vector aleatorio (X,Y), entonces
$$\iint_{[a,b]\times[c,d]} dF_{X,Y}(x,y) = P\left(a < X \leq b,\ c < Y \leq d\right)..$$
2. Si (X,Y) es discreto cuyo recorrido es $A = \left\{(x_i, y_j)\right\}_{i,j}$ y $g : [a,b]\times[c,d] \to R$ es continua, entonces
$$\iint_{[a,b]\times[c,d]} g(x,y)dF_{X,Y}(x,y) = \sum_{(x,y)\in(a,b]\times(c,d]\cap A} g(x,y)p_{X,Y}(x,y)$$
3. Si (X,Y) es absolutamente continuo con función de densidad $f_{X,Y}$ y $g : [a,b]\times[c,d] \to \mathbb{R}$ es continua, entonces
$$\iint_{[a,b]\times[c,d]} g(x,y)dF_{X,Y}(x,y) = \iint_{[a,b]\times[c,d]} g(x,y)f_{X,Y}(x,y)dxdy$$

5.5.2. Integrales múltiples impropias.

Definición 5.30. Dadas $g : \mathbb{R}^n \to \mathbb{R}$ y $F_{X_1,X_2,...,X_n}$ distribución conjunta del vector $(X_1, X_2, ..., X_n)$
$$\int\cdots\int_{\mathbb{R}^n} g(x_1,x_2,...,x_n)dF_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n) =$$
$$\lim_{\substack{a_i \to -\infty \\ b_i \to +\infty \\ \text{para todo } i}} \int\cdots\int_{[a_1,b_1]\times[a_2,b_2]\times...\times[a_n,b_n]} g(x_1,x_2,...,x_n)dF_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)$$

Capítulo 6

Valor esperado.

6.1. Definición.

Un concepto esencial en teoría de la probabilidad y estadística es el concepto de esperanza o valor esperado de una variable aleatoria, el mismo será definido de tal modo que quede un promedio ponderado de los valores que puede tomar la variable. También se verá más adelante, mediante la llamada ley de los grandes números que el valor esperado puede verse también como un valor al cual converge (en cierto sentido) el promedio de una muestra de observaciones tomadas al azar, cuando el tamaño de la muestra (cantidad de observaciones) tiende a infinito. Todo esto va dicho de manera muy informal, pero será precisado más adelante.
Supongamos que tenemos un conjunto formado por 100 personas de las cuales 90 tienen una altura de 170 cms, 5 miden 167 cms y los restantes 5 miden 172 cms. La altura promedio de este conjunto de personas, la calculamos, sumando la altura de las 100 personas, y lo dividimos entre 100 que es el total de personas, así obtenemos que la altura promedio es $\frac{90\times 170+5\times 167+5\times 172}{100} = 169.95$. Si sorteamos un individuo al azar y definimos $X =$ .altura del individuo sorteado", tendríamos que $\text{Rec}(X) = \{167, 170, 172\}$ y su fnción de probabilidad sería $p_X(167) = \frac{5}{100} = 0,05$; $p_X(167) = \frac{90}{100} = 0,9$ y $p_X(172) = \frac{5}{100} = 0,05$ por lo tanto, la altura promedio la podemos escribir como $167\times 0,05+170\times 0,9+172\times 0,05 = 167\times p_X(167)+170\times p_X(170)+172\times p_X(172)$. A este valor le llamaremos esperanza (o valor esperado de X) y lo simbolizaremos como $\mathbb{E}(X)$. Razonando como en este ejemplo, dada una variable aleatoria X discreta, su valor esperado debería ser definido como $\sum\limits_{x\in\text{Rec}(X)} xp_X(x)$, y de ahí, parece natural definirlo para el caso absolutamente continuo como $\int_{-\infty}^{+\infty} xf_X(x)dx$. Aún nos quedaría por definir el valor esperado para una variable aleatoria mixta.

Definición 6.1. Dado $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y $X : \Omega \to \mathbb{R}$ variable aleatoria tal que $\int_{-\infty}^{+\infty} |x|\, dF_X(x) < +\infty$. Definimos

$$\mathbb{E}(X) := \int_{-\infty}^{+\infty} x dF_X(x)$$

y le llamaremos esperanza de X o valor esperado de X.

Diremos también que existe $\mathbb{E}(X)$ cuando se cumple que $\int_{-\infty}^{+\infty} |x|\, dF_X(x) < +\infty$.

Definición 6.2. Dado un $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad, si $A \in \mathcal{A}$ es tal que $P(A) = 1$, diremos que el suceso A ocurre casi seguramente (c.s.).

Observación 6.3. *Si $A \subset \mathbb{R}$ es un boreliano tal que $P(X \in A) = 1$ (es decir si A ocurre c.s.) y existe $\mathbb{E}(X)$, entonces $\mathbb{E}(X) = \int_A x dF_X(x)$, ya que sobre A^c la integral vale 0.*

Observación 6.4. *Si X es discreta, observando que para cada $x \in \mathbb{R}$ se cumple que $p_X(x) = F_X(x) - F_X(x^-)$, entonces existe $\mathbb{E}(X)$ si y sólo si $\sum_{x \in Rec(X)} |x| p_X(x) < +\infty$ y además*

$$\mathbb{E}(X) = \sum_{x \in Rec(X)} x p_X(x).$$

Observación 6.5. *Si X es absolutamente continua, como $F'_X(x) = f_X(x)$ en todo punto x de continuidad de f_X, entonces entonces existe $\mathbb{E}(X)$ si y sólo si $\int_{-\infty}^{+\infty} |x| f_X(x) dx < +\infty$ y además*

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

Observación 6.6. *La convergencia absoluta de la integral que define el valor esperado, se realiza para evitar problemas de convergencia debido a la reordenación de términos en el caso de la serie, o reordenación en los intervalos en el caso absolutamente continuo.*

Cuando $X \geq 0$ casi seguramente, resulta $F_X(x) = 0$ para todo $x < 0$, por lo tanto $\int_{-\infty}^{+\infty} x dF_X(x) = \int_0^{+\infty} x dF_X(x) \geq 0$ lo cual motiva la siguiente definición.

Definición 6.7. Si $X \geq 0$ casi seguramente, y $\int_{-\infty}^{+\infty} |x| dF_X(x) = +\infty$, diremos que $\mathbb{E}(X) = +\infty$.

6.2. Ejemplos.

Ejemplo 6.8. Si $X \sim$Ber(p) entonces $\mathbb{E}(X) = p$ ya que $\mathbb{E}(X) = 0.P(X = 0) + 1.P(X = 1) = p$.

Ejemplo 6.9. Si $X \sim$Bin(n, p) entonces $\mathbb{E}(X) = np$. $\mathbb{E}(X) = \sum_{x=0}^{n} x P(X = x) = \sum_{x=0}^{n} x C_x^n p^x (1-p)^{n-x} = np$. Se deja como ejercicio, verificar la anterior igualdad.

Ejemplo 6.10. Si $X \sim N(\mu, \sigma^2)$ entonces $\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(x-\mu)^2} dx = \mu$. Se deja como ejercicio, verificar la anterior igualdad.

Ejemplo 6.11. Como habíamos observado anteriormente, $F_Y(y) = \begin{cases} 0 & \text{si} \quad y < 1/2 \\ y & \text{si} \quad 1/2 \leq y \leq 1 \\ 1 & \text{si} \quad y > 1 \end{cases}$, F_Y tiene un único salto en $1/2$, y además es derivable en $[1/2, 1]$ con $F'_Y(y) = 1$, por lo tanto, obtenemos

$$\mathbb{E}(Y) = \int_{-\infty}^{+\infty} y dF_Y(y) = \frac{1}{2}\left(F_Y(1/2) - F_Y(1/2^-)\right) + \int_{1/2}^{1} y dy = \frac{5}{8}.$$

6.3. Propiedades.

En las siguientes propiedades se considera dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.

Teorema 6.12. *Si $X : \Omega \to \mathbb{R}$ es variable aleatoria tal que $X \geq 0$ c.s. (es decir que $P(X \geq 0) = 1$) y existe $\mathbb{E}(X)$, entonces $\mathbb{E}(X) \geq 0$.*

Demostración.
Como $X \geq 0$, entonces se tiene que $F_X(x) = 0$ para todo $x < 0$. Entonces, se cumple que

$$0 = \mathbb{E}(X) = \int_{-\infty}^{+\infty} x dF_X(x) = \int_{0}^{+\infty} x dF_X(x) \geq 0.\checkmark$$

Teorema 6.13. *Si $X : \Omega \to \mathbb{R}$ es tal que $X = a$ c.s. (es decir que $P(X = a) = 1$) (X es constante), entonces existe $\mathbb{E}(X)$ y además $\mathbb{E}(X) = a$. Es decir, $\mathbb{E}(a) = a$.*

Demostración.
Observando que $X = a$ es una variable discreta donde $P(X = a) = 1$, entonces

$$\mathbb{E}(a) = aP(X = a) = a.$$

Teorema 6.14. *Si $X : \Omega \to \mathbb{R}$ es variable aleatoria tal que $X \geq 0$ c.s. y $\mathbb{E}(X) = 0$, entonces $X = 0$. c.s.*

Demostración.
Como $X \geq 0$, se deduce se tiene que $F_X(x) = 0$ para todo $x < 0$. Entonces, cualesquiera sean $0 < \alpha < \beta$, se cumple que

$$0 = \mathbb{E}(X) = \int_{-\infty}^{+\infty} x dF_X(x) = \int_{0}^{+\infty} x dF_X(x) \geq \int_{\alpha}^{\beta} x dF_X(x) \geq$$
$$\alpha\left(F_X(\beta) - F_X(\alpha)\right).$$

Entonces $\alpha\left(F_X(\beta) - F_X(\alpha)\right) = 0$, por lo que se deduce que $F_X(\beta) = F_X(\alpha)$ para todos $\alpha, \beta > 0$. Entonces, $F_X(x)$ es constante para $x > 0$, lo cual sumado al hecho de que debe tener límite 1 cuando x tiende a $+\infty$, entonces se obtuvo que $F_X(x) = 1$ para todo $x > 0$, lo cual sumado al hecho de que $F_X(x) = 0$ para todo $x < 0$, y como F_X es continua por derecha en 0, entonces $F_X(0) = 1$, y entonces se obtiene que $P(X = 0) = 1$. ✓

Corolario 6.15. *Si X, Y son variables aleatorias tales que $X \geq Y$ c.s., existen $\mathbb{E}(X)$ y $\mathbb{E}(Y)$, y además $\mathbb{E}(X) = \mathbb{E}(Y)$ entonces $X = Y$ c.s.*

Demostración.
Basta observar que $X - Y \geq 0$ c.s. y que $\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = 0$, luego por el teorema anterior se tiene que $X - Y = 0$ c.s. ✓

Teorema 6.16. *Si $X : \Omega \to \mathbb{R}$ es variable aleatoria, $g : \mathbb{R} \to \mathbb{R}$ es una función boreliana ($g^{-1}(A) \in \mathcal{B}$ para todo $A \in \mathcal{B}$) tal que existe $\mathbb{E}(g(X))$, entonces*

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) dF_X(x).$$

Demostración.
Haremos la demostración suponiendo que g es monótona y biyectiva. El caso general se prueba a partir de teoría de la medida.
Supongamos que g es creciente y biyectiva, el caso decreciente es análogo.

$$F_{g(X)}(y) = P(g(X) \leq y) = P\left(X \leq g^{-1}(y)\right) = F_X(g^{-1}(y)).$$

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} y dF_{g(X)}(y) = \int_{-\infty}^{+\infty} y dF_X(g^{-1}(y))$$

si ahora hacemos el cambio de variable $y = g(x)$, entonces la última integral nos queda

$$\int_{-\infty}^{+\infty} g(x) dF_X(x). \checkmark$$

Observación 6.17. *A partir de esta propiedad, se deduce que existe $\mathbb{E}(X)$ si y sólo si $\mathbb{E}(|X|) < +\infty$.*

Ejemplo 6.18. Si $Y = \max\{X, 1/2\}$ donde $X \sim U(0,1)$, entonces

$$\mathbb{E}(Y) = \int_{-\infty}^{+\infty} \max\{x, 1/2\} f_X(x) dx = \int_0^1 \max\{x, 1/2\} dx = \int_0^{1/2} 1/2 dx + \int_{1/2}^1 x dx = 5/8.$$

Corolario 6.19. *Si $X : \Omega \to \mathbb{R}$ es variable aleatoria tal que existe $\mathbb{E}(X)$, entonces cualquiera sea $\alpha \in \mathbb{R}$,*

$$\textit{existe } \mathbb{E}(\alpha X) \textit{ y además } \mathbb{E}(\alpha X) = \alpha \mathbb{E}(X).$$

Demostración.
La existencia de $\mathbb{E}(\alpha X)$ se deduce de la linealidad de la integral de Riemann Stieltjes ya que $\int_{-\infty}^{+\infty} |\alpha x| dF_X(x) = |\alpha| \int_{-\infty}^{+\infty} |x| dF_X(x)$.
Ahora consideramos $g : \mathbb{R} \to \mathbb{R}$ tal que $g(x) = \alpha x$, entonces g es boreliana y por lo tanto

$$\mathbb{E}(\alpha X) = \int_{-\infty}^{+\infty} \alpha x dF_X(x) = \alpha \int_{-\infty}^{+\infty} x dF_X(x) = \alpha \mathbb{E}(X). \checkmark$$

Teorema 6.20. *Si* $\mathbb{E}(|X|) < +\infty$, *entonces*

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|).$$

Demostración.

$$|\mathbb{E}(X)| = \left| \int_{-\infty}^{+\infty} x dF_X(x) \right| \leq \int_{-\infty}^{+\infty} |x| dF_X(x) = \mathbb{E}(|X|) \checkmark$$

Teorema 6.21. *Si* $X, Y : \Omega \to \mathbb{R}$ *son variables aleatorias y* $g : \mathbb{R}^2 \to \mathbb{R}$ *es boreliana tal que existe* $\mathbb{E}[g(X,Y)]$ *entonces*

$$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y)\, dF_{X,Y}(x,y).$$

Demostración.
Se prueba utilizando teoría de la medida.

Teorema 6.22. *Si* $X, Y : \Omega \to \mathbb{R}$ *son variables aleatorias tales que existen* $\mathbb{E}(X)$ *y* $\mathbb{E}(Y)$, *entonces existe* $\mathbb{E}(X+Y)$ *y además*

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Demostración.

$$\mathbb{E}(|X+Y|) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x+y|\, dF_{X,Y}(x,y) \leq$$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x|\, dF_{X,Y}(x,y) + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |y|\, dF_{X,Y}(x,y) =$$

$$\mathbb{E}(|X|) + \mathbb{E}(|Y|) < +\infty$$

lo cual prueba que existe $\mathbb{E}(X+Y)$.
Definiendo ahora las funciones $g, g_1, g_2 : \mathbb{R}^2 \to \mathbb{R}$ tales que $g(x,y) = x + y$, $g_1(x,y) = x$, $g_2(x,y) = y$, entonces $g = g_1 + g_2$, y por lo tanto usando la linealidad de la integral de Riemann Stieltjes, obtenemos

$$\mathbb{E}(X+Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x+y)\, dF_{X,Y}(x,y) =$$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x dF_{X,Y}(x,y) + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y dF_{X,Y}(x,y) =$$

$$\mathbb{E}(X) + \mathbb{E}(Y).\checkmark$$

Ejemplo 6.23. Si $X \sim$Bin(n,p) entonces $\mathbb{E}(X) = np$. Esto se debe a que definimos para cada $i = 1, 2, 3, ..., n$ las variables

$X_i = \begin{cases} 1 & \text{si hay éxito la vez } i\text{-ésima} \\ 0 & \text{si no} \end{cases}$ entonces cada X_i distribuye como una Ber(p) y además se cumple que $X = X_1 + X_2 + ... + X_n$, luego, aplicando la aditividad de la esperanza nos queda que

$$\mathbb{E}(X) = \mathbb{E}(X_1 + X_2 + ... + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + ... + \mathbb{E}(X_n) = np.$$

Teorema 6.24. *Si $X, Y : \Omega \to \mathbb{R}$ son variables aleatorias tales que $X \leq Y$ c.s. y existen $\mathbb{E}(X)$ y $\mathbb{E}(Y)$, entonces $\mathbb{E}(X) \leq \mathbb{E}(Y)$.*

Demostración.
Como $Y - X \geq 0$, entonces

$$0 \leq \mathbb{E}(Y - X) = \mathbb{E}(Y) - \mathbb{E}(X) . \checkmark$$

Teorema 6.25. *Si $X, Y : \Omega \to \mathbb{R}$ son variables aleatorias independientes, tales que existe $\mathbb{E}(X)$ y $\mathbb{E}(Y)$, entonces existe $\mathbb{E}(XY)$ y además $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*

Demostración.
Debido a la independencia de las variables, $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ para todos x, y. Entonces

$$\mathbb{E}(|XY|) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |xy|\, dF_{X,Y}(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |x|\,|y|\, dF_X(x)dF_Y(y) =$$

$$\int_{-\infty}^{+\infty} |x| dF_X(x) \int_{-\infty}^{+\infty} |y| dF_Y(y) = \mathbb{E}(|X|)\,\mathbb{E}(|Y|) < +\infty.$$

$$\mathbb{E}(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy dF_{X,Y}(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy dF_X(x)dF_Y(y) =$$

$$\int_{-\infty}^{+\infty} x dF_X(x) \int_{-\infty}^{+\infty} y dF_Y(y) = \mathbb{E}(X)\,\mathbb{E}(Y) . \checkmark$$

Observación 6.26. *El corolario 6.18 junto al teorema 6.21, nos indican que si definimos el conjunto*

$$V = \{X : \Omega \to \mathbb{R} \text{ variable aleatoria, tal que existe } \mathbb{E}(X)\}$$

entonces V tiene estructura de espacio vectorial, ya que es un subespacio del conjunto de variables aleatorias definidas en Ω. Además, si definimos $T : V \to \mathbb{R}$ tal que $T(X) = \mathbb{E}(X)$, entonces T es una transformación lineal.

Teorema 6.27. ***Desigualdad de Jensen.***
Dados un $(\Omega, \mathcal{A}, P)$ espacio de probabilidad, X variable aleatoria y $\varphi : \mathbb{R} \to \mathbb{R}$ una función convexa tales que existen el valor esperado de X y de $\varphi(X)$. Entonces

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}[\varphi(X)] .$$

Además, si φ es estrictamente convexa y X no es constante, entonces la desigualdad es estricta.

Demostración.
Dado que φ es convexa, se cumple que existe una recta que pasa por el punto $(\mathbb{E}(X), \varphi(\mathbb{E}(X)))$ tal que el gráfico de φ está por encima de la misma. Entonces, se tiene que $\varphi(X) \geq \varphi(\mathbb{E}(X)) + a(X - \mathbb{E}(X))$ y por lo tanto, tomando esperanzas de ambos lados de la desigualdad obtenemos que $\varphi(\mathbb{E}(X)) \leq \mathbb{E}[\varphi(X)]$.
Por otro lado, definiendo $g(t) = \varphi(\mathbb{E}(X)) + a(t - \mathbb{E}(X))$, al ser φ estrictamente convexa, se cumple que $\varphi(t) \geq g(t)$ para todo t, y además, si $\varphi(t) = g(t)$ entonces $t = \mathbb{E}(X)$. Si se diera $\varphi(\mathbb{E}(X)) = \mathbb{E}[\varphi(X)]$ entonces se tendría que $\mathbb{E}(\varphi(X)) = \mathbb{E}(g(X))$, siendo $\varphi(X) \geq g(X)$ por lo que se deduce que $\varphi(X) = g(X)$ con probabilidad 1, de donde se deduce que debe ser $X = \mathbb{E}(X)$, o sea que X sería constante, lo cual concluye la prueba. ✓

Ejemplo 6.28. Dado que $\varphi(x) = e^x$ es convexa, se tiene que si existen $\mathbb{E}(X)$ y $\mathbb{E}\left(e^X\right)$ entonces $e^{\mathbb{E}(X)} \leq \mathbb{E}\left(e^X\right)$. Adem´as, si X no es constante, la desigualdad es estricta.

