

Notas de Probabilidades y Estadística

Capítulos 1 al 12

Víctor J. Yohai

vyohai@dm.uba.ar

Basadas en apuntes de clase tomados por Alberto Déboli, durante el año 2003
Versión corregida durante 2004 y 2005, con la colaboración de María Eugenia Szretter

5 de Marzo de 2008

Índice general

1. Espacios de Probabilidad.	7
1.1. Experimentos aleatorios. Algunas consideraciones heurísticas.	7
1.2. Axiomas de probabilidad.	8
1.2.1. σ -Álgebras.	8
1.2.2. Espacios de Probabilidad.	10
1.3. σ -Álgebra generada por una familia de conjuntos.	18
1.4. Espacios de probabilidad finitos o numerables.	21
1.5. Probabilidad condicional.	23
1.6. Independencia de eventos.	25
2. Variable Aleatoria.	31
2.1. Concepto de variable aleatoria.	31
2.2. Espacio de probabilidad asociado a una variable aleatoria. . .	32
2.3. Función de distribución de una variable aleatoria.	35
3. Variables aleatorias discretas y continuas.	41
3.1. Variables aleatorias discretas.	41
3.2. Ejemplos de distribuciones discretas.	43
3.2.1. Distribución Binomial.	43
3.2.2. Distribución Binomial Negativa (o Distribución de Pas- cal).	45
3.2.3. Distribución Geométrica.	46
3.2.4. Distribución Hipergeométrica.	47
3.2.5. Distribución de Poisson.	48
3.2.6. Gráfico de la función de distribución asociada a una variable aleatoria discreta.	49
3.3. Variables aleatorias absolutamente continuas.	49
3.4. Ejemplos de distribuciones continuas.	53
3.4.1. Distribución uniforme en un intervalo.	53
3.4.2. Generación de distribuciones a partir de la distribu- ción uniforme en $[0,1]$	55
3.4.3. Distribución Normal $N(\mu, \sigma^2)$	59
3.4.4. Distribución Exponencial.	62

3.5.	Variables aleatorias mixtas.	65
4.	Vectores aleatorios.	69
4.1.	Definición de vector aleatorio.	69
4.2.	Espacio de probabilidad inducido.	70
4.3.	Función de distribución conjunta de un vector aleatorio.	71
4.4.	Algunas propiedades de vectores aleatorios.	78
4.5.	Independencia de variables aleatorias.	80
4.5.1.	Algunas consideraciones heurísticas.	80
4.5.2.	Conservación de la independencia por transformaciones.	86
4.5.3.	Independencia de vectores aleatorios.	86
5.	Vectores aleatorios discretos y continuos.	89
5.1.	Vectores aleatorios discretos.	89
5.1.1.	Función de densidad de probabilidad conjunta.	91
5.1.2.	Caracterización de la función de densidad marginal asociada a un subconjunto de variables.	92
5.2.	Ejemplos de vectores aleatorios con distribución discreta.	94
5.2.1.	Distribución Multinomial.	94
5.2.2.	Distribución Hipergeométrica Multivariada.	96
5.3.	Vectores Aleatorios de tipo absolutamente continuo.	98
6.	Transformaciones de variables y vectores aleatorios.	105
6.1.	Transformaciones monótonas de variables aleatorias.	105
6.1.1.	Distribución Normal	107
6.2.	Transformaciones inyectivas de vectores aleatorios.	109
6.3.	Algunas aplicaciones a la distribución normal.	112
6.4.	Transformaciones no inyectivas	114
6.4.1.	Distribución Chi-cuadrado con un grado de libertad.	115
6.5.	Algunas distribuciones complementarias.	116
6.5.1.	Distribución Gamma.	116
6.5.2.	Distribución beta.	121
6.5.3.	Distribución Chi-cuadrado.	123
6.5.4.	Distribución t de Student	123
7.	Esperanza Matemática.	125
7.1.	Integral de Riemann-Stieltjes.	125
7.1.1.	Definición de la integral.	125
7.2.	Definición de Esperanza Matemática.	128
7.2.1.	Algunas consideraciones heurísticas.	128
7.2.2.	Esperanza de una variable aleatoria discreta.	129
7.2.3.	Definición general de esperanza matemática.	129
7.2.4.	Esperanza matemática para una variable absolutamente continua.	133

7.2.5.	Algunas propiedades de la esperanza matemática . . .	134
7.3.	Esperanza del producto de variables aleatorias independientes.	149
7.4.	Una fórmula general para la esperanza de una variable trans- formada	151
7.5.	Esperanza de distribuciones simétricas	154
7.6.	Mediana de una variable aleatoria.	158
7.7.	Varianza de una variable aleatoria.	161
7.7.1.	Esperanzas y varianzas de distribuciones normales . .	163
7.8.	Covarianza	165
7.9.	Distribución Normal Bivariada.	167
8.	Teoría de la Predicción.	173
8.1.	Error cuadrático medio y predictores óptimos.	173
8.2.	Predictores constantes.	175
8.3.	Predictores lineales.	176
9.	Esperanza y distribución condicional.	179
9.1.	Caso discreto.	179
9.2.	Caso general	187
9.3.	Caso continuo	190
9.4.	Varianza condicional	192
10.	Convergencia de Variables Aleatorias.	195
10.1.	Convergencia de funciones.	195
10.2.	Convergencia casi segura y en probabilidad.	196
10.3.	Preservación de la convergencia por funciones continuas. . . .	199
10.4.	Ley débil de los grandes números.	204
10.5.	Ley fuerte de los grandes números.	207
10.6.	Teorema de la Convergencia Dominada	213
11.	Convergencia en Distribución.	217
11.1.	Definición de convergencia en distribución.	217
11.2.	Funciones características.	220
11.2.1.	Variables aleatorias complejas.	220
11.2.2.	Definición de función característica y propiedades. . .	221
11.3.	Momentos y función característica.	226
11.3.1.	Derivación dentro del signo esperanza.	226
11.3.2.	Derivadas de la función característica y momentos. . .	227
11.4.	Función característica de una distribución normal.	229
11.5.	Teorema Central del Límite.	233
11.5.1.	Caso de variables independientes idénticamente dis- tribuidas	233
11.5.2.	Teorema Central del Límite para variables no idénti- camente distribuidas.	236

11.5.3. Una Aplicación a la Binomial.	240
11.6. Teorema de Slutsky.	242
11.7. Aplicación a intervalos de confianza.	253
11.8. Un teorema útil de Convergencia en Distribución	255
12. Procesos de Poisson.	257
12.1. Procesos de punto.	257
12.2. Axiomática de los Procesos de Poisson	257
12.3. Distribución de un proceso de Poisson.	259
12.4. Tiempos de espera	264
12.5. Procesos de Poisson en el plano.	265

Capítulo 1

Espacios de Probabilidad.

1.1. Experimentos aleatorios. Algunas consideraciones heurísticas.

Se llamará experimento aleatorio a un experimento tal que (i) no se puede preveer el resultado de un solo experimento, (ii) si se repite el experimento varias veces, la frecuencia con la cual el resultado está en un conjunto A converge a un número.

Ejemplo 1.1 *El experimento consiste en arrojar una moneda. En este caso el conjunto de todos los posibles resultados será*

$$\Omega = \{0, 1\},$$

0 corresponde a ceca y 1 a cara. Si se repite experimento muchas veces, la frecuencia con que sale por ejemplo cara, tiende a 0.5

Ejemplo 1.2 *El experimento consiste en lanzar un dado. En este caso el conjunto de todos los posibles resultados será*

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Si se tira el dado muchas veces, por ejemplo la frecuencia con que el resultado está en el conjunto $A \subset \Omega$ será $\#A/6$, donde $\#A$ representa el cardinal de A .

Ejemplo 1.3 *El experimento consiste en lanzar una jabalina y registrar la marca obtenida. En este caso el conjunto de todos los posibles resultados será el conjunto de reales positivos y la frecuencia con que el resultado esté, por ejemplo en un intervalo $[a, b]$, dependerá del atleta.*

Ejemplo 1.4 *Se elige al azar un alumno de primer grado de un colegio y se anota su peso en kilos, x y la altura en metros y En este caso*

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x > 0, y > 0\}.$$

Como puede apreciarse los resultados pueden conformar un conjunto finito o infinito de cualquier cardinalidad.

Supongamos ahora que se hacen n repeticiones del experimento aleatorio. Si $A \subset \Omega$, sea $C_n(A)$ el número de veces que el resultado está en A , luego la frecuencia relativa del conjunto A se define por

$$f_n(A) = \frac{C_n(A)}{n}.$$

En el caso de un experimento aleatorio, cuando n crece, esta frecuencia se aproxima a un número que se llamará probabilidad de A y que denotaremos por $P(A)$.

Claramente

$$0 \leq f_n(A) \leq 1,$$

de manera que

$$P(A) = \lim_{n \rightarrow \infty} f_n(A),$$

y entonces

$$0 \leq P(A) \leq 1.$$

Como veremos, en algunos casos, no se puede definir la probabilidad para todo subconjunto de resultados.

Para precisar este concepto y estudiar sus propiedades formularemos la teoría axiomática de probabilidades.

1.2. Axiomas de probabilidad.

En primer lugar definiremos algunas propiedades que tendrá la familia de todos los conjuntos para los cuales está definida su probabilidad. Esto nos lleva al concepto de σ -álgebra.

1.2.1. σ -Álgebras.

Sea Ω un conjunto. Definiremos el conjunto partes de Ω , por $\mathcal{P}(\Omega) = \{A : A \subset \Omega\}$. Dado un conjunto A , denotaremos por A^c el complemento de A .

Definición 1.1 *Sea una familia \mathcal{A} de subconjuntos de Ω , es decir $\mathcal{A} \subset \mathcal{P}(\Omega)$. Se dice que \mathcal{A} es una σ -álgebra sobre Ω si satisface las siguientes propiedades.*

A1. $\Omega \in \mathcal{A}$.

A2. Dado $A \in \mathcal{A}$ se tiene $A^c \in \mathcal{A}$.

A3. Sea A_1, \dots, A_n, \dots una sucesión de elementos de \mathcal{A} . Entonces

$$A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}.$$

Propiedades de σ -álgebras

Propiedad 1.1 $\emptyset \in \mathcal{A}$.

Demostración. Resulta de A1 y A2. \square

Propiedad 1.2 Si A_1, \dots, A_n son elementos de \mathcal{A} entonces

$$\bigcup_{i=1}^n A_i \in \mathcal{A}.$$

Demostración.

Para ver esto supongamos que $A_i \in \mathcal{A}$; $i = 1, 2, \dots, n$. Probaremos que

$$A = \bigcup_{i=1}^n A_i \in \mathcal{A}.$$

Definamos una sucesión numerable $(B_i)_{i \geq 1}$ agregando el conjunto \emptyset de la siguiente manera

$$\begin{aligned} B_j &= A_j, \quad 1 \leq j \leq n, \\ B_k &= \emptyset \quad \text{si } k > n. \end{aligned}$$

Entonces por ser \mathcal{A} una σ -álgebra se tendrá que $\bigcup_{i=1}^{\infty} B_i \in \mathcal{A}$ y por lo tanto

$$A = \bigcup_{i=1}^n A_i = \bigcup_{i=1}^{\infty} B_i \in \mathcal{A}. \quad \square$$

Propiedad 1.3 Si \mathcal{A} es una σ -álgebra, y A_1, \dots, A_n, \dots es una sucesión de elementos de \mathcal{A} entonces $A = \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

Demostración. Esto resulta de que $A = \left(\bigcup_{i=1}^{\infty} A_i^c \right)^c$. \square

Propiedad 1.4 Si \mathcal{A} es una σ -álgebra, y A_1, \dots, A_n son elementos de \mathcal{A} entonces $A = \bigcap_{i=1}^n A_i \in \mathcal{A}$.