6.4. Teoremas de convergencia.

Supongamos que tenemos una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ y una variable aleatoria X definidas en cierto espacio de probabilidad, tales que $\lim_{n\to+\infty} X_n(w) = X(w)$ para todo $w \in \Omega$. Dado que tenemos convergencia de las X_n a la X en todo punto, es natural preguntarse si será cierto que $\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$. Veremos en el siguiente ejemplo que con la sola convergencia en todo punto w de $X_n(w)$ a $X(w)$, no es suficiente para asegurar que $\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

Ejemplo 6.29. Supongamos que $X \sim U(0,1)$, definimos la sucesión $X_n = n\mathbf{1}_{(0,1/n)}(X)$. Vemos que $\lim_{n\to+\infty} X_n(w) = 0$ para todo $w \in \Omega$, sin embargo, $\mathbb{E}(X_n) = nP(0 < X < 1/n) = 1$ para todo n y por lo tanto, en este caso $X = 0$ y no se cumple que $\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

En lo que sigue veremos dos teoremas de vital importancia en teoría de probabilidad y medida, que bajo cierto conjunto de hipótesis nos permiten aseguran la convergencia de las esperanzas de las X_n a la esperanza de X.

6.4.1. Teorema de convergencia monótona.

Teorema 6.30. ***Teorema de convergencia monótona.***
Dados $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, una sucesión de variables aleatorias $\{X_n\}_{n\in N}$ y una variable aleatoria X tales que existe $\mathbb{E}(X)$, $X_n(w) \geq 0$, $X_n(w) \uparrow X(w)$ para todo $w \in \Omega$, entonces existe $\mathbb{E}(X_n)$ para todo n y además

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

Demostración.
En primer lugar observamos que como $0 < X_n \leq X$, entonces existe $\mathbb{E}(X_n)$ para todo n. Además, dado que $X_n \leq X_{n+1}$ para todo n entonces, $\mathbb{E}(X_n) \leq \mathbb{E}(X_{n+1})$ por lo que la sucesión $\{\mathbb{E}(X_n)\}_{n\in N}$ es creciente y por lo tanto tiene límite. Por otro lado, como $X_n \leq X$ para todo n, entonces $\mathbb{E}(X_n) \leq \mathbb{E}(X)$ para todo n, por lo que $\lim\limits_{n\to+\infty} \mathbb{E}(X_n) \leq \mathbb{E}(X)$.
Entonces será suficiente probar que $\lim\limits_{n\to+\infty} \mathbb{E}(X_n) \geq \mathbb{E}(X)$. Para lograrlo, veremos que dado $\varepsilon > 0$, se cumplirá que $\lim\limits_{n\to+\infty} \mathbb{E}(X_n) \geq \mathbb{E}(X) - \varepsilon$. Fijado $\varepsilon > 0$, aproximaremos X por una variable discreta Y tal que $|X - Y| \leq \varepsilon$.
Definimos los sucesos $B_n = \{n\varepsilon < X \leq (n+1)\varepsilon\}$ para $n = 0, 1, 2, ...$ y definimos la variable $Y(w) = \begin{cases} n\varepsilon & \text{si } n\varepsilon < X(w) \leq (n+1)\varepsilon \\ 0 & \text{si } X(w) = 0 \end{cases}$. Vemos que $X - \varepsilon \leq Y \leq X$ por lo que $\mathbb{E}(X) - \varepsilon \leq \mathbb{E}(Y) \leq \mathbb{E}(X)$. Para obtener el resultado, probaremos que $\lim\limits_{n\to+\infty} \mathbb{E}(X_n) \geq \mathbb{E}(Y)$.
Definimos los sucesos $A_k = \{X_k \geq Y\}$. Si $w \in A_k$ entonces $X_k(w) \geq Y(w)$ pero $X_{k+1}(w) \geq X_k(w)$ por lo que $X_{k+1}(w) \geq Y(w)$, luego $w \in A_{k+1}$ por lo que los A_k son una sucesión creciente de sucesos. Además, para todo $w \in \Omega$, se cumple que $w \in B_n$ para algún n, y como $X_k(w) \to X(w)$ entonces existe un k_0 tal que $X_{k_0}(w) \geq n\varepsilon = Y(w)$, entonces $\cup_{k=1}^{+\infty} A_k = \Omega$.
Por lo tanto, dejando n fijo, los sucesos $A_k \cap B_n$ variando k, crecen a B_n. Por otro lado, observamos que las variables $Y\mathbf{1}_{A_k}$ son discretas, tomando los valores $0, \varepsilon, 2\varepsilon, 3\varepsilon, ...$ por lo que para cualquier m se tiene que

$$\mathbb{E}(Y\mathbf{1}_{A_k}) = \sum_{n=0}^{+\infty} n\varepsilon P(Y\mathbf{1}_{A_k} = n\varepsilon) = \sum_{n=0}^{+\infty} n\varepsilon P(A_k \cap B_n) \geq \sum_{n=0}^{m} n\varepsilon P(A_k \cap B_n).$$

$\lim\limits_{k\to+\infty} \mathbb{E}(Y\mathbf{1}_{A_k}) \geq \lim\limits_{k\to+\infty} \sum_{n=0}^{m} n\varepsilon P(A_k \cap B_n) = \sum_{n=0}^{m} n\varepsilon P(B_n)$ para todo m, entonces $\lim\limits_{k\to+\infty} \mathbb{E}(Y\mathbf{1}_{A_k}) \geq \sum_{n=0}^{+\infty} n\varepsilon P(B_n) = \mathbb{E}(Y)$. Además $Y\mathbf{1}_{A_k} \leq X_k$ entonces $\mathbb{E}(Y\mathbf{1}_{A_k}) \leq \mathbb{E}(X_k)$ por lo que $\lim\limits_{k\to+\infty} \mathbb{E}(X_k) \geq \mathbb{E}(Y)$ lo cual concluye la demostración.
✓

Observación 6.31. *El teorema sigue siendo válido si las hipótesis $X_n > 0$ y $X_n \leq X_{n+1}$ para todo n, se cumplen casi seguramente.*

Observación 6.32. *El teorema sigue valiendo en el caso en que $\mathbb{E}(X) = +\infty$, queda como ejercicio realizar la verificación de la demostración para este caso.*

6.4.2. Teorema de convergencia dominada.

Teorema 6.33. ***Teorema de convergencia dominada.***
Dados $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad, una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ y dos variables aleatorias X e Y tales que que $\lim\limits_{n\to+\infty} X_n(w) = X(w)$ y $|X_n(w)| \leq Y(w)$ para todos n y w. Además existe $\mathbb{E}(Y)$.

Entonces existen las esperanzas de X_n para todo n y la de X y además

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

Demostración.
En primer lugar vemos que como $|X_n| \leq Y$ para todo n, entonces existe la esperanza de las X_n, además tomando límites en la desigualdad, obtenemos que $|X| \leq Y$, por lo que también existe la esperanza de X.
Definimos la sucesión $Y_n = \inf_{k\geq n} X_k$ entonces $Y_n \uparrow X$ (ya que las Y_n tienden a $\sup_n Y_n = \sup_n \inf_{k\geq n} X_k$ que es el límite inferior de la sucesión X_n). Además observamos que $0 \leq Y_n + Y \uparrow X + Y$, por lo que aplicando el teorema de convergencia monótona, obtenemos que

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

Luego, $\lim_{n\to+\infty} \mathbb{E}(Y_n) = \mathbb{E}(X)$. Análogamente, definiendo $Z_n = \sup_{k\geq n} X_k$, vemos que $Z_n \downarrow X$ y como además $0 \leq Y - Z_n \uparrow Y - X$, aplicando nuevamente el teorema de convergencia monótona y utilizando la linealidad del valor esperado, obtenemos que $\lim_{n\to+\infty} \mathbb{E}(Z_n) = \mathbb{E}(X)$.
Para concluir la demostración, basta observar ahora que para todo n y todo w, se cumple que $Y_n(w) \leq X_n(w) \leq Z_n(w)$ por lo que $\mathbb{E}(Y_n) \leq \mathbb{E}(X_n) \leq \mathbb{E}(Z_n)$ y como $\lim_{n\to+\infty} \mathbb{E}(Y_n) = \mathbb{E}(X)$ y $\lim_{n\to+\infty} \mathbb{E}(Z_n) = \mathbb{E}(X)$ se obtiene que $\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$. ✓

Observación 6.34. *Como en el teorema de convergencia monótona, se tiene que basta tomar como hipótesis que* $\lim_{n\to+\infty} X_n = X$ *y* $|X_n| \leq Y$ *se cumplan casi seguramente.*

Corolario 6.35. *Si* $\lim_{n\to+\infty} X_n(w) = X(w)$ *y* $|X_n(w)| \leq k(cte)$ *para todos* n *y* w, *entonces vale el teorema ya que* k *tiene esperanza finita.*

6.4.3. Aplicaciones.

Teorema 6.36. *Si* $f_n : [a,b] \to \mathbb{R}$ *son funciones integrables Riemann en* $[a,b]$ *para todo* $n \in \mathbb{N}$, $g : [a,b] \to \mathbb{R}$ *es integrable Riemann y* $f : [a,b] \to \mathbb{R}$ $\lim_{n\to+\infty} f_n(x) = f(x)$ *y* $|f_n(x)| \leq g(x)$ *para todo* $x \in [a,b]$. *Entonces*

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

Demostración.
Consideramos $X \sim U(0,1)$. Definimos entonces las variables $Y_n = f_n(X)$ e $Y = f(X)$. Entonces $Y_n \xrightarrow{c.s.} Y$, $|Y_n| \leq g(X)$, existe $\mathbb{E}(g(X)) = \frac{1}{b-a}\int_a^b g(x)dx < +\infty$, luego por el teorema de convergencia dominada, se tiene que $\lim_{n\to+\infty} \mathbb{E}(Y_n) = \mathbb{E}(Y)$, ahora vemos que $\mathbb{E}(Y_n) = \mathbb{E}(f_n(X)) = \frac{1}{b-a}\int_a^b f_n(x)dx$ y $\mathbb{E}(Y) = \mathbb{E}(f(X)) = \frac{1}{b-a}\int_a^b f(x)dx$ entonces $\lim_{n\to+\infty} \frac{1}{b-a}\int_a^b f_n(x)dx = \frac{1}{b-a}\int_a^b f(x)dx$, de donde se deduce el resultado.✓

Teorema 6.37. *Dada la sucesión doblemente indizada (sucesión de sucesiones)* $\left\{a_n^{(k)}\right\}_{n,k\in\mathbb{N}}$ $\subset \mathbb{R}$. *Supongamos que existe una sucesión* $\left\{b^{(k)}\right\}_{k\in\mathbb{N}}$ *tal que* $b^{(k)} > 0$, *para todo* k, $\sum_{k=1}^{+\infty} b^{(k)} = L < +\infty$ $\left|a_n^{(k)}\right| \leq b^{(k)}$ *para todos* n, k. *Si* $\underset{n\to+\infty}{lim} a_n^{(k)} = a^{(k)}$, *entonces*

$$\underset{n\to+\infty}{lim} \sum_{k=1}^{+\infty} a_n^{(k)} = \sum_{k=1}^{+\infty} a^{(k)}.$$

Demostración.
Definimos el espacio de probabilidad $\left(\mathbb{N}, 2^{\mathbb{N}}, P\right)$ donde $P(\{k\}) = \frac{b^{(k)}}{L}$.
Definimos la sucesión de variables aleatorias $X_n : \mathbb{N} \to \mathbb{R}$ tales que $X_n(k) = \frac{a_n^{(k)}}{b^{(k)}}$ y $X : \mathbb{N} \to \mathbb{R}$ tal que $X(k) = \frac{a^{(k)}}{b^{(k)}}$. Entonces $X_n \overset{c.s.}{\to} X$ (ya que $X_n(k) \to X(k)$ para todo $k \in \mathbb{N}$). Además

$$P\left(X_n = \frac{a_n^{(k)}}{b^{(k)}}\right) = P(\{k\}) = \frac{b^{(k)}}{L}.$$

Análogamente,

$$P\left(X = \frac{a^{(k)}}{b^{(k)}}\right) = P(\{k\}) = \frac{b^{(k)}}{L}.$$

Además $|X_n(k)| \leq 1$ para todo k. Entonces, aplicando el teorema de convergencia dominada, se deduce que $\lim\limits_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

$$\mathbb{E}(X_n) = \sum_{k=1}^{+\infty} \frac{a_n^{(k)}}{b^{(k)}} P\left(X_n = \frac{a_n^{(k)}}{b^{(k)}}\right) = \sum_{k=1}^{+\infty} \frac{a_n^{(k)}}{b^{(k)}} \frac{b^{(k)}}{L} = \frac{1}{L} \sum_{k=1}^{+\infty} a_n^{(k)}$$

y análogamente,

$$\mathbb{E}(X) = \sum_{k=1}^{+\infty} \frac{a^{(k)}}{b^{(k)}} P\left(X = \frac{a^{(k)}}{b^{(k)}}\right) = \sum_{k=1}^{+\infty} \frac{a^{(k)}}{b^{(k)}} \frac{b^{(k)}}{L} = \frac{1}{L} \sum_{k=1}^{+\infty} a^{(k)}.$$

Entonces obtuvimos que $\lim\limits_{n\to+\infty} \frac{1}{L} \sum_{k=1}^{+\infty} a_n^{(k)} = \frac{1}{L} \sum_{k=1}^{+\infty} a^{(k)}$ de donde se deduce el resultado. ✓
Como aplicación, se deja como ejercicio hallar $\lim\limits_{n\to+\infty} \sum\limits_{k=1}^{+\infty} \frac{1}{n^2k^2}$; $\lim\limits_{n\to+\infty} \int_0^n \left(1 + \frac{x}{n}\right)^n e^{-2x} dx$ y $\lim\limits_{n\to+\infty} \int_0^n \left(1 - \frac{x}{n}\right)^n e^{x/2} dx$.

Capítulo 7

Espacios L^p.

7.1. Definición y propiedades.

Definición 7.1. Espacios L^p.
Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, y $p > 0$, se define el conjunto

$$L^p = \{X : \Omega \to \mathbb{R} \text{ variable aleatoria tal que } \mathbb{E}(|X|^p) < +\infty)\}.$$

Teorema 7.2. *Si* $0 < p < q$ *entonces* $L^q \subset L^p$.

Demostración.
Si $X \in L^q$, entonces

$$\mathbb{E}(|X|^p) = \mathbb{E}\left(|X|^p 1_{\{|X|<1\}}\right) + \mathbb{E}\left(|X|^p 1_{\{|X|\geq 1\}}\right) \leq 1 + \mathbb{E}\left(|X|^q 1_{\{|X|\geq 1\}}\right) \leq$$

$$1 + \mathbb{E}(|X|^q) < +\infty. \checkmark$$

Diremos que X admite momentos de orden p si y sólo si $\mathbb{E}(|X|^p) < +\infty$ o sea, si y sólo si $X \in L^p$.
Del teorema anterior deducimos que si X admite momentos de orden p, entonces admite momentos de cualquier orden menor que p. Por ejemplo, decir que X admite momentos de orden 3, implica que admite momentos de cualquier orden menor que 3.

Teorema 7.3. *Si* $X, Y \in L^p$ *entonces* $\alpha X + \beta Y \in L^p$ *para todos* $\alpha, \beta \in \mathbb{R}$. *Es decir que* L^p *es un espacio vectorial (ya que es subespacio del conjunto de todas las variables aleatorias, que forman un espacio vectorial).*

Demostración.
Si $X \in L^p$ entonces cualquiera sea $\alpha \in \mathbb{R}$, se tiene que $\mathbb{E}(|\alpha X|^p) = |\alpha|^p \mathbb{E}(|X|^p) < +\infty$ por lo que $\alpha X \in L^p$.
Ahora, si $X, Y \in L^p$ observamos que $|X + Y| \leq |X| + |Y| \leq 2\max\{|X|, |Y|\}$ entonces $|X + Y|^p \leq 2^p \max\{|X|^p, |Y|^p\}$, por lo tanto se tiene que
$\mathbb{E}(|X + Y|^p) \leq 2^p \max\{\mathbb{E}|X|^p, \mathbb{E}(|Y|^p) < +\infty\}. \checkmark$

Observación 7.4. *Si* $X,Y\in L^2$, *entonces* $XY\in L^1$, *ya que* $XY=\frac{1}{2}\left[(X+Y)^2-X^2-Y^2\right]$, *es combinación lineal de variables que* $\in L^2$.

Teorema 7.5. ***Desigualdad de Cauchy-Schwartz.***
Si $X,Y\in L^2$,

$$[\mathbb{E}(XY)]^2\leq\mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right).$$

Además se da el igual si y sólo si existe $\lambda_0\in\mathbb{R}$ *tal que*

$$P(X=\lambda_0 Y)=1\,(o\quad P(Y=\lambda_0 X)=1).$$

Demostración.

$$0\leq\mathbb{E}(X-\lambda Y)^2=\lambda^2\mathbb{E}\left(Y^2\right)-2\lambda\mathbb{E}(XY)+\mathbb{E}\left(X^2\right)\text{ para todo }\lambda\in\mathbb{R}.$$

Entonces, si Y no es la función nula casi seguramente, podemos asegurar que nos quedó un polinomio de segundo grado. Como dicho polinomio es ≥ 0 para todo valor de λ, no puede tener dos raíces reales y distintas, por lo que su discriminante debe ser ≤ 0. Entonces $4\left[\mathbb{E}(XY)\right]^2-4\mathbb{E}(X^2)\,\mathbb{E}(Y^2)\leq 0$, de donde se deduce la desigualdad. Además, si fuera $[\mathbb{E}(XY)]^2=\mathbb{E}(X^2)\,\mathbb{E}(Y^2)$, entonces existe un valor de λ donde se anula el polinomio, dicho valor es $\lambda_0=\frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}$, y por lo tanto para dicho valor λ_0, se tiene que $0=\mathbb{E}(X-\lambda_0 Y)^2$, por lo que se tiene que $X=\lambda_0 Y$ casi seguramente.
Si fuera $Y=0$ casi seguramente, entonces también se cumple la igualdad, y además $Y=0X$, lo cual concluye la prueba. ✓
La desigualdad de Cauchy Schwartz recién probada, responde a la conocida desigualdad respecto a espacios vectoriales con producto interno. Para ello definimos la función $\langle\ ,\ \rangle : L^2\times L^2\to\mathbb{R}$ tal que $\langle X,Y\rangle=\mathbb{E}(XY)$, entonces, $\langle\ ,\ \rangle$ es un seudo producto interno, es decir que es una función bilineal simétrica, tal que $\langle X,X\rangle=\mathbb{E}(X^2)\geq 0$ pero no cumple la condición $\langle X,X\rangle=0$ si y sólo si $X=0$, ya que en este caso si $\langle X,X\rangle=\mathbb{E}(X^2)=0$, entonces $X=0$ c.s. por lo que puede haber infinitas (dependiendo del espacio de probabilidad) funciones que cumplan $\langle X,X\rangle=0$. Este problema se puede solucionar si identificamos todas las variables aleatorias que son 0 casi seguramente. Para ello se define la relación $\sim$ tal que $X\sim Y$ si y sólo si $X=Y$ c.s.
Se deja como ejercicio chequear que $\sim$ es una relación de equivalencia, y que si definimos $\langle\ ,\ \rangle : L^2/\sim\times L^2/\sim\to\mathbb{R}$ tal que $\langle[X],[Y]\rangle=\mathbb{E}(XY)$ donde X e Y son representantes de $[X]$ y $[Y]$ respectivamente, entonces la función $\langle\ ,\ \rangle$ está bien definida y define un producto interno en $L^2/\sim$.

7.2. Varianza de una variable aleatoria.

Junto con el valor esperado de una variable aleatoria, en la mayoría de las aplicaciones es necesario tener algún tipo de medida sobre la dispersión que hay entre los valores que puede tomar la variable, y su valor esperado. La definición de varianza apunta en esa dirección.

Definición 7.6. Varianza de una variable aleatoria.
Si $X \in L^2$, entonces se define la varianza de X, como el valor

$$\mathbb{V}(X) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right].$$

Observación 7.7. *Como se ve, si le llamamos $\mu = \mathbb{E}(X)$, entonces la varianza es el valor esperado de la variable $(X-\mu)^2$ que mide la diferencia entre los valores que puede tomar X y su valor esperado, elevado al cuadrado.*
La presencia del cuadrado es para que las diferencias entre X y su valor esperado sean positivas, ya que sin el cuadrado, la esperanza de $(X - \mathbb{E}(X))$ es 0. Por ejemplo, si X es una variable aleatoria discreta tal que $Rec(X) = \{x_1, x_2, ..., x_n\}$ con probabilidades $p_X(x_i) = 1/n$ para todo $i = 1, 2, 3, ..., n$, entonces $\mathbb{E}(X) = \sum_{x \in Rec(X)} x p_X(x) = \frac{x_1+x_2+...+x_n}{n} = \mu$. Luego, $\mathbb{V}(X) = \mathbb{E}\left[(X-\mu)^2\right] =$

$$\sum_{x \in Rec(X)} (x-\mu)^2 p_X(x) = \frac{(x_1-\mu)^2 + (x_2-\mu)^2 + ... + (x_n-\mu)^2}{n}$$

que representa el promedio de las diferencias al cuadrado que existen entre los valores que toma la variable X y su valor esperado.

En las aplicaciones, al calcular la esperanza de $(X - \mathbb{E}(X))^2$, se pierde la unidad de medida de la variable X, la cual queda expresada en unidades al cuadrado. Para salvar este problema se suele considerar la raíz cuadrada de la varianza a la que se le llama desviación típica o estandar de la variable.

Definición 7.8. Desviación típica. Si $X \in L^2$ entonces la desviación típica de X se define como

$$\sigma_X = \sqrt{\mathbb{V}(X)}.$$

Propiedades.

Teorema 7.9. *Si $X \in L^2$, entonces $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$. Aquí se sobreentiende que $\mathbb{E}^2(X) = [\mathbb{E}(X)]^2$.*

Demostración.
Llamémosle $\mu = \mathbb{E}(X)$. Entonces

$$\mathbb{V}(X) = \mathbb{E}\left[(X-\mu)^2\right] = \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right) =$$

$$\mathbb{E}\left(X^2\right) - 2\mu\mathbb{E}(X) + \mu^2 = \mathbb{E}\left(X^2\right) - 2\mu^2 + \mu^2 = \mathbb{E}\left(X^2\right) - \mu^2. \checkmark$$

Teorema 7.10. *Si $X \in L^2$, entonces $\mathbb{V}(aX+b) = a^2\mathbb{V}(X)$.*

Demostración.

$$\mathbb{V}(aX+b) = \mathbb{E}\left[(aX+b)^2\right] - [\mathbb{E}(aX+b)]^2$$

desarrollando ambos cuadrados y simplificando nos queda igual a

$$a^2\left(\mathbb{E}\left(X^2\right) - \mathbb{E}^2(X)\right) = a^2\mathbb{V}(X). \checkmark$$

Teorema 7.11. *Si $X \in L^2$, entonces $\mathbb{V}(X) = 0$ si y sólo si $X = \mathbb{E}(X)$ casi seguramente.*

Demostración.
$\Leftarrow$) Si $X = \mathbb{E}(X) = \mu$, entonces $\mathbb{V}(X) = \mathbb{E}(\mu^2) - \mathbb{E}^2(\mu) = \mu^2 - \mu^2 = 0$.
$\Rightarrow$) Si $\mathbb{V}(X) = 0$, entonces $\mathbb{E}\left[X - \mathbb{E}(X)^2\right] = 0$ y como $(X - \mathbb{E}(X))^2 \geq 0$ casi seguramente y tiene esperanza 0, entonces debe ser $(X - \mathbb{E}(X))^2 \geq 0$ c.s., por lo que debe ser $X = \mathbb{E}(X)$ casi seguramente. ✓

Ejemplo 7.12. Si $X \sim \text{Ber}(p)$, entonces ya vimos que $\mathbb{E}(X) = p$. Además $\mathbb{E}(X^2) = p$ con lo cual obtenemos $\mathbb{V}(X) = p - p^2 = p(1-p)$.

Ejemplo 7.13. Si $X \sim \text{N}(\mu, \sigma^2)$, entonces ya vimos que $\mathbb{E}(X) = \mu$. Ahora, si integramos por partes

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x^2 e^{\frac{-1}{2\sigma^2}(x-\mu)^2} dx$$

nos da $\mu^2 + \sigma^2$, por lo tanto $\mathbb{V}(X) = \sigma^2$.

Veremos en lo que sigue, algunas desigualdades que son muy útiles en la teoría y en la práctica, conocidas como desigualdades de Markov y de Chebyshev.

Teorema 7.14. *Dadas X variable aleatoria, $g : \mathbb{R} \to \mathbb{R}$ monótona creciente, tal que $g(X) \in L^1$, $g \geq 0$ y $a \in \mathbb{R}$ tal que $g(a) > 0$, entonces*

$$P(X > a) \leq \frac{1}{g(a)} \mathbb{E}(g(X)).$$

Demostración.
Consideramos el conjunto $A = \{X > a\}$, entonces, dado que $g \geq 0$, obtenemos que

$$\mathbb{E}(g(X)) = \mathbb{E}(g(X)\mathbf{1}_A) + \mathbb{E}(g(X)\mathbf{1}_{A^c}) \geq \mathbb{E}(g(X)\mathbf{1}_A).$$

Puesto que $g(X)\mathbf{1}_A \geq g(a)\mathbf{1}_A$, ya que g es monótona creciente y por definición del conjunto A, vemos que

$$\mathbb{E}(g(X)\mathbf{1}_A) \geq \mathbb{E}(g(a)\mathbf{1}_A) = g(a)\mathbb{E}(\mathbf{1}_A) = g(a)P(A) = g(a)P(X > a).✓$$

Observación 7.15. *Surge de la demostración, que vale la misma acotación si la probabilidad que se considera es $P(X \geq a)$.*

Corolario 7.16. ***Desigualdad de Markov.*** *Si $X \in L^p$ $(p > 0)$ y $a > 0$, entonces*

$$P(|X| > a) \leq \frac{1}{a^p} \mathbb{E}(|X|^p).$$

Demostración.
Basta tomar $g(x) = x^p$ para $x > 0$ y $g(x) = 0$ para $x \leq 0$ y aplicar la desigualdad anterior a la variable $Y = |X|$.

Corolario 7.17. ***Desigualdad de Chebyshev.*** *Si $X \in L^2$ y $a > 0$, entonces*

$$P\left(|X - \mathbb{E}(X)| > a\right) \leq \frac{1}{a^2}\mathbb{V}(X).$$

Demostración.
Basta usar la desigualdad del corolario anterior, para el caso en que $p = 2$ y para la variable $Y = X - \mathbb{E}(X)$.✓

Observación 7.18. *Como se ve, la desigualdad de Markov nos proporciona una cota para la función de distribución de una variable aleatoria, si se conoce únicamente el momento de algún orden de la variable, por ejemplo, el momento de orden uno.*

Observación 7.19. *La desigualdad de Chebyshev es equivalente a*

$$P\left(|X - \mathbb{E}(X)| \leq a\right) \geq 1 - \frac{1}{a^2}\mathbb{V}(X).$$

y por lo tanto, nos proporciona una cota inferior para la probabilidad de que la variable tome valores en un entorno de su valor esperado, conociendo únicamente el valor esperado y la varianza de la variable.

Observación 7.20. *Las desigualdades de Markov y de Chebyshev, son cotas universales, es decir se cumplen para cualquier tipo de variable aleatoria (con la sóla hipótesis de que admitan momentos de algún orden), por lo que suelen dar cotas groseras de las probabilidades. En cada situación particular, conociendo más información sobre la variable aleatoria X, se suelen conseguir cotas más finas.*

7.3. Covarianza y coeficiente de correlación.

La covarianza y el coeficiente de correlación que definiremos en lo que sigue, sirven como medidas del grado de asociación que hay entre dos variables aleatorias X e Y, ambos conceptos están relacionados como veremos con la independencia entre las variables.

Definición 7.21. Covarianza entre dos variables aleatorias.
Si $X, Y \in L^2$, entonces definimos $\mathbb{COV}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right].$

Propiedades.

1. Si $X, Y \in L^2$, entonces $\mathbb{COV}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$
2. Si $X, Y \in L^2$, entonces $\mathbb{COV}(X,Y) = \mathbb{COV}(Y,X).$
3. Si $X \in L^2$, entonces $\mathbb{COV}(X,X) = \mathbb{V}(X).$
4. Si $X, Y \in L^2$, entonces $\mathbb{COV}(aX + b, Y) = a\mathbb{COV}(X,Y)$ para todos $a, b \in \mathbb{R}$.

5. Si $X, Y, Z \in L^2$, entonces $\mathbb{COV}(X+Y, Z) = \mathbb{COV}(X, Y) + \mathbb{COV}(Y, Z)$.

6. Si $X, Y \in L^2$ y son independientes, entonces $\mathbb{COV}(X, Y) = 0$.

7. Si $X_1, X_2, ..., X_n \in L^2$, entonces

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i) + 2\sum_{i<j} \mathbb{COV}(X_i, X_j).$$

Observación 7.22. $\mathbb{COV}(X, Y) = 0$ *no implica necesariamente que* X *e* Y *sean independientes. Se deja como ejercicio construir un contraejemplo.*

Observación 7.23. *Si* $X_1, X_2, ..., X_n \in L^2$ *son independientes, entonces*

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i).$$

Observación 7.24. *Si* $X, Y \in L^2$*, entonces*

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathbb{COV}(X, Y).$$

Las demostraciones son simplemente operativas y se dejan como ejercicio. Haremos igualmente la demostración de la propiedad 7.

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \mathbb{COV}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{COV}(X_i, X_j)$$

y usando que $\mathbb{COV}(X_i, X_j) = \mathbb{COV}(X_j, X_i)$ y que $\mathbb{COV}(X_i, X_i) = \mathbb{V}(X_i)$, obtenemos

$$\sum_{i=1}^{n} \mathbb{COV}(X_i, X_i) + \sum_{i=1}^{n}\sum_{j\neq i} \mathbb{COV}(X_i, X_j) = \sum_{i=1}^{n} \mathbb{V}(X_i) + 2\sum_{j<i} \mathbb{COV}(X_i, X_j). \checkmark$$

Ejemplo 7.25. Si $X \sim \text{Bin}(n, p)$ entonces vimos que $X = X_1 + X_2 + ... + X_n$ donde las X_i son $\text{Ber}(p)$ e independientes, por lo tanto

$$\mathbb{V}(X) = \mathbb{V}(X_1 + X_2 + ... + X_n) = \mathbb{V}(X_1) + \mathbb{V}(X_2) + ... + \mathbb{V}(X_n) = np(1-p).$$

Definición 7.26. Coeficiente de correlación entre dos variables aleatorias. Si $X, Y \in L^2$ son no constantes, entonces definimos $\rho(X, Y) = \frac{\mathbb{COV}(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$.

Propiedades.
En las propiedades que siguen se consideran $X, Y \in L^2$ no constantes.

1. $-1 \leq \rho(X, Y) \leq 1$.

2. $\rho(X, Y) = 1$ si y sólo si existen $a, b \in \mathbb{R}$, $a > 0$, tales que $Y = aX + b$.

3. $\rho(X,Y) = -1$ si y sólo si existen $a, b \in \mathbb{R}$, $a < 0$, tales que $Y = aX + b$.

4. Si X, Y son independientes, entonces $\rho(X,Y) = 0$.

Demostración.
Aplicando la desigualdad de Cauchy Schwartz, tenemos que

$$|\mathbb{COV}(X,Y)| = |\mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]| \leq$$

$$\sqrt{\mathbb{E}(X - \mathbb{E}(X))^2\,\mathbb{E}(Y - \mathbb{E}(Y))^2} = \sqrt{\mathbb{V}(X)\,\mathbb{V}(Y)}.$$

lo cual es equivalente a decir que $|\rho(X,Y)| \leq 1$. Además sabemos que $|\rho(X,Y)| = 1$ si y sólo si existe $\lambda \in \mathbb{R}$ tal que $X - \mathbb{E}(X) = \lambda(Y - \mathbb{E}(Y))$ donde $\lambda \neq 0$ ya que X no es constante. Por lo tanto $|\rho(X,Y)| = 1$ si y sólo si existen $a \neq 0$ y b tales que $Y = aX + b$. Ahora, aplicando las propiedades de varianza y covarianza, obtenemos que $\rho(X,Y) = \rho(X, aX + b) = \frac{a}{|a|}$ de donde se deduce que $\rho(X,Y)$ es 1 si y sólo si $a > 0$, y -1 si y sólo si $a < 0$. Quedan probadas así las primeras 3 propiedades. La última propiedad es evidente ya que $\rho(X,Y) = 0$ si y sólo si $\mathbb{COV}(X,Y) = 0$. ✓

7.4. Variables i.i.d.

Definición 7.27. Se dice que la sucesión de variables aleatorias $X_1, X_2, ..., X_n, ...$ son v.a.i.i.d, cuando dichas variables son independientes y todas tienen igual función de distribución, es decir cuando son independientes y además $F_{X_1} = F_{X_2} = ... = F_{X_n}$ para todo n.

Cuando n es fijo, se dice también que $X_1, X_2, ..., X_n$ son una M.A.S.c/rep de X de tamaño n (muestra aleatoria simple con reposición). Lo cual significa que las variables son i.i.d con distribución como la de cierta variable X que se toma como representativa.
Supongamos que tenemos $X_1, X_2, ..., X_n$ v.a.i.i.d cuya distribución es como la de cierta $X \in L^2$. Llamémosle en este caso μ y σ^2 a la esperanza y la varianza de X respectivamente. Es decir que $\mathbb{E}(X) = \mu$ y $\mathbb{V}(X) = \sigma^2$.
Se define la media muestral como la siguiente variable aleatoria: $\overline{X}_n := \frac{X_1 + X_2 + ... + X_n}{n}$.
La misma es fundamental desde el punto de vista estadístico, ya que si $X_1, X_2, ..., X_n$ representan n observaciones obtenidas de forma independiente de una cierta variable aleatoria, lo que se llama también una muestra aleatoria simple de tamaño n, entonces $\overline{X}_n$ nos da el promedio de las observaciones obtenidas de la muestra.
Veremos ahora que si $X \in L^2$, entonces $\mathbb{E}(\overline{X}_n) = \mu$ y $\mathbb{V}(\overline{X}_n) = \sigma^2/n$.
Efectivamente, usando la linealidad de la esperanza obtenemos que

$$\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{\mathbb{E}(X_1) + \mathbb{E}(X_2) + ... + \mathbb{E}(X_n)}{n} = \frac{n\mu}{n} = \mu.$$

Ahora, aplicando propiedades de varianza, obtenemos que

$$\mathbb{V}(\overline{X}_n) = \mathbb{V}\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n^2}\mathbb{V}(X_1 + X_2 + ... + X_n) =$$

$$\frac{\mathbb{V}(X_1)+\mathbb{V}(X_2)+...+\mathbb{V}(X_n)}{n^2}=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}.$$

Una aplicación estadística.

Supongamos que deseamos estimar el porcentaje de fumadores en una población. Para obtener el resultado, se encuestarán de manera independiente, n individuos de la población y se calculará el porcentaje de fumadores en la muestra. Podemos pensar entonces que tenemos n variables aleatorias $X_1, X_2, ..., X_n$, definidas como $X_i = \begin{cases} 1 & \text{si la i-ésima persona encuestada fuma} \\ 0 & \text{si no} \end{cases}$. Entonces las variables son independientes con distribución Ber(p), donde p es el porcentaje de fumadores en la población. p es desconocido, que estimaremos mediante el porcentaje de fumadores en la muestra, el cual es $\overline{X}_n = \frac{X_1+,X_2+...+X_n}{n}$ ya que el numerador cuenta el total de fumadores (éxitos).

Supongamos que queremos respondernos a la siguiente pregunta: ¿a cuántos individuos hay que encuestar si deseamos que el porcentaje de la muestra no difiera del real en más de un 1 % con una probabilidad mayor al 95 %?

Por lo tanto queremos hallar n tal que $P\left(\left|\overline{X}_n - p\right| \leq 0,01\right) \geq 0,95$.

Observamos que las variables, al ser Bernoulli están en L^2 y ya vimos que tienen valor esperado p y varianza $p(1-p)$.

Por otro lado, ya vimos que el valor esperado de $\overline{X}_n$ coincide con el de cada X_i, y la varianza de $\overline{X}_n$ es $\sigma^2/n = p(1-p)/n$. O sea que en el caso de las variables Bernoulli, tenemos que $\mathbb{E}\left(\overline{X}_n\right) = p$ y $\mathbb{V}\left(\overline{X}_n\right) = p(1-p)/n$.

Aplicando el corolario 7.16 (desigualdad de Chebyshev) a la variable $\overline{X}_n$, llegamos a que

$$P\left(\left|\overline{X}_n - p\right| \leq 0,01\right) \geq 1 - \frac{1}{0,01^2}\mathbb{V}\left(\overline{X}_n\right) = 1 - \frac{p(1-p)}{n0,01^2}.$$

Puesto que $p(1-p) \leq 1/4$ para todo valor de p, obtenemos que

$$P\left(\left|\overline{X}_n - p\right| \leq 0,01\right) \geq 1 - \frac{p(1-p)}{n0,01^2} \geq 1 - \frac{1}{4n0,01^2}.$$

Entonces eligiendo n tal que $1 - \frac{1}{4n0,01^2} \geq 0,95$, el mismo nos asegurará que $P\left(\left|\overline{X}_n - p\right| \leq 0,01\right) \geq 0,95$. En este caso el menor valor de n que nos asegura esta desigualdad es 50.000.

Capítulo 8

Convergencia en probabilidad, casi segura y en distribución.

Consideremos una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ y una variable aleatoria X definidas sobre un mismo espacio de probabilidad. Dado que las X_n y la X son funciones de Ω en $\mathbb{R}$, hay varias nociones de convergencia de una sucesión de funciones a una función, como la convergencia puntual, la uniforme, la convergencia cuadrática o en el espacio L^p por ejemplo. En teoría de probabilidad, dado que las funciones son aleatorias, es decir que toman valores reales de manera aleatoria, es necesario definir nuevos conceptos de convergencia que involucren el cálculo de la probabilidad de que las X_n esten próximas a X en algún sentido. Definiremos tres conceptos de convergencia que son vitales en teoría de la probabilidad y en estadística matemática, que son la convergencia en probabilidad, la convergencia casi segura y la convergencia en distribución.

8.1. Convergencia en probabilidad y casi segura.

Definición 8.1. Convergencia en probabilidad.
Dadas una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ y una variable aleatoria X definidas sobre cierto $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad, se dice que la sucesión $\{X_n\}_{n\in\mathbb{N}}$ converge en probabilidad a X si y sólo si, para todo $\varepsilon > 0$ se cumple que

$$\lim_{n\to+\infty} P\left(|X_n - X| < \varepsilon\right) = 1.$$

Notación: $X_n \xrightarrow{P} X$.

Observación 8.2. *Equivalentemente, tenemos que* $X_n \xrightarrow{P} X$ *si y sólo si para todo* $\varepsilon > 0$ *se cumple que*

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

Informalmente, la convergencia en probabilidad nos dice que una vez que fijamos el valor de $\varepsilon > 0$ arbitrariamente pequeño, pero fijo, la probabilidad de que X_n tome un valor perteneciente al intervalo $(X-\varepsilon, X+\varepsilon)$ se acerca a uno en la medida de que n se tome suficientemente grande.

Definición 8.3. Convergencia casi segura.
Dadas una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ y una variable aleatoria X definidas sobre cierto $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad se dice que la sucesión $\{X_n\}_{n\in\mathbb{N}}$ converge casi seguramente a X (o en casi todo punto) si y sólo si se cumple que

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

Notación: $X_n \overset{c.s.}{\to} X$.

Observación 8.4. *Dado que el límite de variables aleatorias es variable aleatoria, se verifica que* $\{limX_n = X\}$ *es un suceso.*

Teorema 8.5. $X_n \overset{c.s.}{\to} X$ *si y sólo si* $\underset{k\to+\infty}{lim} P\left(\bigcap_{n=k}^{+\infty}\{|X_n - X| < \varepsilon\}\right) = 1$ *para todo* $\varepsilon > 0$.