Demostración. Se demuestra igual que la Propiedad 1.2. \square

Propiedad 1.5 Si \mathcal{A} es una σ -álgebra, y A_1 y A_2 son elementos de \mathcal{A} , entonces $A_1 - A_2 \in \mathcal{A}$.

Demostración. En efecto $A_1 - A_2 = A_1 \cap A_2^c \in \mathcal{A}$. \square

Propiedad 1.6 La σ -álgebra sobre Ω más chica posible es

$$\mathcal{A}_0 = \{\Omega, \emptyset\},$$

y la más grande es

$$\mathcal{A}_1 = \mathcal{P}(\Omega).$$

Luego si \mathcal{A} es una σ -álgebra sobre Ω , se tendrá

$$\mathcal{A}_0 \subset \mathcal{A} \subset \mathcal{A}_1. \quad \square$$

Observación. En el contexto de la teoría de la medida, un elemento de la σ -álgebra \mathcal{A} se llama un conjunto medible.

Como veremos en la próxima subsección, la probabilidad estará definida para los elementos de una σ -álgebra.

1.2.2. Espacios de Probabilidad.

Definición 1.2 Un espacio de probabilidad es una terna (Ω, \mathcal{A}, P) donde Ω es un conjunto, \mathcal{A} es una σ -álgebra sobre Ω , y $P : \mathcal{A} \rightarrow [0; 1]$ es una función que satisface:

1. $P(\Omega) = 1$.
2. (σ -aditividad). Si $(A_n)_{n \geq 1}$ es una sucesión de elementos de \mathcal{A} disjuntos dos a dos ($A_i \cap A_j = \emptyset$, si $i \neq j$), entonces

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Observaciones.

1. El conjunto Ω se denomina *espacio muestral* y se interpreta como el conjunto de resultados posibles del experimento, los elementos de \mathcal{A} se denominan *eventos*, y corresponden a los subconjuntos de Ω para los cuales la probabilidad está definida. Finalmente P se denomina *función de probabilidad*, y dado $A \in \mathcal{A}$, $P(A)$ se interpreta como la probabilidad de que el resultado del experimento esté en A .
2. En el contexto de la teoría de la medida, la terna (Ω, \mathcal{A}, P) corresponde a un espacio de medida donde la medida P asigna el valor uno al espacio total.
3. Si queremos formalizar la idea intuitiva de la probabilidad como límite de la frecuencia relativa es importante observar que la “frecuencia” tiene la propiedad de σ -aditividad. En principio veamos que debería ser aditiva

Sean A_1, A_2, \dots, A_k eventos disjuntos tomados de a dos, esto es, $A_i \cap A_j = \emptyset$ si $i \neq j$ entonces

$$f_n \left(\bigcup_{i=1}^k A_i \right) = \frac{C_n \left(\bigcup_{i=1}^k A_i \right)}{n} = \frac{\sum_{i=1}^k C_n(A_i)}{n} = \sum_{i=1}^k f_n(A_i).$$

La σ -aditividad ahora se deduce pasando al límite.

Ejemplos de espacios de probabilidad.

Ejemplo 1.5 Sea Ω un conjunto, $\mathcal{A} = \mathcal{P}(\Omega)$. Dado $x_0 \in \Omega$, definimos: $\forall A \subseteq \Omega$

$$P(A) = \begin{cases} 1 & \text{si } x_0 \in A \\ 0 & \text{si } x_0 \notin A. \end{cases}$$

P se denota δ_{x_0} y se dice que la probabilidad está concentrada en x_0 o bien que el único punto de probabilidad positiva es x_0 .

Ejemplo 1.6 Sea $\Omega = \{x_1, x_2, \dots, x_n, \dots\}$ cualquier conjunto numerable, $\mathcal{A} = \mathcal{P}(X)$, y sea $a_i \geq 0$, $i = 1, 2, \dots$, una sucesión tal que

$$\sum_{i=1}^{\infty} a_i = 1.$$

Definimos para todo $A \subseteq \Omega$

$$P(A) = \sum_{\{i: x_i \in A\}} a_i$$

En este caso P define una probabilidad y está completamente determinada por las probabilidades a_i asignadas a cada elemento x_i .

Propiedades de la función de probabilidad.

Propiedad 1.7 $P(\emptyset) = 0$.

Demostración. Es inmediata, pues si tomamos $A_i = \emptyset$, para todo $i \in \mathbb{N}$ entonces por la σ -aditividad

$$0 \leq P(\emptyset) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(\emptyset) \leq 1,$$

y esto sólo se cumple en el caso de que $P(\emptyset) = 0$. \square

Propiedad 1.8 Sean A_1, \dots, A_n eventos disjuntos. Luego $P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$.

Demostración. Tomemos la sucesión $B_j = A_j$ si $j = 1, \dots, n$ y $B_j = \emptyset$ si $j > n$. Aplicando la propiedad de σ -aditividad se obtiene el resultado. \square

Propiedad 1.9 Si $A \in \mathcal{A}$ entonces

$$P(A^c) = 1 - P(A).$$

Demostración. Esto sale teniendo en cuenta que A y A^c son disjuntos y

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c). \quad \square$$

Propiedad 1.10 Consideremos dos eventos A_1 y A_2 . Entonces

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

Demostración. Como

$$A_1 = (A_1 - A_2) \cup (A_1 \cap A_2)$$

se obtiene

$$P(A_1) = P(A_1 - A_2) + P(A_1 \cap A_2),$$

y de ahí sigue el resultado. \square

Proposición 1.1 Si A_1, A_2 son eventos y $A_2 \subset A_1$ entonces

$$P(A_1 - A_2) = P(A_1) - P(A_2).$$

y además

$$P(A_2) \leq P(A_1).$$

Demostración. Por la Propiedad 1.1 y el hecho de que $A_1 \cap A_2 = A_2$ tenemos

$$\begin{aligned} P(A_1 - A_2) &= P(A_1) - P(A_1 \cap A_2) \\ &= P(A_1) - P(A_2) \end{aligned}$$

Además de aquí resulta

$$\begin{aligned} P(A_1) &= P(A_2) + P(A_1 - A_2) \\ &\geq P(A_2). \quad \square \end{aligned}$$

Propiedad 1.11 *Si A_1, A_2 son eventos entonces*

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

Demostración. Escribimos $A_1 \cup A_2$ como la siguiente unión disjunta

$$A_1 \cup A_2 = (A_1 - A_2) \cup (A_1 \cap A_2) \cup (A_2 - A_1).$$

Entonces usando la Propiedad 1.10 resulta

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1 - A_2) + P(A_1 \cap A_2) + P(A_2 - A_1) = \\ &= P(A_1) - P(A_1 \cap A_2) + P(A_1 \cap A_2) \\ &\quad + P(A_2) - P(A_1 \cap A_2) \\ &= P(A_1) + P(A_2) - P(A_1 \cap A_2). \quad \square \end{aligned}$$

Propiedad 1.12 *Sean $A_i \in \mathcal{A}$, $i = 1, 2, \dots, k$. Entonces*

$$P\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{i=1}^k P(A_i).$$

Demostración. De la Propiedad 1.11 se obtiene

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2),$$

y el resultado vale para $k = 2$. El resto de la demostración se hace por inducción y se deja como ejercicio.

Propiedad 1.13 (σ -subaditividad) Sea $(A_n)_{n \geq 1} \subset \mathcal{A}$ y $A = \bigcup_{n=1}^{\infty} A_n$. Entonces

$$P(A) \leq \sum_{n=1}^{\infty} P(A_n).$$

Demostración. Definamos

$$\begin{aligned} B_0 &= \emptyset, \\ B_1 &= A_1, \\ B_2 &= A_2 - A_1, \\ B_3 &= A_3 - (A_1 \cup A_2), \\ &\vdots \\ B_n &= A_n - \bigcup_{i=1}^{n-1} A_i. \end{aligned}$$

Luego es inmediato que los B_i son disjuntos dos a dos y

$$A = \bigcup_{n=1}^{\infty} B_n.$$

Por la σ -aditividad y el hecho de que $B_n \subset A_n$, resulta $P(B_n) \leq P(A_n)$ y entonces

$$P(A) = \sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n). \quad \square$$

Propiedad 1.14 Sea $(A_n)_{n \geq 1}$ una sucesión de eventos tales que $A_n \subset A_{n+1}$ para todo n y

$$A = \bigcup_{i=1}^{\infty} A_i.$$

Luego

$$P(A) = \lim_{n \rightarrow +\infty} P(A_n).$$

Demostración. Como la sucesión es creciente entonces podemos transformar la unión en una unión disjunta definiendo: $B_0 = A_0 = \emptyset$, $B_1 = A_1 - A_0$, $B_2 = A_2 - A_1, \dots, B_k = A_k - A_{k-1}, \dots$ Luego

$$A = \bigcup_{k=1}^{\infty} B_k,$$

y por lo tanto usando la σ -aditividad y la Propiedad 1.1 se tiene

$$\begin{aligned} P(A) &= \sum_{k=1}^{\infty} P(B_k) = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(B_k) = \lim_{n \rightarrow \infty} \sum_{k=1}^n P(A_k - A_{k-1}) \\ &= \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n P(A_k) - \sum_{k=1}^n P(A_{k-1}) \right) = \lim_{n \rightarrow \infty} P(A_n). \quad \square \end{aligned}$$

Propiedad 1.15 Sea $(A_n)_{n \geq 1}$ una sucesión de eventos tal que $A_n \supset A_{n+1}$ para todo n y

$$A = \bigcap_{i=1}^{\infty} A_i.$$

Entonces

$$P(A) = \lim_{n \rightarrow +\infty} P(A_n).$$

Demostración. Sea $B_n = A_n^c$. Luego $(B_n)_{n \geq 1}$ es una sucesión creciente de eventos y $A^c = \bigcup_{i=1}^{\infty} B_i$. Luego por la propiedad anterior tenemos

$$\begin{aligned} 1 - P(A) &= P(A^c) \\ &= \lim_{n \rightarrow +\infty} P(B_n) \\ &= \lim_{n \rightarrow +\infty} (1 - P(A_n)) \\ &= 1 - \lim_{n \rightarrow +\infty} P(A_n), \end{aligned}$$

de donde se obtiene el resultado deseado. \square

Definición 1.3 Se llama límite superior de una sucesión de conjuntos $(A_n)_{n \geq 1} \subset \Omega$ al conjunto

$$\overline{A} = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n,$$

y límite inferior de la sucesión al conjunto

$$\underline{A} = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n.$$

Además

$$\begin{aligned} (\underline{A})^c &= \left(\bigcup_{k \geq 1} \bigcap_{n=k}^{\infty} A_n \right)^c = \bigcap_{k \geq 1} \left(\bigcap_{n=k}^{\infty} A_n \right)^c = \\ &= \bigcap_{k \geq 1} \bigcup_{n=k}^{\infty} A_n^c = \overline{A^c}. \end{aligned}$$

Es decir el complemento del límite inferior de la sucesión $(A_n)_{n \geq 1}$ es el límite superior de la sucesión $(A_n^c)_{n \geq 1}$.

Propiedad 1.16 (*Caracterización de los límites superiores e inferiores*)

(i) Sea

$$A^\infty = \{\omega \in \Omega : \omega \text{ está en infinitos conjuntos } A_n\}.$$

$$\text{Luego } \overline{A} = A^\infty.$$

(ii) Sea

$$A_\infty = \{\omega \in \Omega : \omega \text{ está en todos los } A_n \text{ salvo en un número finito}\}.$$

$$\text{Luego } \underline{A} = A_\infty.$$

(iii) $\underline{A} \subset \overline{A}$

Demostración.

(i) Supongamos que $\omega \in A^\infty$ entonces para todo $k \in \mathbb{N}$ se tiene que $\omega \in \bigcup_{n=k}^{\infty} A_n$ de manera que $\omega \in \overline{A}$. Recíprocamente si $\omega \notin A^\infty$ entonces ω se encuentra en a lo sumo un número finito de conjuntos A_n . Supongamos que A_{n_0} sea el último en el que está, es decir si $n > n_0$ entonces $\omega \notin A_n$ para todo $n > n_0$ de manera que

$$\omega \notin \bigcup_{n=n_0+1}^{\infty} A_n$$

y entonces $\omega \notin \overline{A}$.