Demostración.
Si $w \in \Omega$ es tal que $\lim_{n\to+\infty} X_n(w) = X(w)$ entonces, para todo $\varepsilon > 0$, existe un k tal que para todo $n \geq k$ se cumple que $|X_k(w) - X(w)| < \varepsilon$. Observando que es suficiente en la definición de límite considerar $\varepsilon \in \mathbb{Q}^+$ entonces tenemos que $P\left(\lim_{n\to+\infty} X_n = X\right) = 1$ si y solo si

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

Como la intersección en el conjunto de $\varepsilon \in \mathbb{Q}^+$ es numerable, y tiene probabilidad 1, entonces la última condición es equivalente a

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

Por otro lado, los conjuntos $B_k = \bigcap_{n=k}^{+\infty}\{|X_n - X| < \varepsilon\}$ forman una sucesión creciente de sucesos, entonces, la propiedad de continuidad de las probabilidades nos dice que $P\left(\bigcup_{k=1}^{+\infty} B_k\right) = \lim_{k\to+\infty} P(B_k)$, por lo que

$$P\left(\bigcup_{k=1}^{+\infty}\bigcap_{n=k}^{+\infty}\{|X_n - X| < \varepsilon\}\right) = \lim_{k\to+\infty} P\left(\bigcap_{n=k}^{+\infty}\{|X_n - X| < \varepsilon\}\right).$$

Llegamos así a que

$$X_n \overset{c.s.}{\to} X \text{ si y sólo si} \lim_{k\to+\infty} P\left(\bigcap_{n=k}^{+\infty}\{|X_n - X| < \varepsilon\}\right) = 1 \text{ para todo } \varepsilon \in \mathbb{Q}^+.$$

Finalmente, dado que en la defición de límite es equivalente a trabajar con $\varepsilon > 0$ y observando la demostración, se deduce que $X_n \overset{c.s.}{\to} X$ si y sólo si
$\lim\limits_{k \to +\infty} P\left(\bigcap_{n=k}^{+\infty} \{|X_n - X| < \varepsilon\}\right) = 1$ para todo $\varepsilon > 0$. ✓

Observación 8.6. *La intersección sobre los $\varepsilon \in \mathbb{Q}^+$ se realiza para que podamos asegurar que los conjuntos con los que trabajamos pertenezcan a la σ-álgebra, de otro modo si trabajamos con los $\varepsilon > 0$, la intersección es no numerable y no podemos asegurar que la misma pertenezca a la σ-álgebra.*

Teorema 8.7. *Dados un $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad, una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ y una variable aleatoria X.*

$$\text{Si } X_n \overset{c.s.}{\to} X \text{ entonces } X_n \overset{P}{\to} X.$$

Demostración.
Como $X_n \overset{c.s.}{\to} X$, entonces fijado $\varepsilon > 0$, entonces para todo $k \in \mathbb{N}$ se cumple que

$$\bigcap_{n=k}^{+\infty} \{|X_n - X| < \varepsilon\} \subset \{|X_k - X| < \varepsilon\}$$

entonces

$$P\left(\bigcap_{n=k}^{+\infty} \{|X_n - X| < \varepsilon\}\right) \leq P\left(|X_k - X| < \varepsilon\right)$$

por lo que tomando límite cuando k tiende a $+\infty$ se deduce el resultado. ✓
Veremos en el siguiente ejemplo que la noción de convergencia casi segura es estrictamente más fuerte que la de convergencia en probabilidad.

Ejemplo 8.8. Tomemos un espacio de probabilidad en el cual definimos una variable $Y \sim U\left(0,1\right)$. Consideramos la sucesión de intervalos $I_{m,k} = \left(\frac{k}{2^m}, \frac{k+1}{2^m}\right)$ para $m = 1, 2, 3, ...$ y $k = 0, 1, 2, 3, ..., 2^m - 1$. Definimos I_n ordenando los $I_{m,k}$ dando primero el valor de m y luego, para dicho m, variamos en los distintos valores de $k = 0, 1, 2, 3, ..., 2^m - 1$. Es decir, para $m = 1$, tenemos $k = 0, 1$ por lo que definimos $I_1 = I_{1,0} = \left(0, \frac{1}{2}\right)$; $I_2 = I_{1,1} = \left(\frac{1}{2}, 1\right)$. Luego, para $m = 2$, tenemos $k = 0, 1, 2, 3$ con lo que definimos I_3, I_4, I_5 e I_6 como sigue: $I_3 = I_{2,0} = \left(0, \frac{1}{4}\right)$; $I_4 = I_{2,1} = \left(\frac{1}{4}, \frac{1}{2}\right)$; $I_5 = I_{2,2} = \left(\frac{1}{2}, \frac{3}{4}\right)$ e $I_6 = I_{2,3} = \left(\frac{3}{4}, 1\right)$. Así continuamos sucesivamente.

Definimos ahora la sucesión $X_n = \mathbf{1}_{I_n}(Y)$. Las longitudes de los intervalos I_n tienden a cero por lo que se podría esperar que exista algún tipo de convergencia de las X_n a cero. Dado $\varepsilon > 0$, se tiene que $P\left(|X_n| \geq \varepsilon\right) = \begin{cases} P\left(Y \in I_n\right) & \text{si } \varepsilon < 1 \\ 0 & \text{si } \varepsilon \geq 1 \end{cases}$ y como $P\left(Y \in I_n\right)$ =longitud de $I_n \to 0$, entonces tenemos que $X_n \overset{P}{\to} 0$.
Por otro lado, vemos que cualquier número $\in (0,1)$ pertenece a infinitos de los intervalos I_n y también no pertenece a infinitos de los intervalos I_n. Entonces dado cualquier $w \in \Omega$, se tendrá que $Y\left(w\right) \in (0,1)$ y por lo tanto no existe $\lim\limits_{n\to+\infty} X_n(w)$.
Entonces $\left\{\lim\limits_{n\to+\infty} X_n = 0\right\} = \Phi$ lo cual prueba que X_n no converge casi seguramente a cero.

Algebra de límites en las convergencias en probabilidad y casi segura.
En las siguientes propiedades se consideran dadas las sucesiones de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}, \{Y_n\}_{n\in\mathbb{N}}$ y las variables aleatorias X e Y definidas sobre cierto $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad. Se deja como ejercicio su demostración.

1. Unicidad. Si $X_n \stackrel{P}{\to} X$, $X_n \stackrel{P}{\to} Y$ entonces $X = Y$ c.s.
2. Unicidad. Si $X_n \stackrel{c.s.}{\to} X$, $X_n \stackrel{c.s.}{\to} Y$ entonces $X = Y$ c.s.
3. Si $X_n \stackrel{P}{\to} X$, $Y_n \stackrel{P}{\to} Y$ entonces $\alpha X_n + \beta Y_n \stackrel{P}{\to} \alpha X + \beta Y$ para todos $\alpha, \beta \in \mathbb{R}$.
4. Si $X_n \stackrel{c.s.}{\to} X$, $Y_n \stackrel{c.s.}{\to} Y$ entonces $\alpha X_n + \beta Y_n \stackrel{c.s.}{\to} \alpha X + \beta Y$ para todos $\alpha, \beta \in \mathbb{R}$.
5. Si $X_n \stackrel{P}{\to} X$ y $g : \mathbb{R} \to \mathbb{R}$ es continua, entonces $g(X_n) \stackrel{P}{\to} g(X)$.
6. Si $X_n \stackrel{c.s.}{\to} X$ y $g : \mathbb{R} \to \mathbb{R}$ es continua, entonces $g(X_n) \stackrel{c.s.}{\to} g(X)$.
7. Si $X_n \stackrel{P}{\to} X$, $Y_n \stackrel{P}{\to} Y$ entonces $X_n Y_n \stackrel{P}{\to} XY$.
8. Si $X_n \stackrel{P}{\to} X$, $Y_n \stackrel{P}{\to} Y$ y $P(Y \neq 0) = 1$, entonces $X_n Y_n \stackrel{P}{\to} XY$.
9. Si $X_n \stackrel{c.s.}{\to} X$, $Y_n \stackrel{c.s.}{\to} Y$ y $P(Y \neq 0) = 1$, entonces $X_n Y_n \stackrel{c.s.}{\to} XY$.
10. Si $X_n \stackrel{c.s.}{\to} X$, $Y_n \stackrel{c.s.}{\to} Y$ entonces $X_n Y_n \stackrel{c.s.}{\to} XY$.
11. Si $X_n \stackrel{P}{\to} 0$, existe $k \in \mathbb{R}$ tal que $P(|Y_n| > k) = 0$ para todo n, entonces $X_n Y_n \stackrel{P}{\to} 0$.
12. Si $X_n \stackrel{c.s.}{\to} 0$, existe $k \in \mathbb{R}$ tal que $P(|Y_n| > k) = 0$ para todo n, entonces $X_n Y_n \stackrel{c.s.}{\to} 0$.

8.2. Leyes de los grandes números.

Teorema 8.9. ***Ley débil de los grandes números.***
Dado un $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad. Si las variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ son i.i.d con distribución como la de cierta $X \in L^2$ y le llamamos $\mu = \mathbb{E}(X)$ y $\sigma^2 = \mathbb{V}(X)$.
Entonces

$$\overline{X}_n \stackrel{P}{\to} \mu.$$

Demostración.
Ya vimos sobre el final del capítulo anterior cuando las varaibles son i.i.d. que $\mathbb{E}\left(\overline{X}_n\right) = \mu$ y $\mathbb{V}\left(\overline{X}_n\right) = \sigma^2/n$ para todo n. Entonces aplicando la desigualdad de Chebyshev, obtenemos que, para todo $\varepsilon > 0$,

$$P\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) \leq \frac{\mathbb{V}\left(\overline{X}_n\right)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \underset{n\to+\infty}{\to} 0$$

por lo que $\overline{X}_n \stackrel{P}{\to} \mu$.✓

Observación 8.10. *Como se ve repasando la definición, la misma demostración funciona cambiando las hipótesis de i.i.d por las de que todas las variables, tengan iguales esperanza y varianza, y además sean no correlacionadas.*

Teorema 8.11. ***Ley fuerte de los grandes números.***
Dado un $\left(\Omega, \mathcal{A}, P\right)$ espacio de probabilidad. Si las variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ son i.i.d con distribución como la de cierta $X \in L^4$ y le llamamos $\mu = \mathbb{E}\left(X\right)$.
Entonces

$$\overline{X}_n \stackrel{c.s.}{\rightarrow} \mu.$$

Demostración.
Basta probar el teorema para el caso en que $\mu = 0$, ya que una vez que lo tenemos probado en este caso, para deducir el caso general, definimos para cada n, $Y_n = X_n - \mu$, entonces la sucesión $\{Y_n\}_{n\in\mathbb{N}}$ es i.i.d con distribución como la de $Y = X - \mu$, entonces, $\overline{Y}_n \stackrel{c.s.}{\rightarrow} \mathbb{E}(Y) = 0$, pero $\overline{Y}_n = \overline{X}_n - \mu$, por lo tanto $\overline{X}_n \stackrel{c.s.}{\rightarrow} \mu$.
Suponemos entonces que $\mu = 0$.
Para probar que $\overline{X}_n \stackrel{c.s.}{\rightarrow} 0$, según el teorema 8.4 debemos probar que, dado $\varepsilon > 0$, se cumple que $\lim\limits_{k\to+\infty} P\left(\bigcap_{n=k}^{+\infty}\left\{\left|\overline{X}_n\right| < \varepsilon\right\}\right) = 1$, lo cual es equivalente a probar que $\lim\limits_{k\to+\infty} P\left(\bigcup_{n=k}^{+\infty}\left\{\left|\overline{X}_n\right| > \varepsilon\right\}\right) = 0$.
Dado que $P\left(\bigcup_{n=k}^{+\infty}\left\{\left|\overline{X}_n\right| > \varepsilon\right\}\right) \leq \sum_{n=k}^{+\infty} P\left(\left|\overline{X}_n\right| > \varepsilon\right)$ se deduce que para obtener el resultado es suficiente con probar que $\sum_{n=1}^{+\infty} P\left(\left|\overline{X}_n\right| > \varepsilon\right) < +\infty$.
La idea será entonces acotar $P\left(\left|\overline{X}_n\right| > \varepsilon\right)$ superiormente por una sucesión cuya serie sea convergente.
Como $X \in L^4$, usaremos la desigualdad de Markov con $p = 4$, por lo que

$$P\left(\left|\overline{X}_n\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^4}\mathbb{E}\left(\overline{X}_n^4\right).$$

Por lo tanto será suficiente probar que $\sum_{n=1}^{+\infty} \mathbb{E}\left(\overline{X}_n^4\right) < +\infty$.
$\mathbb{E}\left(\overline{X}_n^4\right) =$

$$\frac{1}{n^4}\mathbb{E}\left[\left(X_1 + X_2 + ... + X_n\right)\left(X_1 + X_2 + ... + X_n\right)\left(X_1 + X_2 + ... + X_n\right)\left(X_1 + X_2 + ... + X_n\right)\right].$$

Desarrolando esta suma, y aplicando linealidad del valor esperado, obtenemos que

$$\mathbb{E}\left[\left(X_1 + X_2 + ... + X_n\right)\left(X_1 + X_2 + ... + X_n\right)\left(X_1 + X_2 + ... + X_n\right)\left(X_1 + X_2 + ... + X_n\right)\right] =$$

$$\sum_{i=1}^{n}\mathbb{E}\left(X_i^4\right) + \sum_{i,j\,:\,i\neq j}\mathbb{E}\left(X_i^3 X_j\right) + \sum_{i,j\,:\,i\neq j}\mathbb{E}\left(X_i^2 X_j^2\right)$$

$$+ \sum_{i,j,k\,:\,i\neq j\neq k,\ i\neq k}\mathbb{E}\left(X_i^2 X_j X_k\right) + \sum_{i,j,k,l\,:\,i\neq j\neq k\neq l,\ j\neq l,\ i\neq k,\ i\neq l}\mathbb{E}\left(X_i X_j X_k X_l\right).$$

Como las variables son i.i.d, tenemos que dentro de cada una de las sumatorias anteriores, los sumandos son todos iguales entre sí, entonces nos queda igual a

$$n\mathbb{E}\left(X_1^4\right)+8C_2^n\mathbb{E}\left(X_1^3X_2\right)+C_2^4C_2^n\mathbb{E}\left(X_1^2X_2^2\right)+6C_2^4C_3^n\mathbb{E}\left(X_1^2X_2X_3\right)+4!C_4^n\mathbb{E}\left(X_1X_2X_3X_4\right).$$

Ahora usando que las variables son i.i.d y recordando que en estos casos, la esperanza de un producto se factoriza como el producto de esperanzas, observamos que $\mathbb{E}\left(X_1^3X_2\right) = \mathbb{E}\left(X_1^3\right)\mathbb{E}\left(X_2\right) = 0$, $\mathbb{E}\left(X_1^2X_2X_3\right) = \mathbb{E}\left(X_1^2\right)\mathbb{E}\left(X_2\right)\mathbb{E}\left(X_3\right) = 0$ y $\mathbb{E}\left(X_1X_2X_3X_4\right) = \mathbb{E}\left(X_1\right)\mathbb{E}\left(X_2\right)\mathbb{E}\left(X_3\right)\mathbb{E}\left(X_4\right) = 0$.
Entonces

$$\mathbb{E}\left(\overline{X}_n^4\right) = \frac{1}{n^4}\left(n\mathbb{E}\left(X_1^4\right) + 3n(n-1)\mathbb{E}\left(X_1^2\right)\mathbb{E}\left(X_2^2\right)\right)$$

por lo que

$$\sum_{n=1}^{+\infty}\mathbb{E}\left(\overline{X}_n^4\right) \sim \sum_{n=1}^{+\infty}\frac{1}{n^2} < +\infty.\checkmark$$

Trabajando con desigualdades más finas, lo cual lleva más trabajo, es posible demostrar que vale el mismo teorema sólo pidiendo que $X \in L^1$. Por lo tanto cuando sea necesaria aplicar la ley, lo haremos simplemente verificando que $X \in L^1$.
Si las variables $\{X_n\}_{n\in\mathbb{N}}$ son i.i.d con distribución como la de cierta $X \notin L^1$, entonces, también tenemos una versión de la ley fuerte.

Teorema 8.12. *Dado un $(\Omega, \mathcal{A}, P)$ espacio de probabilidad. Si las variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ son i.i.d con distribución como la de cierta X tal que $\mathbb{E}\left(|X|\right) = +\infty$, entonces*

$$limsup\ \left|\overline{X}_n\right| = +\infty\ c.s.$$

Demostración.
Como $\mathbb{E}\left(|X|\right) = +\infty$, entonces $\mathbb{E}\left(\frac{|X|}{k}\right) = +\infty$ para todo $k = 1, 2, 3, ...$ Entonces $\sum_{n=1}^{+\infty} P\left(\frac{|X|}{k} \geq n\right) = +\infty$, para todo $k = 1, 2, 3, ...$
Como las variables son idénticamente distribuidas, tenemos que

$$\sum_{n=1}^{+\infty} P\left(\frac{|X|}{k} \geq n\right) = \sum_{n=1}^{+\infty} P\left(\frac{|X_n|}{k} \geq n\right) = \sum_{n=1}^{+\infty} P\left(\frac{|X_n|}{n} \geq k\right) = +\infty \text{ para todo } k = 1, 2, 3, ...$$

Fijado k, se tiene que los sucesos $A_n^{(k)} = \left\{\frac{|X_n|}{n} \geq k\right\}$ son independientes, luego, por el lema de Borel-Cantelli se tiene que

$$P\left(\text{ocurren infinitos } A_n^{(k)}\right) = 1 \text{ para todo } k = 1, 2, 3, ...$$

Entonces, si definimos B_k = "ocurren infinitos $A_n^{(k)}$", tenemos que $P\left(B_k\right) = 1$ para todo $k = 1, 2, 3, ...$ y como intersección numerable de sucesos de probabilidad 1, tiene probabilidad 1, obtenemos que $P\left(\bigcap_{k=1}^{+\infty} B_k\right) = 1$.

Observamos además que $B = \bigcap_{k=1}^{+\infty} B_k =$ "ocurre $A_n^{(k)}$ para infinitos valores de n, para todo k"$= \left\{ \left\{ \frac{|X_n|}{n} \right\}_{n\in\mathbb{N}} \text{ es no acotada} \right\}$. Entonces $P\left(\left\{ \frac{|X_n|}{n} \right\}_{n\in\mathbb{N}} \text{ es no acotada} \right) = 1$.
Ya que existe probabilidad 1 de que la sucesión $\left\{ \frac{|X_n|}{n} \right\}_{n\in\mathbb{N}}$ sea no acotada, para terminar la prueba, definimos $S_n = X_1 + X_2 + ... + X_n$, y bastará con probar que si $\left\{ \frac{|X_n|}{n} \right\}_{n\in\mathbb{N}}$ es no acotada, entonces $\left\{ \left|\overline{X}_n\right| = \frac{|S_n|}{n} \right\}_{n\in\mathbb{N}}$ es no acotada.
Efectivamente, si fuera $\left\{ \frac{|S_n|}{n} \right\}_{n\in\mathbb{N}}$ acotada, entonces también lo sería $\left\{ \frac{|S_{n-1}|}{n} \right\}_{n\in\mathbb{N}}$ ya que $\frac{|S_{n-1}|}{n} = \frac{|S_{n-1}|}{n-1}\frac{n-1}{n}$, entonces, $\frac{|X_n|}{n} = \frac{|S_n - S_{n-1}|}{n} \leq \frac{|S_n|}{n} + \frac{|S_{n-1}|}{n}$, sería acotada, por lo tanto
$\left\{ \frac{|X_n|}{n} \right\}_{n\in\mathbb{N}}$ es acotada lo cual es absurdo. ✓

8.2.1. Aplicaciones.

La cantidad de aplicaciones de la ley fuerte es enorme, veremos en lo que sigue, a modo de ejemplo, algunos corolarios de la ley a modo de aplicación de la misma.

Corolario 8.13. *Si las variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ son i.i.d con distribución Ber(p), entonces*

$$\overline{X}_n \overset{c.s.}{\rightarrow} p.$$

Demostración.
Es obvia ya que las variables Ber(p) están en L^1 y son tales que $\mathbb{E}(X) = p$. ✓
Frecuentemente, en estadística, se tiene un muestreo de alguna variable aleatoria cuya función de distribución es desconocida. Se desea "estimar" a la función F_X dada una muestra aleatoria simple $X_1, X_2, ..., X_n$.
Supongamos entonces que tenemos $X_1, X_2, ..., X_n$, variables aleatorias i.i.d con distribución como la de X. Se define a la distribución empírica asociada a la muestra, a la función $F_n^* : \mathbb{R} \rightarrow \mathbb{R}$ tal que $F_n^*(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{(-\infty,x]}(X_i)$.
Observamos que $\mathbf{1}_{(-\infty,x]}(X_1), \mathbf{1}_{(-\infty,x]}(X_2), ..., \mathbf{1}_{(-\infty,x]}(X_n)$ son independientes (porque las X_i lo son) con distribución Ber$(p = F_X(x))$.
Observamos que $F_n^* : \mathbb{R} \rightarrow \mathbb{R}$ es una función de distribución escalonada, con saltos en los X_i y donde cada salto es de longitud $1/n$ (en el caso en que las X_i sean todas distintas).