(ii) Consideremos la sucesión de los complementos, es decir $(A_n^c)_{n \geq 1}$. Por la observación hecha anteriormente y el punto (i) se tiene que

$$\begin{aligned} \underline{A} &= (\overline{A^c})^c \\ &= \{\omega \in \Omega : \omega \text{ pertenece a infinitos } A_n^c\}^c \\ &= \{\omega \in \Omega : \omega \text{ no pertenece a infinitos } A_n^c\} \\ &= \{\omega \in \Omega : \omega \text{ pertenece a lo sumo a un número finito de conjuntos } A_n^c\} \\ &= \{\omega \in \Omega : \omega \text{ pertenece a todos a todos los } A_n \text{ salvo un número finito}\} \\ &= A_\infty. \end{aligned}$$

(iii) Se obtiene del hecho de que claramente $A_\infty \subset A^\infty$. \square

En lo que sigue $\overline{\lim}_{n \rightarrow \infty} a_n$ y $\underline{\lim}_{n \rightarrow \infty} a_n$ denotarán respectivamente el límite superior e inferior de la sucesión a_n .

Propiedad 1.17 Dada una sucesión de eventos $(A_n)_{n \geq 1}$, se tiene

$$(i) P(\overline{A}) \geq \overline{\lim}_{n \rightarrow \infty} P(A_n).$$

$$(ii) P(\underline{A}) \leq \underline{\lim}_{n \rightarrow \infty} P(A_n).$$

(iii) Se dice que existe el límite de la sucesión $(A_n)_{n \geq 1}$ de conjuntos sii $\underline{A} = \overline{A}$. En tal caso se tiene

$$P(\overline{A}) = P(\underline{A}) = \lim_{n \rightarrow \infty} P(A_n).$$

Demostración.

(i) Como lo hicimos anteriormente consideremos

$$\overline{A} = \bigcap_{k=1}^{\infty} \bigcup_{i \geq k} A_i$$

y escribamos

$$B_k = \bigcup_{i \geq k} A_i.$$

Entonces la sucesión $(B_n)_{n \geq 1}$ es decreciente y

$$\overline{A} = \bigcap_{k \geq 1} B_k.$$

Luego, como para todo $i \geq k$ se tiene $A_i \subset B_k$, podemos escribir

$$P(B_k) \geq \sup_{i \geq k} \{P(A_i)\}$$

y entonces

$$\inf_{k \geq 1} \{P(B_k)\} \geq \inf_{k \geq 1} \sup_{i \geq k} \{P(A_i)\}$$

Luego, como $P(B_k)$ es decreciente, se tiene

$$\begin{aligned} P(\overline{A}) &= \lim_{k \rightarrow \infty} P(B_k) = \inf_{k \geq 1} \{P(B_k)\} \\ &\geq \inf_{k \geq 1} \sup_{i \geq k} \{P(A_i)\} = \overline{\lim}_{i \rightarrow \infty} P(A_i). \end{aligned}$$

(ii) Se deja como ejercicio.

(iii) De (i) y (ii) tenemos que

$$P(\underline{A}) \leq \underline{\lim}_{n \rightarrow \infty} P(A_n) \leq \overline{\lim}_{n \rightarrow \infty} P(A_n) \leq P(\overline{A}).$$

Luego si $\underline{A} = \overline{A}$, resulta $P(\underline{A}) = P(\overline{A})$ y entonces

$$P(\underline{A}) = \underline{\lim}_{n \rightarrow \infty} P(A_n) = \overline{\lim}_{n \rightarrow \infty} P(A_n) = P(\overline{A}).$$

Luego $P(\underline{A}) = P(\overline{A}) = \lim_{n \rightarrow \infty} P(A_n)$. \square

1.3. σ -Álgebra generada por una familia de conjuntos.

En general no se puede tomar como σ -álgebra \mathcal{A} a $\mathcal{P}(\Omega)$ para definir el espacio de probabilidad. Esto siempre es posible si Ω es a lo sumo numerable. El siguiente teorema muestra que dada una familia \mathfrak{S} de subconjuntos de Ω , existe una menor σ -álgebra que contiene a \mathfrak{S} .

Teorema 1.1 *Dado un conjunto Ω y una familia \mathfrak{S} de subconjuntos de Ω , existe una σ -álgebra \mathcal{A}^* sobre Ω tal que (i) $\mathfrak{S} \subset \mathcal{A}^*$ y (ii) Si \mathcal{A} es otra σ -álgebra sobre Ω tal que $\mathfrak{S} \subset \mathcal{A}$, entonces $\mathcal{A}^* \subset \mathcal{A}$. Se dice entonces que \mathcal{A}^* es la σ -álgebra sobre Ω generada por \mathfrak{S} .*

Demostración. Denotaremos a la familia de todas las σ -álgebras sobre Ω que contienen a \mathfrak{S} por \mathcal{R} . Entonces

$$\mathcal{R} = \{\mathcal{A} : \mathcal{A} \text{ es una } \sigma\text{-álgebra sobre } \Omega \text{ y } \mathcal{A} \supset \mathfrak{S}\}.$$

Claramente \mathcal{R} es no vacía, ya que $\mathcal{P}(\Omega) \in \mathcal{R}$. Definamos ahora

$$\mathcal{A}^* = \bigcap_{\mathcal{A} \in \mathcal{R}} \mathcal{A}.$$

Primero mostraremos que \mathcal{A}^* es una σ -álgebra sobre Ω .

Veamos que $\Omega \in \mathcal{A}^*$. En efecto, $\Omega \in \mathcal{A}$, para toda $\mathcal{A} \in \mathcal{R}$, luego $\Omega \in \mathcal{A}^*$.

Sea ahora $A \in \mathcal{A}^*$, mostraremos que $A^c \in \mathcal{A}^*$. En efecto, como $A \in \mathcal{A}$, para toda $\mathcal{A} \in \mathcal{R}$, se tiene $A^c \in \mathcal{A}$, para toda $\mathcal{A} \in \mathcal{R}$. Luego $A^c \in \mathcal{A}^*$.

Sea una sucesión numerable de eventos $A_1, A_2, \dots, A_n, \dots$ que están en \mathcal{A}^* . Mostraremos que $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}^*$. Dado $\mathcal{A} \in \mathcal{R}$, se tiene $A_i \in \mathcal{A}$ para todo i , y luego $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ también. Luego $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, para todo $\mathcal{A} \in \mathcal{R}$ y entonces

$$\bigcup_{i=1}^{\infty} A_i \in \bigcap_{\mathcal{A} \in \mathcal{R}} \mathcal{A} = \mathcal{A}^*.$$

Esto prueba que \mathcal{A}^* es una σ -álgebra. Por otro lado si \mathcal{A} es una σ -álgebra y $\mathcal{A} \supset \mathfrak{S}$, entonces $\mathcal{A} \in \mathcal{R}$, y esto implica que $\mathcal{A}^* \subset \mathcal{A}$. \square

σ -álgebra de Borel sobre los reales. Si tenemos un espacio de probabilidad cuyo espacio muestral es el conjunto de números reales \mathbb{R} , parece natural que la σ -álgebra contenga los conjuntos de la forma $(-\infty, x]$. Esto permitirá calcular la probabilidad de que el resultado del experimento aleatorio correspondiente sea menor o igual que x . Esto motiva la siguiente definición.

Definición 1.4 *La σ -álgebra de Borel sobre \mathbb{R} , que denotaremos por \mathcal{B} , es la σ -álgebra sobre \mathbb{R} generada por los conjuntos de la forma $A_x = (-\infty, x]$, para todo $x \in \mathbb{R}$. Un conjunto $B \in \mathcal{B}$ se denomina boreliano.*

Propiedades de los borelianos.

Propiedad 1.18 *Todo intervalo $(a, b]$ es un boreliano.*

Demostración. Como

$$(a, b] = (-\infty, b] - (-\infty, a],$$

por la Propiedad 1.5 $(a, b]$ es un boreliano \square

Propiedad 1.19 *Dado $x \in \mathbb{R}$, $\{x\} \in \mathcal{B}$.*

Demostración. Para esto se observa que para todo $n \in \mathbb{N}$

$$I_n = (x - \frac{1}{n}, x] \in \mathcal{B}.$$

Puesto que

$$x - \frac{1}{n} \uparrow x$$

resulta que

$$\{x\} = \bigcap_{n=1}^{\infty} I_n \in \mathcal{B},$$

y el resultado se obtiene por las propiedades 1.18 y 1.12. \square

De las propiedades 1.18 y 1.19, se deducen inmediatamente las propiedades 1.20-1.22

Propiedad 1.20 $(a, b) = (a, b] - \{b\} \in \mathcal{B}$.

Propiedad 1.21 $[a, b] = \{a\} \cup (a, b) \in \mathcal{B}$.

Propiedad 1.22 $[a, b) = \{a\} \cup (a, b) \in \mathcal{B}$.

Propiedad 1.23 *Todo abierto es un boreliano*

Demostración. Sea $G \subset \mathbb{R}$ un abierto. Para todo $x \in G$ existe un intervalo (a_x, b_x) tal que $x \in (a_x, b_x) \subset G$ con a_x y b_x racionales. Por lo tanto G puede escribirse como la unión numerable de borelianos

$$G = \bigcup_{x \in G} (a_x, b_x),$$

y por lo tanto $G \in \mathcal{B}$. \square

Propiedad 1.24 *Todo cerrado es un boreliano*

Demostración. Sea F un cerrado. Entonces $F^c = G$ es un abierto y por Propiedad 1.23 se tiene que $F^c \in \mathcal{B}$. Ahora por ser σ -álgebra se obtiene que

$$F = (F^c)^c \in \mathcal{B}. \quad \square$$

σ -álgebra de Borel en \mathbb{R}^n .

Definición 1.5 *La σ -álgebra de Borel sobre \mathbb{R}^n es la σ -álgebra sobre \mathbb{R}^n generada por los conjuntos de la forma*

$$A_{(x_1, x_2, \dots, x_n)} = (-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n],$$

donde (x_1, \dots, x_n) es una n -upla de números reales. Será denotada por \mathcal{B}^n .

Observación. De manera análoga al caso de la σ -álgebra de Borel sobre \mathbb{R} , se pueden mostrar las propiedades 1.25-1.26 cuyas demostraciones se dejan como ejercicio.

Propiedad 1.25 *Cualquier rectángulo en \mathbb{R}^n de la forma*

$$(-a_1, b_1] \times (-a_2, b_2] \times \dots \times (-a_n, b_n]$$

$$(-a_1, b_1) \times (-a_2, b_2) \times \dots \times (-a_n, b_n)$$

$$[-a_1, b_1) \times [-a_2, b_2) \times \dots \times [-a_n, b_n)$$

es un boreliano.

Propiedad 1.26 *Todo abierto y todo cerrado en \mathbb{R}^n es un boreliano.*

1.4. Espacios de probabilidad finitos o numerables.

Definición 1.6 Sea (Ω, \mathcal{A}, P) un espacio de probabilidad con Ω a lo sumo numerable. En este caso podemos tomar como \mathcal{A} el conjunto de partes de Ω ($\mathcal{P}(\Omega)$). Definimos la función de densidad p , asociada a la probabilidad P por

$$p : \Omega \rightarrow [0, 1]$$

de la siguiente manera

$$p(\omega) = P(\{\omega\}).$$

Propiedades de la función de densidad

Propiedad 1.27 La función de densidad determina la función de probabilidad. Para todo $A \subset \Omega$ se tiene

$$P(A) = \sum_{\omega \in A} p(\omega).$$

Demostración. Si $A \subset \Omega$ entonces A se puede escribir como la siguiente unión disjunta

$$A = \bigcup_{\omega \in A} \{\omega\},$$

donde cada conjunto $\{\omega\} \in \mathcal{A}$. Luego

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p(\omega). \quad \square$$

Propiedad 1.28 Si Ω es finito o numerable se cumple que

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

Demostración. En efecto por la Propiedad 1.27

$$1 = P(\Omega) = \sum_{\omega \in \Omega} p(\omega). \quad \square$$

Definición 1.7 Decimos que un espacio finito $\Omega = \{\omega_1, \dots, \omega_n\}$ es equiprobable si

$$p(\omega_i) = p(\omega_j), \quad \forall i, j.$$

Observación. Un espacio de probabilidad infinito numerable no puede ser equiprobable. En efecto, supongamos que $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$, y $p(\omega) = c$. Luego por la Propiedad 1.27 se tendría

$$1 = \sum_{i=1}^{\infty} p(\omega_i) = \sum_{i=1}^{\infty} c,$$

lo que es un absurdo puesto que $\sum_{i=1}^{\infty} c = \infty$ ó 0 según $c > 0$ ó $c = 0$.

Propiedad 1.29 Si Ω es un espacio de probabilidad equiprobable entonces, la probabilidad de cualquier evento A se calcula por

$$P(A) = \frac{\#A}{\#\Omega},$$

donde $\#A$ denota el cardinal de A .

Demostración. Para ver esto supongamos que para todo $\omega \in \Omega$ se tenga $p(\omega) = c$, entonces

$$1 = \sum_{\omega \in \Omega} p(\omega) = \sum_{\omega \in \Omega} c = c \sum_{\omega \in \Omega} 1 = c \#\Omega,$$

y luego,

$$c = \frac{1}{\#\Omega}.$$

Además

$$P(A) = \sum_{w \in A} p(w) = \sum_{w \in A} c = c \sum_{w \in A} 1 = c(\#A) = \frac{\#A}{\#\Omega}.$$

Ejemplo 1.7 Hallar la probabilidad de que dado un conjunto de n personas, dos personas cumplan años el mismo día. Se supondrá que todos los años tienen 365 días y que las probabilidades de nacimiento en cualquier fecha son iguales.

Supongamos que a cada persona se le asigna un número entre 1 y n y sea x_i el día del cumpleaños de la persona i . Luego $1 \leq x_i \leq 365$, y podemos considerar el siguiente espacio muestral

$$\Omega = \{(x_1, x_2, \dots, x_n) : x_i \in \mathbb{N} : 1 \leq x_i \leq 365\}.$$

donde \mathbb{N} es el conjunto de números naturales.

En vez de calcular la probabilidad de que dos personas cumplan el mismo día, calculemos la del complemento, es decir la probabilidad de que todas cumplan años en días distintos

$$A^c = \{(x_1, x_2, \dots, x_n) : 1 \leq x_i \leq 365, x_i \neq x_j \forall i \neq j\}.$$

Se tiene

$$\#\Omega = 365^n$$

Además

$$\#A^c = \binom{365}{n} n!.$$

La importancia de la combinatoria se ve en este punto; es necesario contar con principios de enumeración. En este caso, primero seleccionamos los n días distintos entre los 365 días posibles y luego por cada muestra se obtienen $n!$ formas distintas de distribuirlos entre n personas.

Las probabilidades que se obtienen usando esta fórmula pueden contradecir la intuición. Por ejemplo, si $n = 20$, $P(A) \approx 0,41$, si $n = 30$, $P(A) \approx 0,76$ y si $n = 40$, $P(A) \approx 0,89$.

1.5. Probabilidad condicional.

Sea (Ω, \mathcal{A}, P) un espacio de probabilidad, y consideremos dos eventos $A, B \in \mathcal{A}$, y supongamos que $P(B) \neq 0$.

Queremos estudiar como cambia la probabilidad de ocurrencia de A cuando se conoce que otro evento B ha ocurrido. En este caso habrá que redefinir el espacio muestral considerando solamente los elementos de B como posibles resultados.

Por ejemplo, consideremos el experimento de “tirar un dado” y preguntémosnos acerca de la probabilidad de que salga un seis, sabiendo que el dado escogido es un número par. En este caso la probabilidad no es $1/6$, puesto que tenemos la certeza de que el resultado está en el conjunto $\{2, 4, 6\}$. Como cada uno de estos tres resultados tienen idéntica probabilidad, como se verá, la probabilidad de obtener el 6 sabiendo que el resultado es par será $1/3$.

Vamos a tratar de determinar cual debe ser la probabilidad de un evento A condicional a que se conoce que B ha ocurrido, utilizando interpretación heurística de la probabilidad como límite de la frecuencia con la cual un evento ocurre. Para esto supongamos que se han hecho n repeticiones independientes del experimento y denotemos con

n_B : el número de veces en el que ocurre el resultado B ,

$n_{A \cap B}$: el número de veces en el que ocurre el resultado $A \cap B$.

Heurísticamente la probabilidad condicional de A dado B , será el límite de la frecuencia con la cual A ocurre en los experimentos donde B ocurre, es decir el límite de

$$\frac{n_{A \cap B}}{n_B}.$$

Luego, la “probabilidad de que ocurra A condicional B ” será

$$\lim_{n \rightarrow \infty} \frac{n_{A \cap B}}{n_B} = \lim_{n \rightarrow \infty} \frac{\frac{n_{A \cap B}}{n}}{\frac{n_B}{n}} = \frac{\lim_{n \rightarrow \infty} \frac{n_{A \cap B}}{n}}{\lim_{n \rightarrow \infty} \frac{n_B}{n}} = \frac{P(A \cap B)}{P(B)}.$$

Esto justifica la siguiente definición.

Definición 1.8 Sea (Ω, \mathcal{A}, P) un espacio de probabilidad $A, B \in \mathcal{A}$ tal que $P(B) > 0$. Se define la probabilidad condicional de A dado B por

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

El siguiente teorema muestra que para cada B fijo, $P(\cdot|B)$ es una función de probabilidad.

Teorema 1.2 Fijado el evento $B \in \Omega$, tal que $P(B) > 0$, definamos $\tilde{P} : \mathcal{A} \rightarrow [0, 1]$ por

$$\tilde{P}(A) = P(A|B)$$

para todo $A \in \mathcal{A}$. Luego \tilde{P} es una probabilidad.

Demostración.

(i)

$$\tilde{P}(\Omega) = P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

(ii) Sea $(A_n)_{n \geq 1}$, una sucesión de eventos disjuntos dos a dos, es decir si $i \neq j$, entonces $A_i \cap A_j = \emptyset$. Luego

$$\begin{aligned} \tilde{P}\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} A_n | B\right) = \frac{P\left(\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B\right)}{P(B)} = \\ &= \frac{P\left(\bigcup_{n=1}^{\infty} A_n \cap B\right)}{P(B)} = \frac{\sum_{n=1}^{\infty} P(A_n \cap B)}{P(B)} = \\ &= \sum_{n=1}^{\infty} \frac{P(A_n \cap B)}{P(B)} = \sum_{n=1}^{\infty} P(A_n|B) = \sum_{n=1}^{\infty} \tilde{P}(A_n). \quad \square \end{aligned}$$

1.6. Independencia de eventos.

Definición 1.9 Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y consideremos $A, B \in \mathcal{A}$. Se dice que A y B son independientes si

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

Propiedad 1.30 (i) Si $P(B) > 0$, entonces A y B son independientes si y sólo si $P(A|B) = P(A)$.

(ii) Si $P(B) = 0$, dado cualquier $A \in \mathcal{A}$ se tiene que A y B son independientes.

Demostración. La demostración es inmediata. \square

La propiedad de independencia se generaliza para un número finito de eventos.

Definición 1.10 Se dice que los eventos A_1, \dots, A_k son independientes si para cualquier sucesión de subíndices (i_1, \dots, i_h) , $h \leq k$, con $i_r \neq i_s$ si $r \neq s$ se tiene que

$$P\left(\bigcap_{j=1}^h A_{i_j}\right) = \prod_{j=1}^h P(A_{i_j}).$$

Observaciones.

1. Para que tres eventos A_1, A_2 y A_3 sean independientes se deben cumplir las siguientes igualdades

$$\begin{aligned}P(A_1 \cap A_2) &= P(A_1)P(A_2) \\P(A_1 \cap A_3) &= P(A_1)P(A_3) \\P(A_2 \cap A_3) &= P(A_2)P(A_3) \\P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2)P(A_3).\end{aligned}$$

2. No alcanza la independencia tomados de a dos. Como ejemplo tomemos $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ espacio de probabilidad equiprobable, es decir $P(\{\omega_i\}) = \frac{1}{4}$. Entonces los conjuntos

$$\begin{aligned}A_1 &= \{\omega_1, \omega_2\} \\A_2 &= \{\omega_1, \omega_3\} \\A_3 &= \{\omega_2, \omega_3\}\end{aligned}$$

son independientes tomados de a dos pero no en forma conjunta. Más precisamente, se cumple que

$$\forall j : P(A_j) = \frac{1}{2}$$

$$A_i \cap A_j = \{\omega_k\} \text{ para algún } k$$

y luego

$$P(A_i \cap A_j) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_i) P(A_j).$$

Pero

$$A_1 \cap A_2 \cap A_3 = \emptyset,$$

y por lo tanto

$$0 = P(A_1 \cap A_2 \cap A_3) \neq P(A_1) P(A_2) P(A_3) = \frac{1}{8}.$$

Teorema 1.3 A_1, \dots, A_k son eventos independientes si y sólo si para cualquier sucesión $(i_1, \dots, i_h), h \leq k$, con $i_r \neq i_s$ si $r \neq s$ y tal que

$$P\left(\bigcap_{j=2}^h A_{i_j}\right) > 0,$$

se tiene que

$$P\left(A_{i_1} \left| \bigcap_{j=2}^h A_{i_j} \right.\right) = P(A_{i_1}). \quad (1.1)$$

Demostración. Supongamos primero que A_1, \dots, A_k son independientes y demostraremos que se cumple (1.1). Sean $A_{i_1}, A_{i_2}, \dots, A_{i_h}$ tales que $i_r \neq i_s$ si $r \neq s$ y $P\left(\bigcap_{j=2}^h A_{i_j}\right) > 0$. Entonces

$$P\left(A_{i_1} \left| \bigcap_{j=2}^h A_{i_j} \right.\right) = \frac{P\left(\bigcap_{j=1}^h A_{i_j}\right)}{P\left(\bigcap_{j=2}^h A_{i_j}\right)} = \frac{\prod_{j=1}^h P(A_{i_j})}{\prod_{j=2}^h P(A_{i_j})} = P(A_{i_1}).$$

Supongamos ahora que A_1, \dots, A_k son eventos que satisfacen la propiedad del enunciado. Queremos probar que entonces son independientes, es decir que

$$P\left(\bigcap_{j=1}^h A_{i_j}\right) = \prod_{j=1}^h P(A_{i_j}). \quad (1.2)$$

Lo probaremos por inducción sobre h . Comenzaremos con $h = 2$. Dados A_{i_1} y A_{i_2} con $i_1 \neq i_2$, puede suceder que (a) $P(A_{i_2}) = 0$ o que (b) $P(A_{i_2}) > 0$. En el caso (a) se tiene que como $A_{i_1} \cap A_{i_2} \subset A_{i_2}$, resulta $P(A_{i_1} \cap A_{i_2}) = 0$ y luego

$$P(A_{i_1} \cap A_{i_2}) = P(A_{i_1})P(A_{i_2}) \quad (1.3)$$

En el caso (b) como vale (1.1) se tiene

$$P(A_{i_1}|A_{i_2}) = \frac{P(A_{i_1} \cap A_{i_2})}{P(A_{i_2})} = P(A_{i_1})$$

y luego también vale

$$P(A_{i_1} \cap A_{i_2}) = 0 = P(A_{i_1})P(A_{i_2}).$$

Esto muestra que (1.2) vale para $h = 2$.