Corolario 8.14. ***Aplicación estadística: estimación de una función de distribución desconocida.***
F_n^* converge puntualmente a F_X.
Demostración.
Aplicamos la ley fuerte de los grandes números, se cumple que fijado $x \in \mathbb{R}$, entonces $F_n^*(x) \overset{c.s.}{\rightarrow} \mathbb{E}\left(\mathbf{1}_{(-\infty,x]}(X)\right) = F_X(x)$.✓

Corolario 8.15. ***Cálculo de integrales mediante números aleatorios.***
Dadas $f : [a, b] \to \mathbb{R}$ continua, y $\{X_n\}_{n\in\mathbb{N}}$ i.i.d con distribución $U(a, b)$. Entonces

$$\frac{b-a}{n}\sum_{i=1}^{n} f(X_i) \overset{c.s.}{\to} \int_a^b f(x)dx.$$

Demostración.
Si definimos para cada n las variables $Y_n = (b-a)f(X_n)$, entonces, tendremos que $\{Y_n\}_{n\in\mathbb{N}}$ son i.i.d en L^1 ya que f es continua. Entonces, por la ley fuerte de los grandes números tendremos que

$$\overline{Y}_n \overset{c.s.}{\to} \mathbb{E}(Y) = \mathbb{E}[(b-a)f(X_n)] = (b-a)\int_a^b f(x)\frac{1}{b-a}dx = \int_a^b f(x)dx.\checkmark$$

Corolario 8.16. ***Números normales.***
Dado un número $x \in (0, 1)$ podemos escribirlo en su expresión binaria como $x = \sum\limits_{n=1}^{+\infty} \frac{x_n}{2^n}$ donde $x_i \in \{0, 1\}$ para todo $i = 1, 2, 3, ...$ Si truncamos el número x a sus primeras n cifras en su expansión binaria (sumamos hasta n), observamos que $\overline{x}_n = \frac{x_1+x_2+...+x_n}{n}$ es el porcentaje de veces que aparece el 1, entre los primeros n términos. El número x se dice normal respecto a la base 2, si $\overline{x}_n \to 1/2$.
Probaremos que casi todo punto $\in (0, 1)$ es normal respecto a la base 2 (es decir que si se elige un número aleatorio en $(0, 1)$ con distribución uniforme, entonces el conjunto de números normales tiene probabilidad 1).

Demostración.
Dado $x \in (0, 1)$, escribimos $x = \sum\limits_{n=1}^{+\infty} \frac{x_n}{2^n}$ donde $x_i \in \{0, 1\}$ para todo $i = 1, 2, 3, ...$
Observamos que $x_n = 0$ en una unión de 2^{n-1} intervalos de longitud $(1/2)^n$ y $x_n = 1$ en la unión de los restantes 2^{n-1} intervalos de longitud $(1/2)^n$. Consideramos el siguiente espacio de probabilidad. $\Omega = (0, 1)$, sigma $= \mathbb{B}_{(0,1)}$ y P definida mediante la distribución uniforme.
Definimos la sucesión de variables aleatorias $X_n : (0, 1) \to \mathbb{R}$ tales que $X_n(x) = x_n$.
Entonces, la probabilidad de que X_n tome el valor 1 es la suma de las longitudes de los 2^{n-1} intervalos disjuntos de longitud $(1/2)^n$ lo que es igual a $1/2$. Esto prueba que $X_n \sim$Ber$(p = 1/2)$ para todo n. Además las variables son independientes ya que

$$P(X_{n_1} = \varepsilon_1, X_{n_2} = \varepsilon_2, ..., X_{n_k} = \varepsilon_k) = \frac{1}{2^k} = P(X_{n_1} = \varepsilon_1)P(X_{n_2} = \varepsilon_2)...P(X_{n_k} = \varepsilon_k)$$

cualesquiera sean k, $\varepsilon_1, \varepsilon_2, ..., \varepsilon_k \in \{0, 1\}$ y $n_1 < n_2 < ... < n_k$.
Hemos probado entonces que la sucesión $\{X_n\}_{n\in\mathbb{N}}$ son variables i.i.d con distribución Ber$(p = 1/2)$ por lo tanto, la ley fuerte de los grandes números nos asegura que $\overline{X}_n \overset{c.s.}{\to} p = 1/2$ lo cual significa que casi todo número real perteneciente al intervalo $(0, 1)$ es normal respecto a la base 2.$\checkmark$
De similar forma, se prueba que si se define número normal respecto a la la base

k, cuando el porcentaje de apariciones de cualquier $j \in \{0, 1, 2, ..., k-1\}$ converge a $1/k$, entonces casi todo número $\in (0, 1)$ es normal respecto a la base k. Por ejemplo, en el caso en que $k = 10$, tenemos que casi todo punto es normal respecto a su expansión decimal lo cual significa que el promedio de apariciones de los dígitos $0, 1, 2, ..., 9$ en su expansión decimal tiende a $1/10$.

8.3. Convergencia en distribución.

Apuntamos en lo que sigue a otro concepto de convergencia, de gran utilidad que es la convergencia en distribución. La idea, de la misma es que cuando n tienda a infinito, la función de distribución de las X_n converja a la función de distribución puntualmente en algún conjunto. En el siguiente ejemplo, veremos que la convergencia puntual de $F_n(x)$ a $F(x)$ es muy restrictiva si la pedimos para todo x.

Ejemplo 8.17. Si $\{c_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ es una sucesión decreciente tal que $c_n \to c$ y definimos para cada n las variables $X_n = c_n$ y $X = c$, desearíamos tener una definición de convergencia en distribución tal que X_n converja a X. Las funciones de distribución de estas variables son

$$F_{X_n}(x) = \begin{cases} 0 & \text{si} \quad x < c_n \\ 1 & \text{si} \quad x \geq c_n \end{cases} \quad \text{y } F_X(x) = \begin{cases} 0 & \text{si} \quad x < c \\ 1 & \text{si} \quad x \geq c \end{cases}.$$

Como se ve, $F_{X_n}(c) = 0$ no tiende a $F_X(c) = 1$, mientras que $F_{X_n}(x) \to F_X(x)$ para todo $x \neq c$.
Como se observa, c es el único punto de discontinuidad de F_X.

¿Cuántos puntos de discontinuidad puede tener una cierta función de distribución? Si $F : \mathbb{R} \to \mathbb{R}$ es una función de distribución, veremos que admite a lo sumo una cantidad numerable de discontinuidades.
Para demostrarlo, observamos que

$$\{x \in \mathbb{R} : F \text{ es discontinua en } x\} = \bigcup_{n=1}^{+\infty} \{x \in \mathbb{R} : F(x) - F(x^-) \geq 1/n\}$$

además, para cada n, el conjunto $\{x \in \mathbb{R} : F(x) - F(x^-) \geq 1/n\}$ tiene a lo sumo n elementos, puesto que dado que F es creciente y acotada entre 0 y 1, la suma de los saltos de distintos puntos de discontinuidad no puede exceder a 1. Por lo tanto el conjunto de puntos de discontinuidad de F es numerable por ser unión numerable de conjuntos finitos.
Se deja como ejercicio verificar que si $F : \mathbb{R} \to \mathbb{R}$ es monótona entonces el conjunto de sus puntos de discontinuidad es a lo sumo numerable.

Definición 8.18. Convergencia en distribución.
Dadas $\{X_n\}_{n\in\mathbb{N}}$ variables aleatorias definidas en $(\Omega_n, \mathcal{A}_n, P_n)$ espacios de probabilidad, y X variable aleatoria definida en cierto $(\Omega, \mathcal{A}, P)$ espacio de probabilidad. Se dice que la sucesión $\{X_n\}_{n\in\mathbb{N}}$ converge en distribución a X si y sólo si

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

Notación: $X_n \xrightarrow{d} X$.

También se dice que la sucesión $\{X_n\}_{n\in\mathbb{N}}$ converge débilmente a X, o también que F_{X_n} converge débilmente a F_X.

Observación 8.19. *Como se ve en la definición, no es necesario que las variables X_n y X esten todas definidas en el mismo espacio de probabilidad, ya que lo que importa, es que la convergencia se de entre sus funciones de distribución que son funciones de $\mathbb{R}$ en $\mathbb{R}$.*

Veremos en el siguiente teorema que cuando trabajamos sobre un mismo espacio de probabilidad, la noción de convergencia en distribución es aún más débil que la noción de convergencia en probabilidad.

Teorema 8.20. *Dadas una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ y una variable aleatoria X definidas sobre cierto $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad.*

$$\textit{Si } X_n \xrightarrow{P} X \textit{ entonces } X_n \xrightarrow{d} X.$$

Demostración.
Dado x punto de continuidad de F_X. Fijamos $\varepsilon > 0$ y le llamamos $A_{n,\varepsilon} = \{X - \varepsilon < X_n < X + \varepsilon\}$. Entonces

$$F_{X_n}(x) = P\left(X_n \leq x\right) = P\left(\{X_n \leq x\} \cap A_{n,\varepsilon}\right) + P\left(\{X_n \leq x\} \cap A^c_{n,\varepsilon}\right).$$

Con respecto al primer sumando, tenemos que

$$P\left(\{X_n \leq x\} \cap A_{n,\varepsilon}\right) \leq P\left(\{X - \varepsilon \leq x\} \cap A_{n,\varepsilon}\right) \leq P\left(X - \varepsilon \leq x\right) = F_X(x+\varepsilon).$$

Entonces tenemos que

$$F_{X_n}(x) \leq F_X(x+\varepsilon) + P\left(\{X_n \leq x\} \cap A^c_{n,\varepsilon}\right).$$

Tomando límite en n, el segundo sumando tiende a cero (ya que $P\left(A^c_{n,\varepsilon}\right)$ tiende a cero), por lo que obtenemos la desigualdad $F_{X_n}(x) \leq F_X(x+\varepsilon)$ válida para todo $\varepsilon > 0$. Luego, tomamos límite cuando $\varepsilon \to 0^+$ y usando que F_X es continua por derecha, nos queda

$$\limsup_{n\to+\infty} F_{X_n}(x) \leq F_X(x).$$

Para obtener una desigualdad en el otro sentido razonaremos en forma similar.

$$P\left(\{X_n \leq x\} \cap A_{n,\varepsilon}\right) \geq P\left(\{X \leq x - \varepsilon\} \cap A_{n,\varepsilon}\right).$$

Entonces

$$F_{X_n}(x) \geq P\left(\{X \leq x - \varepsilon\} \cap A_{n,\varepsilon}\right) + P\left(\{X_n \leq x\} \cap A^c_{n,\varepsilon}\right).$$

Si ahora tomamos límite en n, obtenemos que para todo $\varepsilon > 0$,

$$\liminf_{n\to+\infty} F_{X_n}(x) \geq F_X(x-\varepsilon).$$

Ahora usando que x es punto de continuidad de F_X,tomamos límite cuando $\varepsilon \to 0^+$ y obtenemos que $\liminf\limits_{n\to+\infty} F_{X_n}(x) \geq F_X(x)$. Hemos probado entonces que

$$\lim_{n\to+\infty} F_{X_n}(x) = F_X(x).\checkmark$$

Ahora veremos en el siguiente ejemplo que la convergencia en probabilidad es estrictamente más fuerte que la convergencia en distribución.

Ejemplo 8.21. Definimos una sucesión de variables $X, X_1, X_2, ..., X_n, ...$ i.i.d con distribución $N(0,1)$. Entonces $X_n \xrightarrow{d} X$ ya que $F_{X_n} = F_X$ para todo n. Sin embargo la sucesión $\{X_n\}_{n\in\mathbb{N}}$ no converge en probabilidad a X ya que $X_n - X$ tiene distribución $N(0,2)$ para todo n (ya que es combinación lineal de normales independientes), y por lo tanto

$$P(|X_n - X| \leq \varepsilon) = P(-\varepsilon \leq X_n - X \leq \varepsilon) = \Phi\left(\frac{\varepsilon}{\sqrt{2}}\right) - \Phi\left(-\frac{\varepsilon}{\sqrt{2}}\right)$$

esta probabilidad, no depende de n y es menor estricto que 1 por lo que no hay convergencia en probabilidad.

Capítulo 9

Funciones características.

En este capítulo definiremos un concepto que nos permitirá seguir desarrollando el concepto de convergencia en distribución, de hecho veremos más caracterizaciones para esta noción de convergencia, y finalizaremos con un teorema esencial en la teoría y práctica: el teorema central del límite.

Definición 9.1. Función característica. Dado un $\big(\Omega, \mathcal{A}, P\big)$ espacio de probabilidad y $X : \Omega \to \mathbb{R}$ variable aleatoria, se define la función característica de X como $\varphi_X : \mathbb{R} \to \mathbb{C}$ tal que $\varphi_X(t) = \mathbb{E}\left(e^{itX}\right)$.

Observación 9.2. *Dado que* $e^{itX} = \cos(tX) + isen(tX)$, *se tiene que*

$$\mathbb{E}\left(e^{itX}\right) = \int_{-\infty}^{+\infty} e^{itx} dF_X(x) = \int_{-\infty}^{+\infty} \left(\cos(tx) + isen(tx)\right) dF_X(x) =$$

$$\int_{-\infty}^{+\infty} \cos(tx) dF_X(x) + i \int_{-\infty}^{+\infty} sen(tx) dF_X(x) = \mathbb{E}\left(\cos(tX)\right) + i\mathbb{E}\left(sen(tX)\right).$$

Observación 9.3. *La función característica de* X *siempre existe ya que* $\left|e^{itX}\right| = 1$ *para todo* t.

Ejemplo 9.4. Si $X \sim$Poisson(λ), entonces

$$\varphi_X(t) = \mathbb{E}\left(e^{itX}\right) = \int_{-\infty}^{+\infty} e^{itx} dF_X(x) = \sum_{x=0}^{+\infty} e^{itx} p_X(x) = \sum_{x=0}^{+\infty} e^{itx} \frac{e^{-\lambda}\lambda^x}{x!} =$$

$$e^{-\lambda} \sum_{x=0}^{+\infty} \frac{\left(\lambda e^{it}\right)^x}{x!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda\left(e^{it}-1\right)}.$$

Como se verá más adelante, la función característica, juega un papel esencial en la teoría de la convergencia en distribución, convergencia clave en estadística.

9.1. Propiedades.

En todas las siguientes propiedades, se supone dado un espacio de probabilidad $\big(\Omega, \mathcal{A}, P\big)$ y en él, una variable aleatoria $X : \Omega \to \mathbb{R}$.

Proposición 9.5.

$$|\varphi_X(t)| \leq 1 \textit{ para todo } t \in \mathbb{R}.$$

Demostración.
$|\varphi_X(t)| = \left|\mathbb{E}\left(e^{itX}\right)\right| \leq \mathbb{E}\left(\left|e^{itX}\right|\right) = \mathbb{E}\left(1\right) = 1.\checkmark$

Proposición 9.6.

$$\varphi_X\left(0\right) = 1.$$

Demostración.
Obvia.✓

Proposición 9.7.

$$\varphi_{aX+b}(t) = e^{itb}\varphi_X(at) \textit{ para todo } t \in \mathbb{R}.\checkmark$$

Demostración.
$\varphi_{aX+b}(t) = \mathbb{E}\left(e^{it(aX+b)}\right) = \mathbb{E}\left(e^{itaX}e^{itb}\right) = e^{itb}\mathbb{E}\left(e^{iatX}\right) = e^{itb}\varphi_X(at).$

Proposición 9.8. *Si X e Y son independientes, entonces*

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y\left(t\right) \textit{ para todo } t \in \mathbb{R}.$$

Demostración.

$$\varphi_{X+Y}(t) = \mathbb{E}\left(e^{it(X+Y)}\right) = \mathbb{E}\left(e^{itX}e^{itY}\right) \overset{\text{indep}}{=} \mathbb{E}\left(e^{itX}\right)\mathbb{E}\left(e^{itY}\right) = \varphi_X(t)\varphi_Y\left(t\right).\checkmark$$

Proposición 9.9. *φ_X es uniformemente continua.*

Demostración.

$$\varphi_X(t) - \varphi_X(s) = \mathbb{E}\left(e^{itX}\right) - \mathbb{E}\left(e^{isX}\right) = \mathbb{E}\left(e^{itX} - e^{isX}\right) = \mathbb{E}\left(e^{isX}\left(e^{i(t-s)X} - 1\right)\right).$$

Si definimos $g(h) = \mathbb{E}\left(\left|e^{ihX} - 1\right|\right)$, entonces

$$|\varphi_X(t) - \varphi_X(s)| = \left|\mathbb{E}\left(e^{isX}\left(e^{i(t-s)X} - 1\right)\right)\right| \leq \mathbb{E}\left(\left|e^{isX}\left(e^{i(t-s)X} - 1\right)\right|\right) =$$

$$\mathbb{E}\left(\left|e^{i(t-s)X} - 1\right|\right) = g(t-s).$$

Por lo tanto, bastará con ver que g es continua en cero, es decir que $g(h)$ tiende a cero cuando $h \to 0$.
Observamos que $\left|e^{ihx} - 1\right| \leq 2 \in L^1$, y como $e^{ihX} - 1 \underset{h\to 0}{\to} 0$ c.s, entonces por el teorema de convergencia dominada, se tiene que $\lim\limits_{h\to 0}\mathbb{E}\left(e^{ihX} - 1\right) = 0.\checkmark$

Proposición 9.10. *Si $X \in L^k$ para cierto $k \in \mathbb{N}$, $k \geq 1$. Entonces $\varphi_X \in C^k$ y además*
$$\varphi_X^{(k)}(t) = i^k \mathbb{E}\left(X^k e^{itX}\right) \text{ para todo } t \in \mathbb{R}.$$

Demostración.
La prueba se realiza por inducción. Probémoslo para $k = 1$.

$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \frac{\mathbb{E}\left(e^{i(t+h)X}\right) - \mathbb{E}\left(e^{itX}\right)}{h} =$$
$$\frac{\mathbb{E}\left(e^{i(t+h)X} - e^{itX}\right)}{h} = \mathbb{E}\left(\frac{e^{itX}\left(e^{ihX} - 1\right)}{h}\right).$$

Ahora, observamos que $\lim\limits_{h\to 0} \frac{e^{ihx}-1}{h} = ix$, por lo tanto $\frac{e^{itX}\left(e^{ihX}-1\right)}{h} \overset{c.s.}{\to} iXe^{itX}$ cuando $h \to 0$.
Además, $\left|\frac{e^{itx}\left(e^{ihx}-1\right)}{h}\right| = \left|\frac{e^{ihx}-1}{h}\right| = \left|\int_0^x e^{ihs} ds\right| \leq \left|\int_0^x \left|e^{ihs}\right| ds\right| = |x|$ para todos $x, h \in \mathbb{R}$. Entonces $\left|\frac{e^{itX}\left(e^{ihX}-1\right)}{h}\right| \leq |X| \in L^1$, por lo tanto, usando el teorema de convergencia dominada se deduce que

$$\lim_{h\to 0} \frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \lim_{h\to 0} \mathbb{E}\left(\frac{e^{itX}\left(e^{ihX} - 1\right)}{h}\right) = i\mathbb{E}\left(Xe^{itX}\right).$$

Se deja como ejercicio demostrar el paso inductivo y así completar la demostración. ✓

Observación 9.11. *Si Si $X \in L^k$ para cierto $k \in \mathbb{N}$, la proposición anterior nos asegura que podemos derivar respecto a la variable t debajo del signo de la esperanza k veces.*

Observación 9.12. *Si $X \in L^k$ para cierto $k \in \mathbb{N}$, $k \geq 1$, entonces $\varphi_X^{(j)}(0) = i^j \mathbb{E}\left(X^j\right)$ para todo $j = 1, 2, 3, ..., k$. En particular si $X \in L^k$ para todo $k \in \mathbb{N}$, entonces $\varphi_X \in C^\infty$ y además quedan determinados todos los momentos de la variable X a partir de φ_X.*

Observación 9.13. *Se deduce de la demostración que en el caso en que $X \in L^k$ para cierto k, entonces φ_X es uniformemente continua.*

Ejemplo 9.14. Si $X \sim N\left(\mu, \sigma^2\right)$, entonces

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

Para demostrarlo, en primer lugar probaremos que si $X \sim N(0,1)$, probaremos que $\varphi_X(t) = e^{-t^2/2}$. Para lograrlo, demostraremos que si definimos la función h como $h(t) := e^{t^2/2}\varphi_X(t)$, entonces $h(t) = 1$ para todo t.
Como $h(0) = 1$, bastará probar que $h'(t) = 0$ para todo t. En efecto, dado que

podemos derivar debajo del signo de esperanza en la función característica, obtenemos $h'(t) = te^{t^2/2}\mathbb{E}\left(e^{itX}\right)+e^{t^2/2}\mathbb{E}\left(iXe^{itX}\right)$. Entonces, resta probar que $\mathbb{E}\left((t+iX)\,e^{itX}\right) = 0$.

$$\mathbb{E}\left((t+iX)\,e^{itX}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(t+ix)e^{itx}e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(t+ix)e^{itx-x^2/2}dx =$$

$$-ie^{itx-x^2/2}|_{-\infty}^{+\infty} = 0.$$

Ahora, para demostrar el caso en que $X \sim N(\mu, \sigma^2)$, escribimos $X = \sigma Z + \mu$ donde $Z \sim N(0,1)$. Entonces,

$$\varphi_X(t) = \varphi_{\sigma Z+\mu}(t) = e^{it\mu}\varphi_Z(\sigma t) = e^{it\mu - t^2\sigma^2/2}.$$

9.2. Fórmula de inversión.

En esta sección probaremos una fórmula que nos pemite obtener F_X si conocemos φ_X, de aquí se deducirá que la función característica de una variable aleatoria, "caracteriza" a la función de distribución, es decir que $F_X = F_Y$ si y sólo si $\varphi_X = \varphi_Y$.