Supongamos ahora que (1.2) vale para h y probemos que también vale para $h + 1$. Elegimos $A_{i_1}, A_{i_2}, \dots, A_{i_h}, A_{i_{h+1}}$ eventos. Consideramos dos casos

- (a) Supongamos que $P\left(\bigcap_{j=2}^{h+1} A_{i_j}\right) = 0$. En tal caso por la suposición que (1.2) vale para h conjuntos se tiene que

$$0 = P\left(\bigcap_{j=2}^{h+1} A_{i_j}\right) = \prod_{j=2}^{h+1} P(A_{i_j}).$$

Luego

$$\prod_{j=1}^{h+1} P(A_{i_j}) = 0, \quad (1.4)$$

y como $\bigcap_{j=1}^{h+1} A_{i_j} \subset \bigcap_{j=2}^{h+1} A_{i_j}$ se tendrá que

$$P\left(\bigcap_{j=1}^{h+1} A_{i_j}\right) = 0. \quad (1.5)$$

De (1.4) y (1.5) obtenemos que

$$P\left(\bigcap_{j=1}^{h+1} A_{i_j}\right) = \prod_{j=1}^{h+1} P(A_{i_j}).$$

- (b) Supongamos ahora que $P\left(\bigcap_{j=2}^{h+1} A_{i_j}\right) > 0$. Entonces como estamos suponiendo que (1.1) vale se tiene

$$P\left(A_{i_1} \left| \bigcap_{j=2}^{h+1} A_{i_j} \right.\right) = P(A_{i_1}),$$

y luego

$$\frac{P\left(\bigcap_{j=1}^{h+1} A_{i_j}\right)}{P\left(\bigcap_{j=2}^{h+1} A_{i_j}\right)} = P(A_{i_1}).$$

Equivalentemente

$$P\left(\bigcap_{j=1}^{h+1} A_{i_j}\right) = P(A_{i_1}) P\left(\bigcap_{j=2}^{h+1} A_{i_j}\right),$$

y como por la hipótesis inductiva (1.2) vale para h , se deduce

$$P\left(\bigcap_{j=1}^{h+1} A_{i_j}\right) = P(A_{i_1}) \prod_{j=2}^{h+1} P(A_{i_j}) = \prod_{j=1}^{h+1} P(A_{i_j}). \quad \square$$

Definición 1.11 Sea I un conjunto finito o numerable, una sucesión $\{A_i\}_{i \in I}$ se dice una partición de Ω sii

1.

$$\bigcup_{i \in I} A_i = \Omega$$

2. Si $i \neq j$ entonces

$$A_i \cap A_j = \emptyset$$

Teorema 1.4 (Teorema de la Probabilidad Total) Sea (Ω, \mathcal{A}, P) un espacio de probabilidad, $\{A_n\}_{n \in I} \subset \mathcal{A}$ una partición de Ω con $P(A_i) > 0$, para todo $i \in I$ y $B \in \mathcal{A}$ tal que $P(B) > 0$. Entonces

$$P(B) = \sum_{i \in I} P(A_i) P(B|A_i)$$

Demostración. Como B se puede escribir como la siguiente unión disjunta

$$B = \bigcup_{i \in I} (B \cap A_i),$$

entonces como $P(B|A_i) = P(B \cap A_i)/P(A_i)$, se tiene $P(B \cap A_i) = P(A_i)P(B|A_i)$ y por lo tanto

$$P(B) = \sum_{i \in I} P(A_i) P(B|A_i). \quad \square$$

Teorema 1.5 (Bayes) Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y $\{A_i\}_{1 \leq i \leq k} \subset \mathcal{A}$ una partición de Ω con $P(A_i) > 0$, $1 \leq i \leq k$. Sea $B \in \mathcal{A}$ con $P(B) > 0$. Supongamos conocidas a priori las probabilidades $P(B|A_i)$ y $P(A_i)$ para todo i . Entonces

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^k P(A_j)P(B|A_j)}.$$

Demostración. Usando el teorema de la probabilidad total teniendo en cuenta que $\{A_j\}_{1 \leq j \leq k}$ es una partición y aplicando la definición de probabilidad condicional y el Teorema 1.4 se obtiene

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} \\ &= \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^k P(A_j)P(B|A_j)}. \quad \square \end{aligned}$$

Ejemplo de aplicación del Teorema de Bayes.

Consideremos un test que detecta pacientes enfermos de un tipo específico de enfermedad. La detección corresponde a que el test de positivo. El resultado de un test negativo se interpreta como no detección de enfermedad.

Sea

A_1 : el evento “el paciente seleccionado no tiene la enfermedad”

A_2 : el evento “el paciente seleccionado tiene la enfermedad ”

Entonces $\{A_1, A_2\}$ constituye una partición del espacio de probabilidad

Consideremos además

T_+ : el evento “el test da positivo”

T_- : el evento “el test da negativo”

Supongamos conocidas las probabilidades de ser sano o enfermo antes de hacer el test (probabilidades apriori).

$$P(A_1) = 0,99; P(A_2) = 0,01.$$

Ademas supongamos que

$$P(T_+|A_1) = 0,01; P(T_+|A_2) = 0,99.$$

Observemos que para un test perfecto se pediría

$$P(T_+|A_1) = 0; P(T_+|A_2) = 1.$$

Es decir, estamos suponiendo que el test no es perfecto.

Calculemos la probabilidad de que dado que el test detecta enfermedad el paciente sea efectivamente enfermo (esta probabilidad se denomina probabilidad a posteriori). De acuerdo al Teorema de Bayes se tiene

$$P(A_2|T_+) = \frac{P(A_2)P(T_+|A_2)}{P(A_1)P(T_+|A_1) + P(A_2)P(T_+|A_2)} = 0,5.$$

y

$$P(A_1|T_+) = 1 - P(A_2|T_+) = 0,5$$

La conclusión es que si el test da positivo, no hay una evidencia fuerte de que el paciente esté enfermo o sano ya que ambas probabilidades condicionales son iguales a 0.50. Luego un test como el descrito no es útil para detectar la enfermedad.

Si logramos tener

$$P(T_+|A_1) = 0,001; \quad P(T_+|A_2) = 0,999$$

la situación cambia; en tal caso resulta $P(A_2|T_+) = 0,91$, que es más aceptable que la anterior.

Capítulo 2

Variable Aleatoria.

2.1. Concepto de variable aleatoria.

En muchos casos interesa conocer solamente alguna característica numérica del resultado del experimento aleatorio. Demos dos ejemplos:

1. El experimento consiste en tirar dos dados y los posibles resultados son $\Omega = \{ (x, y) : x \in I_6, y \in I_6 \}$ donde $I_k = \{1, 2, \dots, k\}$ y para cada resultado (x, y) interesa solo la suma de los dados $x + y$.
2. El experimento consiste en un tiro al blanco y el conjunto de los resultados es $\Omega = \{ (x, y) : x \in \mathbb{R}, y \in \mathbb{R} \}$, x e y son la abcisa y ordenada del punto donde pegó el tiró tomando origen $(0, 0)$ el punto correspondiente al blanco. En este ejemplo solo interesa la distancia al blanco, es decir $(x^2 + y^2)^{1/2}$

Definición 2.1 Sea (Ω, \mathcal{A}, P) un espacio de probabilidad. Una variable aleatoria es una función $X : \Omega \rightarrow \mathbb{R}$ tal que para todo $x \in \mathbb{R}$

$$X^{-1}((-\infty, x]) \in \mathcal{A}. \tag{2.1}$$

Observaciones.

1. La condición (2.1) permite calcular

$$P(\{\omega : X(\omega) \leq x\}) = P(X^{-1}((-\infty, x])).$$

2. El concepto de variable aleatoria es esencialmente el mismo que el de función medible en teoría de la medida. Si $(\Omega, \mathcal{A}, \mu)$ es un espacio de medida $f : \mathcal{A} \rightarrow \mathbb{R}$ se dice medible sii para todo x vale que $f^{-1}((-\infty, x]) \in \mathcal{A}$.

3. Si \mathcal{A} es el conjunto de partes de Ω , como es usual cuando Ω es finito o numerable, la condición (2.1) se cumple trivialmente.

Teorema 2.1 *Sea X una variable aleatoria sobre un espacio de probabilidad (Ω, \mathcal{A}, P) . Entonces vale que $X^{-1}(B) \in \mathcal{A}$ para todo $B \in \mathcal{B}$. (\mathcal{B} es el conjunto de borelianos en \mathbb{R}).*

Demostración. Como por definición $X^{-1}((-\infty, x]) \in \mathcal{A}$, basta con verificar que

$$\Phi = \{A \subset \mathbb{R} : X^{-1}(A) \in \mathcal{A}\}$$

es una σ -álgebra. Si esto es cierto se tendrá que $\mathcal{B} \subset \Phi$, puesto que la σ -álgebra de Borel es la más chica que contiene a las semirectas. Veamos que esto es cierto.

- (a) $\mathbb{R} \in \Phi$ pues

$$X^{-1}(\mathbb{R}) = \Omega \in \mathcal{A}.$$

- (b) Si $A \in \Phi$, entonces $A^c \in \Phi$. Como $X^{-1}(A) \in \mathcal{A}$, se tendrá que

$$X^{-1}(A^c) = [X^{-1}(A)]^c \in \mathcal{A}.$$

- (c) Sea $\{A_n\}_{n \in \mathbb{N}} \subset \Phi$. Luego $X^{-1}(A_n) \in \mathcal{A}$ para todo n y como \mathcal{A} es un σ -álgebra se tendrá que

$$\bigcup_{n \in \mathbb{N}} X^{-1}(A_n) \in \mathcal{A}.$$

Luego

$$X^{-1}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \bigcup_{n \in \mathbb{N}} X^{-1}(A_n) \in \mathcal{A}.$$

- (a), (b) y (c) prueban que Φ es una σ -álgebra. \square

2.2. Espacio de probabilidad asociado a una variable aleatoria.

Sea un espacio de probabilidad (Ω, \mathcal{A}, P) y sea $X : \Omega \rightarrow \mathbb{R}$ una variable aleatoria. Asociada a esta variable podemos definir un nuevo espacio de probabilidad $(\mathbb{R}, \mathcal{B}, P_X)$ donde para todo $B \in \mathcal{B}$ se define

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

Obsérvese que $P(X^{-1}(B))$ está definido ya que $X^{-1}(B)$ está en \mathcal{A} . Vamos a mostrar que P_X es efectivamente una probabilidad. La función P_X se denomina *probabilidad inducida por X* o *distribución de X* .

Si a uno le interesa sólo el resultado de la variable aleatoria, esto permite trabajar en un espacio de probabilidad donde el espacio muestral es \mathbb{R} y la σ -álgebra es \mathcal{B} , la σ -álgebra de Borel.

Teorema 2.2 P_X es efectivamente una función de probabilidad.

Demostración.

(a)

$$P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1.$$

(b) Si $\{B_i\}_{i \in \mathbb{N}} \subset \mathcal{B}$ es una sucesión disjunta dos a dos, entonces $\{X^{-1}(B_i)\}_{i \in \mathbb{N}}$ también lo es. Luego

$$\begin{aligned} P_X\left(\bigcup_{i \in \mathbb{N}} B_i\right) &= P\left(X^{-1}\left(\bigcup_{i \in \mathbb{N}} B_i\right)\right) = P\left(\bigcup_{i \in \mathbb{N}} X^{-1}(B_i)\right) = \\ &= \sum_{i \in \mathbb{N}} P(X^{-1}(B_i)) = \sum_{i \in \mathbb{N}} P_X(B_i). \quad \square \end{aligned}$$

Definiremos el concepto de función medible

Definición 2.2 Una función $g : \mathbb{R} \rightarrow \mathbb{R}$, se dice medible Borel sii para todo $x \in \mathbb{R}$

$$g^{-1}((-\infty, x]) \in \mathcal{B}.$$

Observaciones.

1. Trabajaremos en este curso con funciones medibles Borel, de manera que a veces nos referiremos a ellas simplemente con el nombre de medibles.
2. Si $B \in \mathcal{B}$ resultará $g^{-1}(B) \in \mathcal{B}$. Este resultado se demuestra como el análogo para variables aleatorias.
3. Considerando un espacio de probabilidad con $\Omega = \mathbb{R}$ y $\mathcal{A} = \mathcal{B}$ es inmediato que g es medible Borel es equivalente a que g es una variable aleatoria.

Ejercicio. Demostrar los siguientes resultados:

Propiedad 2.1 Si $g : \mathbb{R} \rightarrow \mathbb{R}$ es continua entonces g es medible.

Propiedad 2.2 Si $g : \mathbb{R} \rightarrow \mathbb{R}$ es monótona entonces g es medible.

Propiedad 2.3 Si B es boreliano, su función característica I_B es medible.

Propiedad 2.4 Sea $\{f_n\}_{n \geq 1}$ es una sucesión de funciones medibles. Entonces

(i) Las siguientes funciones son medibles

$$f(x) = \inf_{n \in \mathbb{N}} \{f_n(x)\},$$

$$f(x) = \sup_{n \in \mathbb{N}} \{f_n(x)\}.$$

1. También son medibles

$$f(x) = \overline{\lim}_{n \rightarrow \infty} f_n(x),$$

$$f(x) = \underline{\lim}_{n \rightarrow \infty} f_n(x).$$

En particular si existe el límite puntual

$$f(x) = \lim_{n \rightarrow \infty} f_n(x)$$

es medible.

El siguiente teorema muestra que la composición de una variable aleatoria con una función medible es una variable aleatoria.

Teorema 2.3 Si $g : \mathbb{R} \rightarrow \mathbb{R}$ es medible y $X : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria, entonces $g(X) : \Omega \rightarrow \mathbb{R}$ es también una variable aleatoria.

Demostración. Basta con observar que dado $B \in \mathcal{B}$

$$[g(X)]^{-1}(B) = X^{-1}(g^{-1}(B))$$

Como $C = g^{-1}(B) \in \mathcal{B}$, resulta que también $X^{-1}(g^{-1}(B)) \in \mathcal{B}$. \square

Como consecuencia de este teorema si g es continua y X es una variable aleatoria resulta que $g(X)$ también es una variable aleatoria. Por ejemplo si X es una variable aleatoria, entonces $\text{seno}(X)$, $\text{coseno}(X)$, a^X , con a constante son variables aleatorias.

Teorema 2.4 Si X, Y son variables aleatorias entonces

(i) $X + Y$, $X Y$ son variables aleatorias.

(ii) Si $P(Y \neq 0) = 1$ entonces X/Y es una variable aleatoria.

Demostración. Las demostraciones de (i) y (ii) se verán más adelante.

2.3. Función de distribución de una variable aleatoria.

Definición 2.3 Sea X una variable aleatoria. Se define la función de distribución asociada a X como la función $F_X : \mathbb{R} \rightarrow [0, 1]$ dada por

$$F_X(x) = P_X((-\infty, x]) = P(X^{-1}((-\infty, x])).$$

Observación. Como veremos, la importancia de F_X es que caracteriza la distribución de X . Es decir F_X determina el valor de $P_X(B)$ para todo $B \in \mathcal{B}$

Propiedades de la función de distribución.

Las cuatro propiedades que probaremos en el Teorema 2.5 van a caracterizar a las funciones de distribución.

Teorema 2.5 Sea X una variable aleatoria sobre (Ω, \mathcal{A}, P) y sea F_X su función de distribución. Entonces se tiene

1. F_X es monótona no decreciente, es decir $x_1 < x_2$ implica $F_X(x_1) \leq F_X(x_2)$.
2. $\lim_{x \rightarrow \infty} F_X(x) = 1$.
3. $\lim_{x \rightarrow -\infty} F_X(x) = 0$.
4. F_X es continua a derecha en todo punto de \mathbb{R} .

Demostración.

1. Si $x < x'$ entonces

$$(-\infty, x] \subset (-\infty, x'],$$

y por lo tanto

$$F_X(x) = P((-\infty, x]) \leq P((-\infty, x']) = F_X(x').$$

2. En primer lugar veamos que

$$\lim_{n \rightarrow \infty} F_X(n) = 1.$$

Consideremos la sucesión monótona creciente de conjuntos

$$A_n = (-\infty, n], \quad n \in \mathbb{N}.$$

Entonces

$$\bigcup_{n \in \mathbb{N}} A_n = \mathbb{R}.$$

Luego de acuerdo con la propiedad para sucesiones crecientes de eventos

$$\lim_{n \rightarrow \infty} F_X(n) = \lim_{n \rightarrow \infty} P_X(A_n) = P_X\left(\bigcup_{n \in \mathbb{N}} A_n\right) = P_X(\mathbb{R}) = 1.$$

Ahora veamos que efectivamente $\lim_{n \rightarrow \infty} F_X(x) = 1$, esto es para todo $\varepsilon > 0$ existe $x_0 > 0$ tal que si $x > x_0$ entonces se cumple $|F_X(x) - 1| < \varepsilon$. O equivalentemente

$$1 - \varepsilon < F_X(x) < 1 + \varepsilon.$$

Por $0 \leq F_X(x) \leq 1$, se cumple que para cualquier $\varepsilon > 0$, $F_X(x) < \varepsilon + 1$. Por lo tanto sólo tenemos que mostrar que existe $x_0 > 0$ tal que si $x > x_0$ entonces se cumple

$$1 - \varepsilon < F_X(x).$$

Sabemos que dado $\varepsilon > 0$ existe un $n_0 \in \mathbb{N}$ tal que si $n > n_0$ entonces

$$1 - \varepsilon < F_X(n).$$

Tomando $x_0 = n_0$ y teniendo en cuenta la monotonía de F_X , se tendrá que si $x > x_0$ entonces

$$1 - \varepsilon < F_X(n_0) \leq F_X(x).$$

3. Se demuestra de manera similar a (2). En primer lugar se prueba que

$$\lim_{n \rightarrow \infty} F_X(-n) = 0.$$

Luego se considera la sucesión monótona decreciente que converge a \emptyset

$$A_n = (-\infty, -n],$$

y se obtiene

$$\lim_{n \rightarrow \infty} P_X(A_n) = 0.$$

Luego se procede como en (2).

4. Queremos ver que F_X es continua a derecha en cualquier punto $x_0 \in \mathbb{R}$. Es decir, dado $\varepsilon > 0$ existe $\delta > 0$ tal que si

$$0 < x - x_0 < \delta$$

entonces

$$F_X(x_0) - \varepsilon \leq F_X(x) \leq F_X(x_0) + \varepsilon.$$

La primer inecuación es válida siempre ya que como $x_0 < x$ entonces $F_X(x_0) - \varepsilon \leq F_X(x_0) \leq F_X(x)$. Basta entonces probar que $F_X(x) \leq F_X(x_0) + \varepsilon$. Consideremos la sucesión decreciente de conjuntos

$$A_n = \left(-\infty, x_0 + \frac{1}{n} \right]$$

que satisface

$$\bigcap_{n \in \mathbb{N}} A_n = (-\infty, x_0].$$

Entonces

$$\begin{aligned} \lim_{n \rightarrow \infty} F_X \left(x_0 + \frac{1}{n} \right) &= \lim_{n \rightarrow \infty} P_X(A_n) = P_X \left(\bigcap_{n \in \mathbb{N}} A_n \right) \\ &= P_X((-\infty, x_0]) = F_X(x_0) \end{aligned}$$

Luego existe $n_0 \in \mathbb{N}$ tal que si $n > n_0$ entonces

$$F_X \left(x_0 + \frac{1}{n} \right) \leq F_X(x_0) + \varepsilon$$

Si tomamos $\delta < 1/n_0$, entonces para todo x tal que $0 < x - x_0 < \delta$ se tendrá

$$F_X(x) \leq F_X(x_0 + \delta) \leq F_X \left(x_0 + \frac{1}{n_0} \right) \leq F_X(x_0) + \varepsilon. \square$$

Dada una función $g : \mathbb{R} \rightarrow \mathbb{R}$, denotemos por $\lim_{x \rightarrow x_0^-} g(x)$ el límite de $g(x)$ cuando x tiende a x_0 por la izquierda. Entonces tenemos la siguiente propiedad de la función de distribución.

Propiedad 2.5 *Para todo $x_0 \in \mathbb{R}$ se tiene que*

$$\lim_{x \rightarrow x_0^-} F_X(x) = F_X(x_0) - P_X(\{x_0\}).$$

Demostración. Sea $a = F_X(x_0) - P_X(\{x_0\})$. Tenemos que mostrar que dado $\varepsilon > 0$ existe $\delta > 0$ tal que si $x_0 - \delta < x < x_0$, entonces

$$a - \varepsilon \leq F_X(x) \leq a + \varepsilon. \quad (2.2)$$

Tenemos que

$$a = P_X((-\infty, x_0]) - P_X(\{x_0\}) = P_X((-\infty, x_0)).$$

Como $x_0 - \varepsilon < x < x_0$ implica que $(-\infty, x] \subset (-\infty, x_0)$, se tendrá que

$$F_X(x) = P_X((-\infty, x]) \leq P_X((-\infty, x_0)) = a.$$

Luego, para probar (2.2) bastará probar que $x_0 - \delta < x < x_0$ implica

$$a - \varepsilon \leq F_X(x). \quad (2.3)$$

Como la sucesión de intervalos $A_n = (-\infty, x_0 - 1/n]$ es creciente y

$$\bigcup_{n \in \mathbb{N}} A_n = (-\infty, x_0),$$

se tendrá

$$\begin{aligned} \lim_{n \rightarrow \infty} F_X(x_0 - 1/n) &= \lim_{n \rightarrow \infty} P_X(A_n) = P_X((-\infty, x_0)) \\ &= a. \end{aligned}$$

Luego existe n_0 tal que $F_X(x_0 - 1/n_0) \geq a - \varepsilon$. Sea $\delta = 1/n_0$ y tomemos $x_0 - \delta < x < x_0$. Por la monotonía de F_X se tendrá

$$a - \varepsilon \leq F_X(x_0 - 1/n_0) = F_X(x_0 - \delta) \leq F_X(x),$$

y por lo tanto (2.3) se cumple. Esto prueba la Propiedad 2.5. \square

Propiedad 2.6 F_X es continua a izquierda en x_0 si y sólo si $P_X(\{x_0\}) = 0$.

Demostración. El resultado es inmediato a partir de la Propiedad 2.5. \square

Demostración.

Teorema 2.6 Sea F_X la función de distribución de una v.a X . Entonces el conjunto de puntos de discontinuidad de F_X es a lo sumo numerable.

Demostración. De acuerdo a la Propiedad 2.6, el conjunto de puntos de discontinuidad está dado por

$$A = \{x : P_X(\{x\}) > 0\}.$$

Para todo $k \in \mathbb{N}$ sea

$$A_k = \left\{ x : P_X(\{x\}) > \frac{1}{k} \right\}.$$

Entonces es fácil mostrar que

$$\bigcup_{k=1}^{\infty} A_k = A.$$

Luego para demostrar el teorema bastará probar que para $k \in \mathbb{N}$ se tiene que $\#A_k < \infty$. En efecto, supongamos que para algún k_0 existen infinitos puntos $\{x_n\}_{n \geq 1}$ tal que para todo $n \in \mathbb{N}$ se cumpla

$$P_X(\{x_n\}) > \frac{1}{k_0}.$$

Entonces si

$$B = \bigcup_{i \in \mathbb{N}} \{x_i\}$$

se tendrá

$$P_X(B) = \sum_{i=1}^{\infty} P_X(\{x_i\}) > \sum_{i=1}^{\infty} \frac{1}{k_0} = \infty,$$

lo que es un absurdo. \square

Veremos ahora que toda función con las cuatro propiedades del Teorema 2.5 es una función de distribución para cierta variable aleatoria X (no única). Para eso se requiere el siguiente teorema que daremos sin demostración.