Teorema 9.15. ***Fórmula de inversión.***

Dado un $(\Omega, \mathcal{A}, P)$ espacio de probabilidad y $X : \Omega \to \mathbb{R}$ variable aleatoria, entonces

$$F_X(x) = \lim_{z\downarrow x}\lim_{y\to-\infty}\lim_{h\to+\infty}\frac{1}{2\pi}\int_{-h}^{h}\frac{e^{-ity}-e^{-itz}}{it}\varphi_X(t)dt \text{ para todo } x,$$

donde los límites en y y en z se realizan sobre puntos de continuidad de F_X.
Demostración.
En primer lugar fijamos $y < z$ puntos de continuidad de F_X.
Definimos

$$I(h) := \int_{-h}^{h}\frac{e^{-ity}-e^{-itz}}{it}\varphi_X(t)dt = \int_{-h}^{h}\left(\int_{-\infty}^{+\infty}\frac{e^{-ity}-e^{-itz}}{it}e^{itx}dF_X(x)\right)dt.$$

Dado que la función integrando $f(t,z) = \frac{e^{-ity}-e^{-itz}}{it}e^{itx}$ es continua, ya que $\lim_{t\to 0}\frac{e^{-ity}-e^{-itz}}{it} = y - z$, por lo tanto $|f(t,x)| \leq c$ para todo $(t,x) \in [-h,h]\times\mathbb{R}$ y entonces $\int_{-h}^{h}\left(\int_{-\infty}^{+\infty}|f(t,x)|\,dF_X(x)\right)dt \leq 2hc$, por lo que podemos intercambiar el orden de integración (Fubini), obteniendo que $I(h) = \int_{-\infty}^{+\infty}\left(\int_{-h}^{h}\frac{e^{it(x-y)}-e^{it(x-z)}}{it}dt\right)dF_X(x)$.
Ahora, observando que $\frac{\cos(at)}{t}$ es impar y $\frac{sen(at)}{t}$ es par para todo $a \in \mathbb{R}$, nos queda que

$$I(h) = \int_{-\infty}^{+\infty}\left(2\int_{0}^{h}\frac{sent(x-y)}{t}dt - 2\int_{0}^{h}\frac{sent(x-z)}{t}dt\right)dF_X(x) = \mathbb{E}\left(g_h(X)\right)$$

siendo $g_h(x) = 2\int_0^h \frac{sent(x-y)}{t}dt - 2\int_0^h \frac{sent(x-z)}{t}dt$.
Tomaremos límite cuando $h \to +\infty$ y veremos que podemos aplicar el teorema de convergencia dominada.

Utilizando el valor de la integral de Dirichlet $\int_0^{+\infty} \frac{sen(at)}{t}dt = \begin{cases} \pi/2 & \text{si } a > 0 \\ 0 & \text{si } a = 0 \\ -\pi/2 & \text{si } a < 0 \end{cases}$,
entonces el límite puntual de g_h es

$$\lim_{h\to+\infty} g_h(x) = 2\pi \mathbf{1}_{\{y<x<z\}} + \pi \mathbf{1}_{\{x=y\}} + \pi \mathbf{1}_{\{x=z\}}.$$

Observando que $\left|\int_0^h \frac{sen(at)}{t}dt\right| \leq \sup_{h>0} \left|\int_0^h \frac{sent}{t}dt\right| \overset{\text{def}}{:=} M$, entonces

$$|g_h(x)| = \left|2\int_0^h \frac{sent(x-y)}{t}dt - 2\int_0^h \frac{sent(x-z)}{t}dt\right| \leq 4M$$

entonces por el teorema de convergencia dominada se obtiene que

$$\lim_{h\to+\infty} I(h) = \lim_{h\to+\infty} \mathbb{E}\left(g_h\left(X\right)\right) = \mathbb{E}\left(2\pi \mathbf{1}_{\{y<X<z\}} + \pi \mathbf{1}_{\{X=y\}} + \pi \mathbf{1}_{\{X=z\}}\right)$$

y como y, z son puntos de continuidad de F_X entonces

$$\lim_{h\to+\infty} I(h) = 2\pi P\left(y < X < z\right) = 2\pi\left(F_X(z) - F_X(y)\right).$$

Entonces

$$F_X(z) - F_X(y) = \frac{1}{2\pi}\lim_{h\to+\infty} I(h) = \frac{1}{2\pi}\lim_{h\to+\infty}\int_{-h}^h \frac{e^{-ity} - e^{-itz}}{it}\varphi_X(t)dt.$$

Si tomamos límite cuando $y \to -\infty$ (siendo y punto de continuidad de F_X) en la anterior igualdad, obtenemos

$$F_X(z) = \frac{1}{2\pi}\lim_{y\to-\infty}\lim_{h\to+\infty}\int_{-h}^h \frac{e^{-ity} - e^{-itz}}{it}\varphi_X(t)dt \text{ para todo } z \text{ punto de continuidad de } F_X.$$

Para concluir, basta fijar cualquier $x \in \mathbb{R}$ y tomar límite en la anterior igualdad cuando $z \to x^+$ tomando z puntos de continuidad de F_X (esto es posible debido a que por ser F_X una función monótona, la cantidad de puntos de discontinuidad es numerable).
Entonces nos queda

$$F_X(x) = \frac{1}{2\pi}\lim_{z\to x^+}\lim_{y\to-\infty}\lim_{h\to+\infty}\int_{-h}^h \frac{e^{-ity} - e^{-itz}}{it}\varphi_X(t)dt \text{ para todo } x \in \mathbb{R},$$

donde el límite en las variables y, z se hacen sobre puntos de continuidad de F_X.✓

Corolario 9.16. *Dado un* $\left(\Omega, \mathcal{A}, P\right)$ *espacio de probabilidad y* $X, Y : \Omega \to \mathbb{R}$ *variables aleatorias. Entonces*

$$F_X = F_Y \text{ si y sólo si } \varphi_X = \varphi_Y.$$

Demostración.
Es consecuencia inmediata de la fórmula de inversión.✓

9.3. Caracterización de la convergencia en distribución.

En el siguiente teorema, probaremos que la convergencia en distribución es equivalente a la convergencia puntual de las funciones características.

Teorema 9.17. *Si para cada n, $X_n : \Omega \to \mathbb{R}$ es variable aleatoria sobre $(\Omega_n, \mathcal{A}_n, P_n)$ y $X : \Omega \to \mathbb{R}$ es variable aleatoria sobre $(\Omega, \mathcal{A}, P)$. Entonces son equivalentes:*

(a) $X_n \xrightarrow{d} X$.

(b) $\mathbb{E}\left(g\left(X_n\right)\right) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left(g\left(X\right)\right)$ *para toda $g : \mathbb{R} \to \mathbb{R}$ continua y acotada.*

(c) $\varphi_{X_n}(t) \underset{n\to+\infty}{\longrightarrow} \varphi_X(t)$ *para todo $t \in \mathbb{R}$.*

Demostración.
(a) $\Rightarrow$ (b)
Para simplificar la escritura, le llamamos F_n a la función de distribución de las X_n y F a la función de distribución de X. Tomemos $g : \mathbb{R} \to \mathbb{R}$ continua y acotada, tal que $|g(x)| \leq c$ para todo $x \in \mathbb{R}$, entonces para cualesquiera $a < b$, tenemos

$$\left|\mathbb{E}\left(g\left(X_n\right)\right) - \mathbb{E}\left(g\left(X\right)\right)\right| = \left|\int_{-\infty}^{+\infty} gdF_n - \int_{-\infty}^{+\infty} gdF\right| \leq$$

$$\left|\int_{-\infty}^{+\infty} gdF_n - \int_a^b gdF_n\right| + \left|\int_a^b gdF_n - \int_a^b gdF\right| + \left|\int_a^b gdF - \int_{-\infty}^{+\infty} gdF\right| := I_1 + I_2 + I_3.$$

Fijemos un $\varepsilon > 0$ arbitrario.

$$I_3 = \left|\int_{-\infty}^{a} gdF + \int_b^{+\infty} gdF\right| \leq \left|\int_{-\infty}^{a} gdF\right| + \left|\int_b^{+\infty} gdF\right| \leq \int_{-\infty}^{a} |g|\, dF + \int_b^{+\infty} |g|\, dF \leq$$

$$\int_{-\infty}^{a} cdF + \int_b^{+\infty} cdF = c\left(F\left(a\right) + 1 - F\left(b\right)\right).$$

Dado que $c\left(F\left(a\right) + 1 - F\left(b\right)\right) \to 0$ cuando $a \to -\infty$ y $b \to +\infty$, elegimos a suficientemente pequeño y b suficientemente grande tal que $c\left(F\left(a\right) + 1 - F\left(b\right)\right) < \varepsilon$. Por conveniencia tomaremos a, b puntos de continuidad, ya que lo necesitaremos para acotar I_1 e I_2.
Acotamos de manera similar I_1 y obtenemos

$$I_1 = \left|\int_{-\infty}^{+\infty} gdF_n - \int_a^b gdF_n\right| \leq c\left(F_n\left(a\right) + 1 - F_n\left(b\right)\right).$$

Para los a y b obtenidos, dado que son puntos de continuidad de F , se deduce que $c\left(F_n\left(a\right) + 1 - F_n\left(b\right)\right) \underset{n\to+\infty}{\longrightarrow} c\left(F\left(a\right) + 1 - F\left(b\right)\right) < \varepsilon$, por lo tanto existe $k \in \mathbb{N}$ tal que $c\left(F_n\left(a\right) + 1 - F_n\left(b\right)\right) < 2\varepsilon$ para todo $n \geq k$. Por ahora obtenemos $I_1 + I_3 < 3\varepsilon$ para todo $n \geq k$.

Para culminar la demostración, probaremos que $I_2 < 3\varepsilon$ para todo n suficientemente grande.
Como g es continua en $[a,b]$, entonces es absolutamente continua, por lo que podemos elegir una partición de $[a,b]$, $a = x_0 < x_1 < x_2 < ... < x_N = b$ tal que $x_1, x_2, ..., x_{N-1}$ sean puntos de continuidad de F_X y $|g(x) - g(x_i)| < \varepsilon$ para todo $x \in [x_i, x_{i+1}]$ para todo $i = 0, 1, 2, ..., N-1$.

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

$$m_{ni} \stackrel{\text{def}}{=} (g(x_i) - \varepsilon) \left(F_n\left(x_{i+1}\right) - F_n(x_i) \right) \leq \int_{x_i}^{x_{i+1}} g(x) dF_n(x) \leq$$

$$(g(x_i) + \varepsilon) \left(F_n\left(x_{i+1}\right) - F_n(x_i) \right) \stackrel{\text{def}}{=} M_{ni}$$

$$m_i \stackrel{\text{def}}{=} (g(x_i) - \varepsilon) \left(F\left(x_{i+1}\right) - F(x_i) \right) \leq \int_{x_i}^{x_{i+1}} g(x) dF(x) \leq$$

$$(g(x_i) + \varepsilon) \left(F\left(x_{i+1}\right) - F(x_i) \right) \stackrel{\text{def}}{=} M_i.$$

Entonces

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

y sumando en todos los intervalos, obtenemos que

$$\sum_{i=0}^{N-1} (m_{ni} - M_i) \leq \int_a^b g(x) dF_n(x) - \int_a^b g(x) dF(x) \leq \sum_{i=0}^{N-1} (M_{ni} - m_i).$$

Ahora, observamos que como los x_i son puntos de continuidad de F_X, se obtiene que $m_{ni} \underset{n\to+\infty}{\to} m_i$ y $M_{ni} \underset{n\to+\infty}{\to} M_i$ para todo $i = 0, 1, 2, ..., N-1$, por lo que

$$\sum_{i=0}^{N-1} (m_{ni} - M_i) \underset{n\to+\infty}{\to} \sum_{i=0}^{N-1} (m_i - M_i) =$$

$$-2\varepsilon \sum_{i=0}^{N-1} \left(F\left(x_{i+1}\right) - F(x_i) \right) = -2\varepsilon \left(F(b) - F(a) \right) \geq -2\varepsilon$$

y

$$\sum_{i=0}^{N-1} (M_{ni} - m_i) \underset{n\to+\infty}{\to} \sum_{i=0}^{N-1} (M_i - m_i) =$$

$$2\varepsilon \sum_{i=0}^{N-1} \left(F\left(x_{i+1}\right) - F(x_i) \right) = 2\varepsilon \left(F(b) - F(a) \right) \leq 2\varepsilon.$$

Entonces a partir de cierto n suficientemente grande, se tiene que

$$-3\varepsilon \leq \int_a^b g(x)dF_n(x) - \int_a^b g(x)dF(x) \leq 3\varepsilon$$

lo que prueba que $I_2 \leq 3\varepsilon$ concluyendo así la prueba.
(b) ⇒ (c)
Fijado $t \in \mathbb{R}$, consideramos las funciones $g_1(x) = sen\,(tx)$ y $g_2(x) = \cos(tx)$ ambas son continuas y acotadas, por lo que $\mathbb{E}\left(g_1\left(X_n\right)\right) = \mathbb{E}\left(sen(tX_n)\right) \underset{n\to+\infty}{\rightarrow} \mathbb{E}\left(g_1\left(X\right)\right) = \mathbb{E}\left(sen(tX)\right)$, y $\mathbb{E}\left(g_2\left(X_n\right)\right) = \mathbb{E}\left(\cos(tX_n)\right) \underset{n\to+\infty}{\rightarrow} \mathbb{E}\left(g_2\left(X\right)\right) = \mathbb{E}\left(\cos(tX)\right)$. Entonces $\mathbb{E}\left(e^{itX_n}\right) \underset{n\to+\infty}{\rightarrow} \mathbb{E}\left(e^{itX}\right)$ y como t es arbitrario, entonces $\varphi_{X_n}(t) \underset{n\to+\infty}{\rightarrow} \varphi_X(t)$ para todo $t \in \mathbb{R}$.
(c) ⇒ (a)
Nuevamente, por simplicidad, le llamamos F_n a la función de distribución de X_n y F a la función de distribución de X. Para demostrar que $F_n \xrightarrow{d} F$, bastará con probar que existe una subsucesión tal que $F_{n_j} \xrightarrow{d} F$. Esto se debe a que una vez probado que $F_{n_j} \xrightarrow{d} F$, si $\{F_n\}_{n\in\mathbb{N}}$ no convergiera débilmente a F, entonces, existiría x_0 punto de continuidad de F tal que $F_n(x_0) \nrightarrow F(x_0)$, entonces como $\{F_n(x_0)\}_{n\in\mathbb{N}}$ es una sucesión acotada, existe una subsucesión $\{F_{n_k}\}_{k\in\mathbb{N}}$ tal que $F_{n_k}(x_0) \underset{k\to+\infty}{\rightarrow} a$ para cierto $a \neq F(x_0)$. Entonces extraemos una subsucesión de $\{F_{n_k}\}_{k\in\mathbb{N}}$, que converge débilmente a F, $F_{n_{k_j}} \xrightarrow{d} F$. Entonces, dado que x_0 es punto de continuidad de F, se tendría que $F_{n_{k_j}}(x_0) \underset{j\to+\infty}{\rightarrow} F(x_0)$, pero $\left\{F_{n_{k_j}}(x_0)\right\}_{j\in\mathbb{N}}$ es subsucesión de $\{F_{n_k}(x_0)\}_{k\in\mathbb{N}}$ y por lo tanto $F_{n_{k_j}}(x_0) \underset{j\to+\infty}{\rightarrow} a \neq F(x_0)$, lo cual es absurdo.

En lo que sigue, construiremos una subsucesión $\left\{F_{n_j}\right\}_{j\in\mathbb{N}}$ de $\{F_n\}_{n\in\mathbb{N}}$ tal que $F_{n_j} \xrightarrow{d} F$. Consideramos una numeración de los racionales, $\mathbb{Q} = \{q_k\}_{k\in\mathbb{N}}$. Para cada k, existe una subsucesión de $\{F_n(q_k)\}_{n\in\mathbb{N}}$ que es convergente, llamémosle g_k a dicho límite. Mediante el procedimiento de la diagonal, podemos asegurar que existe una sucesión de naturales $n_1 < n_2 < \ldots < n_j < \ldots$ tal que $F_{n_j}(q_k) \underset{j\to+\infty}{\rightarrow} g(q_k)$ para todo k. Definimos la función $G : \mathbb{R} \to \mathbb{R}$ tal que $G(x) = \begin{cases} g(q_k) & \text{si} \quad x = q_k \\ \lim\limits_{q\downarrow x\ q\in\mathbb{Q}} g(q) & \text{si} \quad x \notin \mathbb{Q} \end{cases}$. En primer lugar debemos ver que G está bien definida, es decir que existe el límite para el caso en que x es irracional. Para ello, observamos que G restringida a $\mathbb{Q}$, es monótona creciente, esto se debe a que si $q < q'$ entonces $F_{n_j}(q) \leq F_{n_j}(q')$ para todo j, luego, se toma límite en j. De aquí se deduce que G es monótona creciente. Podría no ser continua por derecha, pero veamos en lo que sigue, que $F_{n_j}(x) \underset{j\to+\infty}{\rightarrow} G(x)$ en todo punto de continuidad de G.
En efecto, si x es punto de continuidad de G, entonces, dado $\varepsilon > 0$, existen dos racionales q y q' tales que $q < x < q'$ con $G(q') - \varepsilon < G(x) < G(q) + \varepsilon$, entonces

$$G(x) - \varepsilon < G(q) = \lim_{j\to+\infty} F_{n_j}(q) \leq \liminf F_{n_j}(x) \leq$$

$$\limsup F_{n_j}(x) \leq \lim_{j\to+\infty} F_{n_j}(q') = G(q') < G(x) + \varepsilon$$

de donde se deduce que $\lim_{j\to+\infty} F_{n_j}(x) = G(x)$. En los puntos donde G no sea continua, la podemos redefinir de modo que quede continua por derecha (esto es posible porque G es creciente).
Probaremos que ésta función G redefinida de modo que quede continua por derecha, es una función de distribución, para lo cual bastará ver que tiene límites 0 y 1 a $-\infty$ y $+\infty$ respectivamente.
Como $\varphi_{X_{n_j}} \to \varphi_X$ en todo punto, entonces, por el teorema de convergencia dominada dado que $\left|\varphi_{X_{n_j}}(s)\right| \leq 1$ para todo s, obtenemos

$$\int_0^t \varphi_{X_{n_j}}(s)ds \underset{j\to+\infty}{\to} \int_0^t \varphi_X(s)ds \text{ para todo } t.$$

Por otro lado, observamos que

$$\int_0^t \varphi_X(s)ds = \int_0^t \left(\int_{-\infty}^{+\infty} e^{isu}dF\left(u\right)\right) ds \overset{\text{Fubini}}{=} \int_{-\infty}^{+\infty} \left(\int_0^t e^{isu}ds\right) dF(u) =$$

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

Además, observando que la demostración de que $(a) \Rightarrow (b)$ sigue valiendo si la convergencia débil, es definida sobre funciones acotadas, si definimos $g_t : \mathbb{R} \to \mathbb{R}$ tal que $g_t(u) = \frac{e^{iut}-1}{iu}$, entonces, dado que para todo t, g_t es continua y acotada, se tiene que $\mathbb{E}\left(g_t(X_{n_j})\right) \underset{j\to+\infty}{\to} \mathbb{E}\left(g_t(X)\right)$, es decir

$$\int_{-\infty}^{+\infty} \left(\frac{e^{iut}-1}{iu}\right) dF_{n_j}(u) \underset{j\to+\infty}{\to} \int_{-\infty}^{+\infty} \left(\frac{e^{iut}-1}{iu}\right) dG(u) \text{ para todo } t.$$