Teorema 2.7 (de Extensión) *Sea $F : \mathbb{R} \rightarrow [0, 1]$ una función con las cuatro propiedades del Teorema 2.5. Luego existe una única probabilidad P sobre $(\mathbb{R}, \mathcal{B})$ tal que para todo $x \in \mathbb{R}$ se tiene*

$$P((-\infty, x]) = F(x).$$

Este Teorema no se demostrará en este curso ya que requiere teoría de la medida. La la probabilidad P se denomina extensión de la función F .

Veremos ahora algunas consecuencias del Teorema de Extensión.

Corolario 2.1 *Si X y X^* son variables aleatorias tales que $F_X = F_{X^*}$. Entonces para todo $B \in \mathcal{B}$ se tendrá*

$$P_X(B) = P_{X^*}(B).$$

Demostración. Es consecuencia de la unicidad del teorema de extensión. \square

Corolario 2.2 *Si F satisface las cuatro propiedades del Teorema 2.5, entonces existe una variable aleatoria X (no necesariamente única) tal que $F = F_X$.*

Demostración. De acuerdo al teorema de extensión se puede definir un espacio de probabilidad $(\mathbb{R}, \mathcal{B}, P)$ de forma tal que para todo $x \in \mathbb{R}$

$$F(x) = P((-\infty, x]).$$

Ahora consideramos la función identidad $X : \mathbb{R} \rightarrow \mathbb{R}$ definida como $X(x) = x$ para todo $x \in \mathbb{R}$. Entonces se cumple que

$$F_X(x) = P_X((-\infty, x]) = P(X^{-1}((-\infty, x])) = P((-\infty, x]) = F(x). \square$$

Capítulo 3

Variables aleatorias discretas y continuas.

Existen varios tipos de variables aleatorias. En este curso sólo estudiaremos con detalle las discretas y las (absolutamente) continuas.

3.1. Variables aleatorias discretas.

Definición 3.1 *Se dice que una v.a. X es discreta si existe $A \subset \mathbb{R}$ finito o numerable tal que $P_X(A) = 1$.*

Observación. Ese conjunto A no tiene que ser único. Si se le agrega un conjunto finito o numerable de probabilidad cero, seguirá teniendo esta propiedad. A continuación vamos a encontrar el conjunto más chico que tiene esta propiedad.

Definición 3.2 *Sea X una variable aleatoria discreta. Se define el rango de X como el conjunto de los puntos de discontinuidad de la función de distribución, es decir por*

$$R_X = \{x \in \mathbb{R} : P_X(\{x\}) > 0\}.$$

Teorema 3.1 *Sea X una variable aleatoria discreta. Luego (i) $P_X(R_X) = 1$, (ii) Si $P_X(A) = 1$, entonces $R_X \subset A$.*

Demostración.

- (i) Sea A un conjunto a lo sumo numerable tal que $P_X(A) = 1$. Luego A se puede escribir como la siguiente unión disjunta

$$A = (A \cap R_X) \cup (A - R_X).$$

Entonces

$$\begin{aligned}
 1 &= P_X(A) \\
 &= P_X((A \cap R_X) \cup (A - R_X)) \\
 &= P_X(A \cap R_X) + P_X(A - R_X). \tag{3.1}
 \end{aligned}$$

Luego basta probar que

$$P_X(A - R_X) = 0. \tag{3.2}$$

El conjunto $A - R_X$ es finito o infinito numerable. Además para todo $x \in A - R_X$ se tiene que $P_X(\{x\}) = 0$. Luego, como

$$A - R_X = \bigcup_{x \in A - R_X} \{x\},$$

resulta que

$$P_X(A - R_X) = \sum_{x \in P_X(A - R_X)} P_X(\{x\}) = 0.$$

Luego hemos demostrado (3.2). Luego por (3.1) se tiene $P_X(A \cap R_X) = 1$, y luego también $P(R_X) = 1$.

- (ii) Sea un conjunto A numerable tal que $P_X(A) = 1$. Supongamos que exista $x_0 \in R_X$ tal que $x_0 \notin A$ entonces consideramos $\tilde{A} = A \cup \{x_0\}$ y se obtiene que

$$P_X(\tilde{A}) = P_X(A) + P_X(\{x_0\}) > P_X(A) = 1,$$

lo cual es un absurdo. \square

La importancia de R_X reside en el hecho de que para calcular la probabilidad de un evento B solo interesan los puntos de B que están en R_X . En este sentido se dice que la probabilidad se concentra en R_X .

Teorema 3.2 *Para todo $B \in \mathcal{B}$ se tiene*

$$P_X(B) = P_X(R_X \cap B).$$

Demostración. Podemos escribir a B como la siguiente unión disjunta

$$B = (R_X \cap B) \cup (B - R_X), \tag{3.3}$$

y tomando probabilidad en ambos miembros se obtiene

$$P_X(B) = P_X(R_X \cap B) + P_X(B - R_X).$$

Pero

$$B - R_X \subset (R_X)^c,$$

de manera que

$$P_X(B - R_X) \leq P_X((R_X)^c) = 0.$$

Luego $P_X(B - R_X) = 0$ y el teorema resulta de (3.3). \square

Definición 3.3 Sea X una variable aleatoria discreta. Se define la función de densidad de probabilidad asociada a la variable X como la función

$$p_X : \mathbb{R} \rightarrow [0, 1]$$

tal que

$$p_X(x) = P_X(\{x\}).$$

También p_X se suele llamar función de probabilidad puntual de X o función de frecuencia de X .

Observación. La función de densidad satisface $p_X(x) > 0$ sii $x \in R_X$ y determina totalmente la probabilidad P_X .

Para ver esto probaremos el siguiente teorema.

Teorema 3.3 Si $B \in \mathcal{B}$ entonces

$$P_X(B) = \sum_{x \in B \cap R_X} p_X(x).$$

Demostración. $B \cap R_X$ se puede escribir como la siguiente unión disjunta

$$B \cap R_X = \bigcup_{x \in B \cap R_X} \{x\}.$$

Como $B \cap R_X$ es finito o numerable se tiene

$$P_X(B) = P_X(R_X \cap B) = \sum_{x \in B \cap R_X} p_X(x) \quad \square.$$

3.2. Ejemplos de distribuciones discretas.

3.2.1. Distribución Binomial.

Supongamos que se repite n veces un experimento que puede dar lugar a dos resultados: éxito o fracaso. Supongamos que todos los experimentos son independientes y tienen la misma probabilidad de éxito θ . Sea X la variable aleatoria definida como el número total de éxitos. La distribución de esta variable se denomina binomial con n repeticiones y probabilidad de éxito θ . La denotaremos con $\text{Bi}(\theta, n)$.

Para formalizar este experimento aleatorio tomaremos como espacio muestral

$$\Omega = \{(\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in \{0, 1\}\},$$

donde $\omega_i = 1$ indicará que el i -ésimo experimento resultó éxito y $\omega_i = 0$ que fue fracaso. Como Ω es finito podemos tomar como σ -álgebra \mathcal{A} el conjunto de partes de Ω .

La variable X se puede definir por

$$X((\omega_1, \omega_2, \dots, \omega_n)) = \sum_{i=1}^n \omega_i.$$

El rango de esta variable es $R_X = \{0, 1, \dots, n\}$. Obtendremos seguidamente su función de densidad. Sea $0 \leq x \leq n$, el evento $\{X = x\}$ está dado por

$$A_x = \{(\omega_1, \omega_2, \dots, \omega_n) \in \Omega : \sum_{i=1}^n \omega_i = x\}.$$

En primer lugar determinaremos la cantidad de elementos del conjunto A_x . Claramente un elemento de A_x queda determinado por los x lugares entre los n posibles donde aparecen los unos. De manera que

$$\#(A_x) = \binom{n}{x}.$$

Obsérvese que el espacio muestral no es equiprobable, por lo que la probabilidad no se determina con el esquema “casos favorables / casos igualmente posibles”.

Sea ω el resultado de un experimento cualquiera. Si $\omega = 0$ entonces $P(\omega) = 1 - \theta$ y si $\omega = 1$ entonces $P(\omega) = \theta$. Esto puede escribirse de manera más compacta de la siguiente manera

$$P(\omega) = \theta^\omega (1 - \theta)^{1-\omega}.$$

En primer lugar calculemos la probabilidad de un elemento arbitrario del espacio muestral. Teniendo en cuenta la independencia de los resultados de los distintos experimentos y que la ocurrencia de $(\omega_1, \omega_2, \dots, \omega_n)$ involucra una intersección de eventos se tiene que

$$\begin{aligned} P((\omega_1, \omega_2, \dots, \omega_n)) &= P\left(\bigcap_{i=1}^n \{\text{en el experimento } i \text{ el resultado es } \omega_i\}\right) \\ &= \prod_{i=1}^n P(\omega_i) \\ &= \prod_{i=1}^n \theta^{\omega_i} (1 - \theta)^{1-\omega_i} = \\ &= \theta^{\sum_{i=1}^n \omega_i} (1 - \theta)^{n - \sum_{i=1}^n \omega_i}. \end{aligned}$$

Ahora si $\omega = (\omega_1, \omega_2, \dots, \omega_n) \in A_x$ entonces $\sum_{i=1}^n \omega_i = x$ y queda que la probabilidad de ocurrencia de cualquier elemento ω de A_x es

$$p_X(\omega) = p_X((\omega_1, \omega_2, \dots, \omega_n)) = \theta^x (1 - \theta)^{n-x}$$

En definitiva como A_x se puede escribir como la siguiente unión disjunta

$$A_x = \bigcup_{\omega \in A_x} \{\omega\}$$

entonces

$$\begin{aligned} p_X(\omega) &= P(\{\omega : X(\omega) = x\}) \\ &= P(A_x) \\ &= \sum_{\omega \in A_x} P(\{\omega\}) = \\ &= \#(A_x) \theta^x (1 - \theta)^{n-x} \\ &= \binom{n}{x} \theta^x (1 - \theta)^{n-x}. \end{aligned}$$

3.2.2. Distribución Binomial Negativa (o Distribución de Pascal).

Consideremos, como en el caso de la distribución binomial, un experimento aleatorio cuyo resultado es éxito con probabilidad θ y fracaso con probabilidad $1 - \theta$. Supongamos que se hacen repeticiones independientes del experimento hasta que ocurran k éxitos. Los parámetros de esta distribución son θ : “probabilidad de éxito” y k : “el número de éxitos buscado”. Llamaremos X a la variable aleatoria definida como el número de experimentos que hay que realizar para obtener los k éxitos. La distribución de esta variable se denomina binomial negativa o de Pascal y se la denotará con $\text{BN}(\theta, k)$. El rango de X es

$$R_X = \{m \in \mathbb{N} : m \geq k\}$$

el cual es infinito numerable.

Consideremos la sucesión variables aleatorias independientes Z_i , $i \in \mathbb{N}$ definidas por

$$Z_i = \begin{cases} 1 & \text{si el } i\text{-ésimo experimento es éxito} \\ 0 & \text{si el } i\text{-ésimo experimento es fracaso,} \end{cases}$$

y definimos las variables

$$Y_i = \sum_{j=1}^i Z_j,$$

Claramente Y_i cuenta la cantidad de éxitos que se alcanzaron en los primeros i experimentos. Luego su distribución es $\text{Bi}(\theta, i)$.

El evento $\{X = x\}$, o sea el evento definido como “la cantidad de experimentos necesarios para alcanzar k éxitos es x ”, puede escribirse como una intersección de dos eventos

$$\{X = x\} = \{Y_{x-1} = k - 1\} \cap \{Z_k = 1\}.$$

Los dos eventos del lado derecho de la última ecuación son independientes. Luego, usando el hecho que Y_{x-1} tiene distribución $\text{Bi}(\theta, x - 1)$ resulta para $x \geq k$.

$$\begin{aligned} p_X(x) &= P(X = x) \\ &= P(Y_{x-1} = k - 1) P(Z_k = 1) \\ &= \binom{x-1}{k-1} \theta^{k-1} (1-\theta)^{x-k} \theta \\ &= \binom{x-1}{k-1} \theta^k (1-\theta)^{x-k}. \end{aligned} \tag{3.4}$$

3.2.3. Distribución Geométrica.

Se llama distribución geométrica a la $\text{BN}(\theta, k)$, con $k = 1$. Luego es la distribución de la variable aleatoria X definida como “el número de experimentos necesarios para alcanzar el primer éxito”. A esta distribución la denotaremos como $\mathcal{G}(\theta)$.