Entonces obtuvimos

$$\int_0^t \varphi_X(s)ds = \int_0^t \left(\int_{-\infty}^{+\infty} e^{isu}dF\left(u\right)\right) ds = \int_0^t \left(\int_{-\infty}^{+\infty} e^{isu}dG\left(u\right)\right) ds$$

para todo t. Luego

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

y tomando límite cuando $t \to 0$ se obtiene que $1 = \varphi\left(0\right) = \int_{-\infty}^{+\infty} dG(u) = G\left(+\infty\right) - G\left(-\infty\right)$ y como además G es creciente y acotada entre 0 y 1, entonces necesariamente $G\left(+\infty\right) = 1$ y $G\left(-\infty\right) = 0$. Se concluye entonces que G es una función de distribución.
Ahora, como tenemos que $F_{n_j} \xrightarrow{d} G$, sabemos que existe un espacio de probabilidad y en él una variable aleatoria Y tal que $G = F_Y$. Como (a) implica (c), se deduce que

$\varphi_{X_n}(t) \underset{n\to+\infty}{\rightarrow} \varphi_Y(t)$ para todo t, pero por hipótesis $\varphi_{X_n}(t) \underset{n\to+\infty}{\rightarrow} \varphi_X(t)$ para todo t, por lo tanto $\varphi_X = \varphi_Y$, lo cual implica que $F_X = F_Y$, es decir $F = G$.
Queda probado hasta ahora que existe una subsucesión de $\{F_n\}_{n\in\mathbb{N}}$ tal que $F_{n_j} \xrightarrow{d} F$. Para concluir la prueba debemos ver que $F_n \xrightarrow{d} F$. Ahora, si $\{F_n\}_{n\in\mathbb{N}}$ no convergiera en distribución a F, entonces existiría $a \in \mathbb{R}$ punto de continuidad de F y una subsucesión $\left\{F_{n_j}\right\}_{j\in\mathbb{N}}$ tal que $F_{n_j}(a) \underset{j\to+\infty}{\nrightarrow} F(a)$. Podemos suponer que $\left\{F_{n_j}(a)\right\}_{j\in\mathbb{N}}$ es convergente ya que de lo contrario como es una sucesión acotada en $\mathbb{R}$, admiitiría una subsucesión convergente y trabajaríamos con dicha subsucesión si fuera necesario. Suponemos entonces que $\lim_{j\to+\infty} F_{n_j}(a) = b \neq F(a)$. Por lo recién probado, existe una subsucesión de $\left\{F_{n_j}\right\}_{j\in\mathbb{N}}$ que converge en distribución a cierta función de distribución G. Observamos además que debe ser $G = F$ ya que por hipótesis, las funciones características asociadas a esta subsucesión convergen a la función característica asociada a F.
Entonces como a es punto de continuidad de F, esta subsucesión evaluada en a, debería converger a $F(a)$, pero por ser subsucesión de $\left\{F_{n_j}(a)\right\}_{j\in\mathbb{N}}$ converge a b. ✓

9.4. Teorema Central del Límite.

El teorema central del límite es un equivalente en importancia a la ley de los grandes números en lo que respecta al límite en distribución de la sucesión $\overline{X}_n$.

Teorema 9.18. *Si $\{X_n\}_{n\geq 1}$ es una sucesión de v.a.i.i.d con distribución F_X, $X \in L^2$, $E(X) = \mu$, $V(X) = \sigma^2$. Entonces*

$$\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \xrightarrow{d} N(0,1).$$

Demostración.
Suponemos en un primer caso que $\mu = 0$ y $\sigma = 1$.
Recordando que la función característica de $N(0,1)$ es $\varphi(t) = e^{-t^2/2}$ para todo $t \in \mathbb{R}$, y usando el teorema que caracteriza la convergencia en distribución mediante la convergencia de las funciones características para todo t, bastará probar que

$$\varphi_{\sqrt{n}\overline{X}_n}(t) \underset{n\to+\infty}{\rightarrow} e^{-t^2/2} \quad \forall t \in \mathbb{R}.$$

Usando que $\varphi_{aX}(t) = \varphi_X(at)$ y luego que las X_i son independientes e idénticamente distribuídas, se obtiene

$$\varphi_{\sqrt{n}\overline{X}_n}(t) = \varphi_{\frac{X_1+X_2+\ldots+X_n}{\sqrt{n}}}(t) = \varphi_{X_1+X_2+\ldots+X_n}\left(t/\sqrt{n}\right) = \prod_{i=1}^{n} \varphi_{X_i}\left(t/\sqrt{n}\right) = \left[\varphi_X\left(t/\sqrt{n}\right)\right]^n.$$

Ahora si tenemos en cuenta que φ admite dos derivadas continuas (ya que $X \in L^2$) desarrollamos por Taylor alrededor de cero y obtenemos

$$\varphi_X(t) = \varphi_X(0) + \varphi_X'(0)\,t + \frac{\varphi_X''(c_t)\,t^2}{2} \text{ donde } |c_t| \leq |t|$$

Pero $\varphi_X(0) = 1$, $\varphi'_X(0) = iE(X) = 0$, $\varphi''_X(0) = -E(X^2) = -1$, entonces queda

$$\varphi_{\sqrt{n}\overline{X}n}(t) = \left[\varphi_X\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left[1 + \frac{\varphi''_X(c_{t,n})}{2n}t^2\right]^n = e^{n\ln\left(1+\frac{\varphi''_X(c_{t,n})}{2n}t^2\right)}.$$

Ahora, teniendo en cuenta que φ''_X es continua y que $|c_{t,n}| \leq |t|/\sqrt{n}$, se deduce que $\varphi''_X(c_{t,n}) \underset{n\to+\infty}{\to} \varphi''_X(0) = -1$.
Entonces

$$\lim_{n\to+\infty}\varphi_{\sqrt{n}\overline{X}n}(t) = \lim_{n\to+\infty} e^{n\ln\left(1+\frac{\varphi''_X(c_{t,n})}{2n}t^2\right)} = \lim_{n\to+\infty} e^{n\frac{\varphi''_X(c_{t,n})}{2n}t^2} = e^{-t^2/2}.$$

lo que concluye la prueba en el caso $\mu = 0$ y $\sigma = 1$.
El caso general se deduce definiendo las variables $Y_n := \frac{X_n-\mu}{\sigma}$. Entonces $\{Y_n\}_{n\geq 1}$ es una sucesión de v.a.i.i.d con distribución F_Y, $Y \in L^2$, $E(Y) = 0$, $V(Y) = 1$. Entonces se tiene que $\sqrt{n}\overline{Y_n} = \sqrt{n}\ \frac{\overline{X}_n-\mu}{\sigma} \xrightarrow{d} N(0,1)$ lo que concluye la prueba. ✓

Observación 9.19. *Si $X_1, X_2, \ldots$ son variables i.i.d en L^2 con esperanza μ y varianza σ^2, el teorema central del límite nos dice que*

$$\mathit{lim}_{n\to+\infty} P\left(\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - \mu\right) \leq x\right) = \Phi(x).$$

Entonces, si n es suficientemente grande, podemos realizar la siguiente aproximación

$$P\left(\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - \mu\right) \leq x\right) = P\left(\overline{X}_n \leq \mu + \frac{\sigma}{\sqrt{n}}x\right) \overset{aprox}{\simeq} \Phi(x)$$

luego, si le llamamos $t = \mu + \frac{\sigma}{\sqrt{n}}x$, entonces $F_{\overline{X}_n}(t) \simeq \Phi\left(\frac{\sqrt{n}}{\sigma}(t-\mu)\right)$ que es la función de distribución de una variable $N(\mu, \sigma^2/n)$, por lo tanto si n es suficientemente grande, entonces podemos aproximar la distribución de $\overline{X}_n$ por $N(\mu, \sigma^2/n)$.

Observación 9.20. *A partir de la observación anterior deducimos que, si n es suficientemente grande, podemos aproximar la distribución de $X_1 + X_2 + \ldots + X_n$ por $N(n\mu, n\sigma^2)$.*

Ejemplo 9.21. Si $X \sim$Bin(n,p) y n es suficientemente grande, entonces X es aproximadamente $N(np, np(1-p))$ ya que podemos escribir X como $X = X_1+X_2+\ldots+X_n$ donde $X_1, X_2, \ldots, X_n$ son i.i.d $\sim$Ber(p).

Ejemplo 9.22. Si tiramos 100 veces una moneda, calcularemos de manera aproximada mediante le empleo del teormea central del límite la probabilidad de obtener entre 40 y 60 caras.
Para el cálculo, definimos $X =$ çantidad de caras en los 100 lanzamientos", entonces $X \sim$Bin$(n = 100, p = 1/2)$. Deseamos hallar $P(40 \leq X \leq 60)$. Dado que $np = 50$ y $np(1-p) = 25$, tenemos que la distribución de X es aproximadamente $N(50; 25)$ y por lo tanto $P(40 \leq X \leq 60) \cong \Phi\left(\frac{60-50}{5}\right) - \Phi\left(\frac{40-50}{5}\right) = 0,954\,50$. El valor exacto en este caso es 0,9648.

Como aplicación, podemos volver a calcular n, de forma aproximada, tal que $P\left(\left|\overline{X}_n - p\right| \leq 0,01\right) \geq 0,95$ para el caso en que $X_1, X_2, ..., X_n$ son i.i.d $\sim$Ber(p). Esto ya fue resuelto como aplicación de la desigualdad de Chebyshev, ahora podremos dar otra solución, aproximada, mediante el empleo del teorema central del límite. Aproximando la distribución de $\overline{X}_n$ por $N\left(p, p(1-p)/n\right)$ obtenemos

$$P\left(\left|\overline{X}_n - p\right| \leq 0,01\right) = P\left(p - 0,01 \leq \overline{X}_n \leq p + 0,01\right) \cong$$

$$\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(\frac{-0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) = 2\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1$$

y usando que $p(1-p) \leq 1/4$ obtenemos

$$2\Phi\left(\frac{0,01\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \geq 2\Phi\left(0,02\sqrt{n}\right) - 1$$

por lo que bastará con hallar n tal que $2\Phi\left(0,02\sqrt{n}\right) - 1 \geq 0,95$ lo cual se cumple si y sólo si $0,02\sqrt{n} \geq \Phi^{-1}\left(0,975\right) = 1,96$, es decir que basta con tomar $n \geq \left(\frac{1,96}{0,02}\right)^2 = 9604$.

Observación 9.23. *El hecho de que aplicando el teorema central del límite, resulte un valor de n (aunque aproximado) notoriamente más pequeño que el obtenido por aplicación de la desigualdad de Chebyshev, se debe a que como ya fue dicho en su momento, la desigualdad de Chebysehv es una desigualdad universal, aplicable a toda variable aleatoria en L^2 y por lo tanto es natural esperar que en ciertas situaciones nos de acotaciones groseras de la probabilidad buscada.*

Capítulo 10

Estimación puntual.

10.1. Estadísticos y estimadores.

Cuando $X_1, X_2, ..., X_n$ son variables i.i.d con distribución como la de cierta X, se dice que $X_1, X_2, ..., X_n$ es una M.A.S (muestra aleatoria simple) de tamaño n de X.
En estadística aplicada, es frecuente encontrarse con números $x_1, x_2, ..., x_n$ producto de un muestreo sobre alguna característica de cierta población, por ejemplo, ingreso de los hogares de cierta ciudad, diámetro de las células de cierta población observada al microscopio, altura o peso de ciertos animales, etc. En todas estas situaciones, la variable a estudiar, no se conoce su distribución, por lo que interesa manipular la información que nos brinda la muestra $x_1, x_2, ..., x_n$ para poder estimar diversos parámetros de interés.

Definición 10.1. Si $X_1, X_2, ..., X_n$ es una M.A.S de cierta X para un determinado n, se le llama estadístico a la función $T(X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^k$ para cierto k, donde $T : \mathbb{R}^n \to \mathbb{R}^k$ es una función boreliana que no depende de parámetros desconocidos.

Se pide que la función T sea boreliana para que $T(X_1, X_2, ..., X_n)$ sea variable aleatoria, y se pide que no dependa de parámetros desconocidos porque dada una muestra realizada (u observada) $x_1, x_2, ..., x_n$, el valor $T(x_1, x_2, ..., x_n)$ pueda ser utilizado para estimar parámetros desconocidos por ejemplo.

Definición 10.2. Si $X_1, X_2, ..., X_n$ es una M.A.S de cierta X con distribución $F_X(x, \theta)$ con $\theta \in \Theta \subset \mathbb{R}^k$. Al conjunto Θ se le denomina espacio paramétrico.

Cuando tenemos $X_1, X_2, ..., X_n$ una M.A.S de cierta X con distribución $F_X(x, \theta)$ con $\theta \in \Theta \subset \mathbb{R}^k$, es decir que la distribución de la variable de estudio (X) es completamente conocida salvo por un parámetro θ, se dice que estamos en estadística paramétrica, mientras que si la distribución de X es totalmente desconocida, estamos en presencia de estadística no paramétrica.

Definición 10.3. Estimador.
Si $X_1, X_2, ..., X_n$ es una M.A.S de cierta X con distribución $F_X(x,\theta)$ con $\theta \in \Theta \subset \mathbb{R}^k$, se dice que $\widehat{\theta} : \Omega \to \Theta$ es un estimador de θ si y sólo si $\widehat{\theta}(X_1, X_2, ..., X_n)$ es un estadístico que es usado para "estimar" el verdadero valor de θ.

En general para abreviar, le llamaremos $\widehat{\theta}$ a $\widehat{\theta}(X_1, X_2, ..., X_n)$.
Observamos que $\widehat{\theta}$ depende de n y es importante tener un estimador que cumpla propiedades de convergencia al verdadero valor de θ cuando el tamaño de muestra $n \to +\infty$.

Definición 10.4. Estimador consistente.
Si $X_1, X_2, ..., X_n$ es una M.A.S de cierta X con distribución $F_X(x,\theta)$ con $\theta \in \Theta \subset \mathbb{R}^k$, se dice que $\widehat{\theta} = \widehat{\theta}(X_1, X_2, ..., X_n)$ es un estimador débilmente consistente si y sólo si $\widehat{\theta} \xrightarrow{P} \theta$ y se dice que es fuertemente consistente si y sólo si $\widehat{\theta} \xrightarrow{c.s.} \theta$.

Definición 10.5. Estimador insesgado.
Si $X_1, X_2, ..., X_n$ es una M.A.S de cierta X con distribución $F_X(x,\theta)$ con $\theta \in \Theta \subset \mathbb{R}^k$, se dice que $\widehat{\theta} = \widehat{\theta}(X_1, X_2, ..., X_n)$ es un estimador insesgado si y sólo si $\mathbb{E}\left(\widehat{\theta}\right) = \theta$ y asintóticamente insesgado si y sólo si $\lim\limits_{n\to+\infty} \mathbb{E}\left(\widehat{\theta}\right) = \theta$.

Si $X_1, X_2, ..., X_n$ es una M.A.S de cierta $X \in L^1$, por la ley fuerte de los grandes números, sabemos que $\overline{X}_n \xrightarrow{c.s.} \mu = \mathbb{E}(X)$ lo cual nos dice que $\widehat{\mu} = \overline{X}_n$ es un estimador fuertemente consistente de μ, además, sabemos que $\mathbb{E}\left(\overline{X}_n\right) = \mu$ lo que prueba que el estimador es además insesgado.
Por otro lado, si $X \in L^2$, el estimador natural de σ^2 es la varianza muestral, es decir $S_n^2 = \frac{1}{n}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$, se deja como ejercicio verificar que $S_n^2 \xrightarrow{c.s.} \sigma^2$, lo cual prueba que $\widehat{\sigma^2} = S_n^2$ es un estimador fuertemente consistente de σ^2, además se deja como ejercicio también verificar que $\mathbb{E}\left(S_n^2\right) = \frac{n-1}{n}\sigma^2$ lo que prueba que es asintóticamente insesgado.
Se observa que $\frac{n}{n-1}S_n^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$ es un estimador fuertemente consistente y además insesgado de σ^2.

10.2. Métodos de estimación.

Ya vimos que podemos estimar de manera fuertemente consistente e insesgada, a la esperanza y la varianza de una variable aletoria. Ahora ¿cómo se estima otro tipo de parámetros? Sería importante tener métodos que nos permitan obtener estimadores, por lo que veremos los dos más populares, el método de los momentos y el de máxima verosimilitud.

10.2.1. Método de los momentos.

Si $X_1, X_2, ..., X_n$ es una M.A.S de cierta $X \in L^k$ con distribución $F_X(x,\theta)$ con $\theta \in \Theta \subset \mathbb{R}^k$ entonces se plantean las siguientes k ecuaciones

$$\begin{cases} \mathbb{E}(X) = \overline{X}_n \\ \mathbb{E}(X^2) = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \\ \vdots \\ \mathbb{E}\left(X^k\right) = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^k \end{cases}$$

Observamos que las k igualdades se pueden ver como un sistema de k ecuaciones con k incógnitas, donde las incógnitas son $\theta_1, \theta_2, ..., \theta_k$ que aparecen del lado izquierdo en las igualdades, ya que al depender la distribución de X de los parámetros $\theta_1, \theta_2, ..., \theta_k$, entonces sus momentos $\mathbb{E}(X), \mathbb{E}(X^2), ..., \mathbb{E}\left(X^k\right)$ quedan en función de $\theta_1, \theta_2, ..., \theta_k$.
Si estas k ecuaciones con k incógnitas, admitieran una solución, $\widehat{\theta}_1,\ \widehat{\theta}_2, ...,\ \widehat{\theta}_k$, esta solución quedará en función de $\overline{X}_n, \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2, ..., \frac{1}{n}\sum\limits_{i=1}^{n} X_i^k$ quedando así los llamados estimadores por momentos de $\theta_1, \theta_2, ..., \theta_k$.
Se observa que éste método está basado en la ley de los grandes números ya que la misma nos afirma que $\overline{X}_n$ converge casi seguramente a $\mathbb{E}(X)$, $\frac{1}{n}\sum\limits_{i=1}^{n} X_i^2$ converge casi seguramente a $\mathbb{E}(X^2)$ ...$\frac{1}{n}\sum\limits_{i=1}^{n} X_i^k$ converge casi seguramente a $\mathbb{E}\left(X^k\right)$ por lo que parece natural pensar que si este sistema admite solución, la misma se debería esperar que sea fuertemente consistente.

Ejemplo 10.6. Si $X_1, X_2, ..., X_n$ es una M.A.S de cierta $X \sim U(0,b)$ entonces para hallar el estimador por el método de los momentos, dado que hay un sólo parámetro a estimar, planteamos una ecuación con una incógnita: $\mathbb{E}(X) = \overline{X}_n$, la misma nos queda $\frac{b}{2} = \overline{X}_n$ por lo que el estimador por momentos de b nos queda $\widehat{b} = 2\overline{X}_n$.
Como se observa en este caso, el estimador queda fuertemente consistente ya que

$$\overline{X}_n \overset{c.s.}{\to} \mathbb{E}(X) = \frac{b}{2} \text{ por lo que } 2\overline{X}_n \overset{c.s.}{\to} 2\frac{b}{2} = b.$$

Además es insegado ya que

$$\mathbb{E}\left(\widehat{b}\right) = \mathbb{E}\left(2\overline{X}_n\right) = 2\mathbb{E}\left(\overline{X}_n\right) = 2\frac{b}{2} = b.$$

Bajo ciertas hipótesis de regularidad, se puede probar que el estimador de un parámetro $\theta = (\theta_1, \theta_2, ..., \theta_k)$ por momentos, en caso de existir es fuertemente consistente y asintóticamente insesgado.

10.2.2. Método de máxima verosimilitud.

Si $X_1, X_2, ..., X_n$ es una M.A.S de cierta X discreta con función de probabilidad $p_X(x,\theta)$ (o absolutamente continua con función de densidad $f_X(x,\theta)$) se define la

función de verosimilitud de la muestra a la función $L\left(x_1, x_2, ..., x_n, \theta\right) = \prod_{i=1}^{n} p_X(x_i, \theta)$ o $L\left(x_1, x_2, ..., x_n, \theta\right) = \prod_{i=1}^{n} f_X(x_i, \theta)$ según el caso.
El método de máxima verosimilitud, consiste en resolver el siguiente problema de optimización:
dada $X_1, X_2, ..., X_n$ M.A.S de cierta X con distribución $F_X(x, \theta)$ con $\theta \in \Theta \subset \mathbb{R}^k$ el estimador máximo verosímil de θ es la solución al problema (si existe)

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} L\left(X_1, X_2, ..., X_n, \theta\right).$$

Es decir que para hallar el estimador máximo verosímil de θ, se debe maximizar la función $L\left(X_1, X_2, ..., X_n, \theta\right)$ como función de θ (generalmente se la llama $L\left(\theta\right)$ para recordar que miramos la función de veosimilitud como función de θ) y luego el valor de θ donde se obtiene dicho máximo (que depende de la muestra) es el estimador buscado.
Dado que la función logaritmo es creciente, el valor de θ donde se maximiza $L\left(\theta\right)$ es el mismo que el valor de θ donde se maximiza $h\left(\theta\right) = \log L\left(\theta\right)$ (el logaritmo es neperiano) muchas veces es más sencillo maximizar h.
Supongamos que luego de realizado el muestreo, obtuvimos la muestra $(x_1, x_2, ..., x_n)$ es decir que $(x_1, x_2, ..., x_n)$ es la realización de una M.A.S $(X_1, X_2, ..., X_n)$. Supongamos además que X es discreta con función de probabilidad $p_X(x, \theta)$, entonces

$$L\left(x_1, x_2, ..., x_n, \theta\right) = \prod_{i=1}^{n} p_X(x_i, \theta) = \prod_{i=1}^{n} P(X = x_i, \theta) \overset{\text{i.d.}}{=} \prod_{i=1}^{n} P(X_i = x_i, \theta) \overset{\text{indep}}{=}$$

$$P\left(X_1 = x_1, X_2 = x_2, ..., X_n = x_n, \theta\right) = P\left((X_1, X_2, ..., X_n) = (x_1, x_2, ..., x_n)\right).$$

Es decir que la función de verosimilitud es la probabilidad (en función de θ) de que la muestra $(X_1, X_2, ..., X_n)$ sea $(x_1, x_2, ..., x_n)$, que es la muestra realmente observada. Entonces, dado que es intuitivo, aunque no necesariamente cierto, pensar de que si se observó la muestra $(x_1, x_2, ..., x_n)$, entonces la misma, debería tener una probabilidad alta de ocurrir, por lo tanto como método se busca aquel valor de θ donde se maximice esta probabilidad.
Podría no existir el estimador máximo verosímil en algunas situaciones, pero vale la pena observar que si bien la función L podría no tener máximo, al menos en el caso discreto es acotada superiormente, por lo que admite supremo.

Ejemplo 10.7. Si $X_1, X_2, ..., X_n$ M.A.S de $X \sim$Ber(p), hallaremos el estimador máximo verosímil de p.

$$h(p) = \sum_{i=1}^{n} \log\, p_X(x_i, p) = \sum_{i=1}^{n} \log\, p^{x_i}\left(1-p\right)^{1-x_i} = \sum_{i=1}^{n} \left[x_i \log p + (1-x_i)\log\left(1-p\right)\right] =$$

$$\sum_{i=1}^{n} x_i \log p \ + \ \left(n - \sum_{i=1}^{n} x_i\right) \log \ (1-p)\,.$$

Luego,

$$h'(p) = \sum_{i=1}^{n} x_i \frac{1}{p} \ - \ \left(n - \sum_{i=1}^{n} x_i\right) \frac{1}{1-p}.$$

Entonces $h'(p) = 0$ si y sólo si $p = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$. Dado que para cada i se tiene que $x_i \in \{0,1\}$, entonces $\overline{X}_n \in \{0,1\}$ para todo n, entonces analizando el signo de h' vemos que h se maximiza para $\widehat{p} = \overline{X}_n$.

Ejemplo 10.8. Si $X_1, X_2, ..., X_n$ M.A.S de $X \sim U\,(0,b)$, hallaremos el estimador máximo verosímil de b.

$$L\,(b) = \prod_{i=1}^{n} f_X(x_i,\theta) = \prod_{i=1}^{n} \left\{ \begin{array}{ccc} 1/b & \text{si} & 0 < x_i < b \\ 0 & \text{si} & \text{no} \end{array} \right. = \left\{ \begin{array}{ccc} 1/b^n & \text{si} & 0 < x_1, x_2, ..., x_n < b \\ 0 & \text{si} & \text{no} \end{array} \right. .$$

Dado que L es una función decreciente cuando $b > x_1, x_2, ..., x_n$ (es decir cuando $b >$max$\{x_1, x_2, ..., x_n\}$) y 0 cuando no, se deduce que la función L se optimiza para $\widehat{b}$ =max $\{X_1, X_2, ..., X_n\}$.

Bajo ciertas condiciones de regularidad, es posible demostrar que existe el estimador máximo verosímil y es fuertemente consistente, también es posible demostrar la convergencia en distribución a una variable normal.

Capítulo 11

Intervalos de confianza.

11.1. Definición.

Dada una $X_1, X_2, ..., X_n$ muestra aleatoria simple de X cuya función de distribución es $F_X(x, \theta)$ siendo $\theta \in \Theta \subset \mathbb{R}$, en lugar de estimar el parámetro θ dando un valor numérico a partir de los datos de la muestra, daremos una región (en general un intervalo) con "probabilidad tan alta como se desee" de que el verdadero parámetro pertenezca a dicha región (intervalo).

Definición 11.1. Si $X_1, X_2, ..., X_n$ es una muestra aleatoria simple de X cuya función de distribución es $F_X(x, \theta)$ siendo $\theta \in \Theta \subset \mathbb{R}$.
Dado $\alpha \in (0,1)$, supongamos que $a(X_1, X_2, ..., X_n)$ y $b(X_1, X_2, ..., X_n)$ son dos estadísticos tales que $P(\theta \in [a(X_1, X_2, ..., X_n); b(X_1, X_2, ..., X_n)]) = 1 - \alpha$, diremos entonces que $I = [a(X_1, X_2, ..., X_n); b(X_1, X_2, ..., X_n)]$ es un intervalo de confianza de nivel $1 - \alpha$ para el parámetro θ.

Observación 11.2. *Observamos que en la práctica el valor de α (o equivalentemente el nivel de confianza $1 - \alpha$) está determinado por el investigador, por lo que es un valor fijo.*

Observación 11.3. *Observemos también que una vez que la muestra $X_1, X_2, ..., X_n$ es realizada en los números $x_1, x_2, ..., x_n$, el intervalo $I = [a(x_1, x_2, ..., x_n); b(x_1, x_2, ..., x_n)]$ no es aleatorio y por lo tanto, la probabilidad de que $\theta \in I$ es 0 o 1 según el parámetro $\theta \in I$ o $\theta \notin I$, entonces vale observar que el intervalo $I = [a(X_1, X_2, ..., X_n); b(X_1, X_2, ..., X_n)]$ es aleatorio, mientras que el intervalo $I = [a(x_1, x_2, ..., x_n); b(x_1, x_2, ..., x_n)]$ es fijo, para distinguir una situación de otra, se le suele llamar a éste último, intervalo de confianza mientras que al otro se le suele denominar intervalo aleatorio. En lo que sigue, seremos informales en la escritura y les llamaremos a ambos intervalos de confianza, a pesar de que debemos tener clara su diferencia.*

11.2. Construcción de intervalos de confianza en algunos casos particulares.

En esta sección, construiremos intervalos de confianza en algunos casos particulares. Si $X_1, X_2, ..., X_n$ es una muestra de $X \sim N(\mu, \sigma^2)$. Supongamos que conocemos σ^2. Construiremos un intervalo de confianza para el parámetro desconocido μ.
Sabemos por la ley fuerte de los grandes números que $\overline{X}_n \stackrel{c.s.}{\to} \mu$, por lo que es razonable formar un intervalo de la forma $\left[\overline{X}_n - k; \overline{X}_n + k\right]$ siempre y cuando podamos hallar k de modo que $P\left(\mu \in \left[\overline{X}_n - k; \overline{X}_n + k\right]\right) = 1 - \alpha$ cumpliendo además que k no dependa de parámetros desconocidos.
Recordamos que $\overline{X}_n \sim N(\mu, \sigma^2/n)$ por ser combinación lineal de normales. Entonces

$$P\left(\mu \in \left[\overline{X}_n - k; \overline{X}_n + k\right]\right) = P\left(\mu - k \leq \overline{X}_n \leq \mu + k\right) =$$
$$\Phi\left(\frac{\mu + k - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{\mu - k - \mu}{\sigma/\sqrt{n}}\right) =$$
$$\Phi\left(\frac{k\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{-k\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{k\sqrt{n}}{\sigma}\right) - 1 = 1 - \alpha,$$

luego $\Phi\left(\frac{k\sqrt{n}}{\sigma}\right) = 1 - \alpha/2$ por lo que $\frac{k\sqrt{n}}{\sigma} = \Phi^{-1}(1 - \alpha/2)$ de donde obtenemos $k = \frac{\Phi^{-1}(1-\alpha/2)\sigma}{\sqrt{n}}$. Dado que en este caso el valor de σ^2 se supone conocido, tenemos que k no depende de parámetros desconocidos.
Si además para cada $p \in (0, 1)$ le llamamos $z_p = \Phi^{-1}(p)$ tendremos entonces el intervalo de confianza para este caso como sigue

$$\left[\overline{X}_n - \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}; \overline{X}_n + \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}\right].$$

En el caso en que $X_1, X_2, ..., X_n$ es una muestra aleatoria simple de $X \sim N(\mu, \sigma^2)$ donde σ^2 es desconocido, si bien la igualdad calculada es válida, carece de valor ya que en este caso $a(X_1, X_2, ..., X_n) = \overline{X}_n - \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}$ no es un estadístico (tampoco lo es b) por lo que no es válido como intevalo de confianza. Para obtener un intervalo en estos caso introducimos dos nuevas familias de variables aleatorias.

Definición 11.4. Se dice que X tiene una distribución $t-$student con n grados de libertad cuando tiene la siguiente densidad:

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}$$

Notación: $X \sim t_n$.
Se observa que si $X \sim t_n$ entonces $\mathbb{E}(X) = 0$ para $n > 1$ (si $n \leq 1$, entonces no existe la esperanza) y se puede verificar que $\mathbb{V}(X) = \frac{n}{n-2}$ para $n > 2$ (si $n \leq 2$, no admite momentos de orden 2).

Definición 11.5. Se dice que X tiene una distribución χ^2 con n grados de libertad cuando tiene la siguiente densidad:

$$f_X(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} \mathbf{1}_{\{x>0\}}$$

Notación: $X \sim \chi^2_n$.
Se puede verificar que si $X \sim \chi^2_n$, entonces $\mathbb{E}(X) = n$ y $\mathbb{V}(X) = 2n$.
Para obtener un intervalo de confianza para μ en estos casos, nos serviremos del siguiente teorema (que no demostraremos).

Teorema 11.6. *Si $X_1, X_2, ..., X_n$ es una muestra de $X \sim N(\mu, \sigma^2)$ entonces $\frac{\sqrt{n-1}\left(\overline{X}_n - \mu\right)}{S_n} \sim T(n-1)$ ($t-$student con $n-1$ grados de libertad) siendo $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$ la varianza muestral.*

Entonces, veamos que en este caso podemos determinar k, no dependiendo de parámetros desconocidos, de modo que el intervalo de confianza de nivel $1-\alpha$ sea de la forma $\left[\overline{X}_n - kS_n; \overline{X}_n + kS_n\right]$. Para abreviar, le llamamos T a la variable $\frac{\sqrt{n-1}\left(\overline{X}_n - \mu\right)}{S_n}$. Entonces

$$P\left(\mu \in \left[\overline{X}_n - kS_n; \overline{X}_n + kS_n\right]\right) = P\left(\left|\overline{X}_n - \mu\right| \leq kS_n\right) =$$
$$P\left(\frac{\sqrt{n-1}\left|\overline{X}_n - \mu\right|}{S_n} \leq \sqrt{n-1}k\right) =$$
$$P(|T| \leq k) = P\left(-k \leq T \leq \sqrt{n-1}k\right) =$$
$$F_T(\sqrt{n-1}k) - F_T(-\sqrt{n-1}k) = 2F_T(\sqrt{n-1}k) - 1$$

ya que por simetría de la distribución de Student, se tiene que $F_T(-t) = 1 - F_T(t)$. Entonces $2F_T(\sqrt{n-1}k) - 1 = 1 - \alpha$ de donde obtenemos $k = F_T^{-1}(1-\alpha/2)$ que no depende de parámetros desconocidos. Nuevamente si le llamamos $t_p(n) = F^{-1}(p)$ para $p \in (0,1)$ y F función de distribución correspondiente a una variable $t-$Student con n grados de libertad, tenemos que $\sqrt{n-1}k = F_T^{-1}(1-\alpha/2) = t_{1-\alpha/2}(n-1)$ por lo que tenemos el intervalo de confianza en la forma

$$\left[\overline{X}_n - \frac{S_n t_{1-\alpha/2}(n-1)}{\sqrt{n-1}}; \overline{X}_n + \frac{S_n t_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right].$$

Para completar el caso de la variable normal, construiremos en lo que sigue un intervalo de confianza para σ^2. Para ello nos serviremos del siguiente teorema (que no demostraremos).

Teorema 11.7. *Si $X_1, X_2, ..., X_n$ es una muestra de $X \sim N(\mu, \sigma^2)$ entonces $\frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1)$ (χ^2 con $n-1$ grados de libertad) siendo $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$ la varianza muestral.*

Dado que $S_n^2 \overset{c.s.}{\rightarrow} \sigma^2$ podríamos nuevamente intentar buscar un intervalo en la forma $[S_n^2 - k; S_n^2 + k]$ pero la idea no funciona, por lo tanto veremos si podemos encontrar valores a y b tales que el intervalo quede en la forma $[aS_n^2; bS_n^2]$. Planteamos entonces la ecuación $1-\alpha = P\left(\sigma^2 \in [aS_n^2; bS_n^2]\right)$ y hallaremos a y b tales que $P\left(\sigma^2 < aS_n^2\right) = \alpha/2$ y $P\left(\sigma^2 > bS_n^2\right) = \alpha/2$. Para simplificar le llamaremos χ^2 a la distribución de $\frac{nS_n^2}{\sigma^2}$. Entonces

$$P\left(\sigma^2 < aS_n^2\right) = P\left(S_n^2 > \frac{\sigma^2}{a}\right) = P\left(\frac{nS_n^2}{\sigma^2} > \frac{n}{a}\right) = 1 - F_{\chi^2}\left(\frac{n}{a}\right) = \alpha/2$$

por lo que $\frac{n}{a} = F_{\chi^2}^{-1}(\alpha/2)$ de donde obtenemos $a = \frac{n}{F_{\chi^2}^{-1}(\alpha/2)}$ y nuevamente, llamándole $\chi_p^2(n) = F^{-1}(p)$ siendo la función de distribución asociada a una variable $\chi^2(n)$ y observando que en este caso la variable con la cual estamos distribuye $\chi^2(n-1)$ obtenemos $a = \frac{n}{\chi_{\alpha/2}^2(n-1)}$. Trabajando análogamente con la otra igualdad se obtiene que $b = \frac{n}{\chi_{1-\alpha/2}^2(n-1)}$ y por lo tanto el intervalo de confianza para σ^2 de nivel $1-\alpha$ nos queda

$$\left[\frac{nS_n^2}{\chi_{\alpha/2}^2(n-1)}; \frac{nS_n^2}{\chi_{1-\alpha/2}^2(n-1)}\right].$$

En numerosas situaciones, se tiene una muestra $X_1, X_2, ..., X_n$ de cierta X desconocida. Si el tamaño de muestra es grande, y suponemos que $X \in L^2$ y deseamos estimar $\mu = \mathbb{E}(X)$ mediante un intervalo de confianza, entonces podemos aplicar el teorema central del límite y realizar algunos cálculos similares a los realizados, obteniéndose así intervalos de confianza de nivel aproximadamente iguales a $1-\alpha$.

Efectivamente, debido al teorema central del límite podemos afirmar que (en el caso n suficientemente grande) la distribución de $\overline{X}_n$ es aproximadamente $N(\mu, \sigma^2/n)$. Por lo tanto, $P\left(\mu \in \left[\overline{X}_n - k; \overline{X}_n + k\right]\right) =$

$$P\left(\mu - k \leq \overline{X}_n \leq \mu + k\right) \overset{\text{TCL}}{\cong} \Phi\left(\frac{\mu + k - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{\mu - k - \mu}{\sigma/\sqrt{n}}\right) =$$

$$\Phi\left(\frac{k\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{-k\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{k\sqrt{n}}{\sigma}\right) - 1 = 1 - \alpha,$$

por lo que obtendríamos que el intervalo $I = \left[\overline{X}_n - \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}; \overline{X}_n + \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}\right]$ es tal que $P(\mu \in I) \cong 1-\alpha$. Ahora, como el intervalo depende de un parámetro desconocido (σ), no nos sirve como intervalo de confianza, pero recordando que n es grande, podemos sustituir σ por un estimador consistente del mismo, por ejemplo $S_n = \sqrt{S_n^2}$ obteniéndose de esa forma el intervalo

$$\left[\overline{X}_n - \frac{S_n z_{1-\alpha/2}}{\sqrt{n}}; \overline{X}_n + \frac{S_n z_{1-\alpha/2}}{\sqrt{n}}\right]$$

que es, ahora si, un intervalo de confianza de un nivel aproximadamente igual a $1-\alpha$. Como caso particular podemos obtener un intervalo de confianza aproximado para p

cuando X ~Ber(p) en el caso en que n es grande. Efectivamente, cuando X ~Ber(p) entonces $\mu = \mathbb{E}(X) = p$, además, como $X_i = X_i^2$, entonces $S_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2 = \overline{X}_n - \overline{X}_n^2 = \overline{X}_n\left(1 - \overline{X}_n\right)$, obteniéndose así un intervalo de confianza para p

$$\left[\overline{X}_n - \frac{\sqrt{\overline{X}_n\left(1-\overline{X}_n\right)}z_{1-\alpha/2}}{\sqrt{n}}; \overline{X}_n + \frac{\sqrt{\overline{X}_n\left(1-\overline{X}_n\right)}z_{1-\alpha/2}}{\sqrt{n}}\right]$$

cuyo nivel es aproximadamente $1 - \alpha$.

11.3. Resumen.

Recordemos que dado $p \in (0,1)$ usamos las siguientes notaciones para $F^{-1}(p)$: z_p si F es la función de distribución de una variable $N(0,1)$; $t_p(n)$ si F es la distribcuión de una variable t_n ($t-$student con n grados de libertad) y χ_p^2 cuando F es la fucnión de distribución de una variable χ_n^2 dada $X_1, X_2, ..., X_n$ muestra de X, hemos obtenido intervalos de confianza para los siguientes casos.

1. Intervalo de confianza para $\mu = \mathbb{E}(X)$ al nivel $1 - \alpha$.

   a) Si $X \sim N(\mu, \sigma^2)$ con σ^2 conocido,
   $$\left[\overline{X}_n - \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}; \overline{X}_n + \frac{\sigma z_{1-\alpha/2}}{\sqrt{n}}\right].$$

   b) Si $X \sim N(\mu, \sigma^2)$ con σ^2 desconocido,
   $$\left[\overline{X}_n - \frac{S_n t_{1-\alpha/2}(n-1)}{\sqrt{n-1}}; \overline{X}_n + \frac{S_n t_{1-\alpha/2}(n-1)}{\sqrt{n-1}}\right].$$

   c) Si $X \in L^2$ y n es suficientemente grande, un intervalo aproximado es
   $$\left[\overline{X}_n - \frac{S_n z_{1-\alpha/2}}{\sqrt{n}}; \overline{X}_n + \frac{S_n z_{1-\alpha/2}}{\sqrt{n}}\right].$$

2. Intervalo de confianza para p al nivel $1 - \alpha$ cuando X ~Ber(p) y n es suficientemente grande, un intervalo aproximado es
   $$\left[\overline{X}_n - \frac{\sqrt{\overline{X}_n\left(1-\overline{X}_n\right)}z_{1-\alpha/2}}{\sqrt{n}}; \overline{X}_n + \frac{\sqrt{\overline{X}_n\left(1-\overline{X}_n\right)}z_{1-\alpha/2}}{\sqrt{n}}\right].$$

3. Intervalo de confianza para σ^2 en el caso en que $X \sim N(\mu, \sigma^2)$.
   $$\left[\frac{nS_n^2}{\chi_{\alpha/2}^2(n-1)}; \frac{nS_n^2}{\chi_{1-\alpha/2}^2(n-1)}\right].$$