El rango de los valores posibles para la v.a. X es

$$R_X = \{1, 2, \dots, n, \dots\}.$$

Reemplazando $k = 1$ en (3.4) se obtiene

$$p_X(x) = \binom{x-1}{0} \theta (1-\theta)^{x-1} = \theta (1-\theta)^{x-1}.$$

Podemos verificar que

$$\begin{aligned} \sum_{x=1}^{\infty} p_X(x) &= \sum_{x=1}^{\infty} \theta (1-\theta)^{x-1} = \theta \sum_{x=1}^{\infty} (1-\theta)^{x-1} \\ &= \theta \sum_{j=0}^{\infty} (1-\theta)^j = \theta \frac{1}{1-(1-\theta)} = 1. \end{aligned}$$

3.2.4. Distribución Hipergeométrica.

Consideremos una urna que contiene N bolillas de las cuales D son negras y $N - D$ blancas. Se extraen secuencialmente (una a una) n bolillas y se define la variable X como el número total de bolillas negras extraídas. Si cada bolilla obtenida es repuesta en la urna antes de obtener la siguiente, el resultado de cada extracción es independiente de las anteriores, ya que esos resultados no modifican la composición de la urna. Luego en este caso X tendrá distribución $\text{Bi}(\theta, n)$ con $\theta = D/N$, ya que este número es la probabilidad de sacar cada vez una bolilla negra.

Si después de cada extracción la bolilla obtenida no se repone, no hay independencia en los resultados de las extracciones y la distribución de X se denomina hipergeométrica. La denotaremos por $\mathcal{H}(N, D, n)$.

Estudiemos el rango de esta distribución. Por un lado podemos observar que X no puede ser un número negativo, ni tampoco mayor que n , la cantidad total de bolillas extraídas. Por lo tanto:

$$0 \leq X \leq n. \quad (3.5)$$

Por otro lado, claramente a lo sumo se pueden extraer D negras, y luego

$$X \leq D. \quad (3.6)$$

Además el número de total de bolillas blancas extraídas debe ser menor que $N - D$. Por lo tanto también tenemos

$$n - X \leq N - D. \quad (3.7)$$

En definitiva de (3.5), (3.6) y (3.7) obtenemos

$$R_X = \{x \in N : \text{máx}(0, n - N + D) \leq x \leq \text{mín}(n, D)\}.$$

Podemos pensar que las D bolillas negras están numeradas de 1 a D , y las blancas de $D + 1$ a N . Luego si denotamos

$$I_N = \{x \in N : 1 \leq x \leq N\},$$

el resultado de extraer n bolillas será un subconjunto de I_N con cardinal n . Luego, podemos tomar como espacio muestral

$$\Omega = \{A \subset I_N : \#A = n\}.$$

Como todos estos subconjuntos tienen la misma probabilidad de ser extraídos, estaremos en un caso de resultados equiprobables. El cardinal de Ω es

$$\binom{N}{n}.$$

El evento $\{X = x\}$ corresponderá a aquellos subconjuntos A que contienen x bolillas negras y $n - x$ blancas. Para obtener el cardinal de $\{X = x\}$ procedamos de la siguiente manera. Primero consideremos el número de subconjuntos de x bolas negras elegidas entre las D posibles. Este número es

$$\binom{D}{x}.$$

Para cada uno de estos subconjuntos de x bolas negras hay

$$\binom{N - D}{n - x}$$

formas de elegir las restantes $n - x$ blancas. Luego

$$\#\{X = x\} = \binom{D}{x} \binom{N - D}{n - x},$$

y por lo tanto

$$p_X(x) = \frac{\#A_x}{\#\Omega} = \frac{\binom{D}{x} \binom{N - D}{n - x}}{\binom{N}{n}}.$$

Ejercicio.

Sea $n \in \mathbb{N}$ fijo y consideremos una sucesión de distribuciones hipergeométricas $\mathcal{H}(N, D_N, n)$, $N \in \mathbb{N}$ tales que

$$\lim_{N \rightarrow \infty} \frac{D_N}{N} = \theta.$$

Entonces si p_N^H es la densidad de probabilidad de una distribución $\mathcal{H}(N, D_N, n)$ y p^B la de una $\text{Bi}(\theta, n)$, se tiene

$$\lim_{N \rightarrow \infty} p_N^H(x) = p^B(x).$$

Es decir para N suficientemente grande la distribución $\mathcal{H}(N, D_N, n)$ se puede aproximar por la distribución $\text{Bi}(\theta, n)$. Heurísticamente, este resultado puede interpretarse como que debido a que n es pequeño con respecto a N , la reposición o no de las bolillas extraídas no cambia substancialmente la composición de la urna.

3.2.5. Distribución de Poisson.

La distribución de Poisson se presenta cuando se considera el número de veces que ocurre cierto evento en un intervalo determinado de tiempo. Por ejemplo

- (a) El número de clientes que entran en un determinado banco durante un día.

- (b) El número de accidentes automovilísticos que ocurren en la ciudad de Buenos Aires por mes.
- (c) El número total de llamadas telefónicas que llegan a una central telefónica entre las 15 hs. y 16 hs. de los días hábiles.

Para que las distribuciones de estas variables sean de Poisson, se requiere un conjunto de supuestos que trataremos con mayor detalle más adelante (ver el capítulo 12).

Por ahora sólo indicamos su función de densidad. Para cada $\lambda > 0$, se define la distribución de Poisson con parámetro λ que simbolizaremos por $\mathcal{P}(\lambda)$ por la siguiente densidad de probabilidad

$$p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!} \text{ para } x \in \mathbb{N}_{\geq 0},$$

donde $\mathbb{N}_{\geq 0}$ es el conjunto de enteros no negativos.

Es claro que

$$\sum_{x=0}^{\infty} p_X(x) = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = e^0 = 1.$$

3.2.6. Gráfico de la función de distribución asociada a una variable aleatoria discreta.

Supongamos que el rango de X sea finito $R_X = \{x_1, \dots, x_n\}$ y $x_1 < \dots < x_n$. En tal caso la función de distribución F_X es una función no decreciente escalonada, en los puntos de probabilidad positiva, x_j , $0 \leq j \leq n$.

Sea

$$c_i = \sum_{j=1}^i p_X(x_j); \quad 1 \leq i \leq n.$$

Luego se tendrá

$$F_X(x) \begin{cases} 0 & \text{si } x \in (-\infty, x_1) \\ c_i & \text{si } x \in [x_i, x_{i+1}), \quad 1 \leq i \leq n-1 \\ 1 & \text{si } x \in [x_n, \infty). \end{cases}$$

Ejercicio. Graficar la F_X para una $\text{Bi}(1/4, 10)$.

3.3. Variables aleatorias absolutamente continuas.

Definición 3.4 Se dice que una variable aleatoria X es continua *sii* F_X es continua para todo $x \in \mathbb{R}$.

Observación. Esto es equivalente a pedir que “la probabilidad en todo punto es cero.”

Definición 3.5 Se dice que F_X es absolutamente continua sii existe una función $f_X : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ tal que f_X es integrable Riemann sobre \mathbb{R} y para todo $x \in \mathbb{R}$ se tiene

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

La función f_X se denomina función de densidad de la probabilidad asociada a X .

Propiedades de las Distribuciones Continuas.

Propiedad 3.1 (a) Si f_X es una función de densidad de probabilidad para una variable aleatoria X entonces

$$\int_{-\infty}^{+\infty} f_X(t) dt = 1.$$

(b) Recíprocamente si $f \geq 0$ es integrable Riemann sobre \mathbb{R} y cumple que

$$\int_{-\infty}^{+\infty} f(t) dt = 1,$$

entonces definiendo

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

se obtiene una función que resulta ser la función de distribución de alguna variable aleatoria X .

Demostración.

(a) Resulta de

$$\begin{aligned} \int_{-\infty}^{+\infty} f_X(t) dt &= \lim_{x \rightarrow \infty} \int_{-\infty}^x f_X(t) dt \\ &= \lim_{x \rightarrow \infty} F_X(x) = 1. \end{aligned}$$

(b) Usando propiedades de las integrales de Riemann se puede mostrar que F_X satisface las cuatro propiedades del Teorema 2.5 . Luego este resultado se obtiene del Corolario 2.2 del Teorema 2.7. \square

Propiedad 3.2 *Supongamos que F_X es absolutamente continua. Entonces*

$$P_X((a, b]) = \int_a^b f_X(t) dt.$$

Demostración.

$$\begin{aligned} P_X((a, b]) &= P_X((-\infty, b]) - P_X((-\infty, a]) \\ &= F_X(b) - F_X(a) \\ &= \int_{-\infty}^b f_X(t) dt - \int_{-\infty}^a f_X(t) dt \\ &= \int_a^b f_X(t) dt. \quad \square \end{aligned}$$

Propiedad 3.3 *Si F_X es absolutamente continua entonces es continua.*

Demostración. Primero supondremos que f_X es acotada en un entorno del punto x . Luego existe $\delta > 0$ y M positivo tal que $f(x) \leq M$ para todo $x \in [x - \delta, x]$. Luego para todo $\varepsilon \leq \delta$ tenemos

$$\begin{aligned} P_X(\{x\}) &\leq P((x - \varepsilon, x]) \\ &= \int_{x-\varepsilon}^x f_X(t) dt \\ &\leq \varepsilon M. \end{aligned}$$

Como esto vale para todo $\varepsilon \leq \delta$, resulta $P_X(\{x\}) = 0$. Luego F_X es continua en x .

Supongamos ahora que f_X no es acotada en ningún entorno del punto x . Luego

$$F_X(x) = \int_{-\infty}^x f_X(t) dt$$

se define por

$$\begin{aligned} \int_{-\infty}^x f_X(t) dt &= \lim_{y \uparrow x} \int_{-\infty}^y f_X(t) dt \\ &= \lim_{y \uparrow x} F_X(y), \end{aligned}$$

y luego F_X es continua en x . \square

El nombre densidad nos recuerda “la cantidad de masa por unidad de longitud, área o volumen” según el caso. En este caso se puede decir que $f_X(x)$ indica la probabilidad por unidad de longitud “en las cercanías del punto x ”. Más precisamente podemos enunciar el siguiente teorema.

Teorema 3.4 Sea f_X una función de densidad continua en x_0 , entonces

$$\lim_{h \rightarrow 0} \frac{P_X([x_0 - h, x_0 + h])}{2h} = \lim_{h \rightarrow 0} \frac{1}{2h} \int_{x_0-h}^{x_0+h} f_X(t) dt = f_X(x_0).$$

Demostración. Sea

$$M_h = \max\{f_X(x) : x \in [x_0 - h; x_0 + h]\}$$

y

$$m_h = \min\{f_X(x) : x \in [x_0 - h; x_0 + h]\}.$$

Por continuidad

$$f_X(x_0) = \lim_{h \rightarrow 0} M_h = \lim_{h \rightarrow 0} m_h. \quad (3.8)$$

Por otro lado valen las desigualdades

$$2hm_h \leq \int_{x_0-h}^{x_0+h} f_X(t) dt \leq 2hM_h,$$

y dividiendo por $2h$ en todos los miembros queda:

$$m_h \leq \frac{1}{2h} \int_{x_0-h}^{x_0+h} f_X(t) dt \leq M_h.$$

Luego, teniendo en cuenta (3.8) y pasando al límite cuando $h \rightarrow 0$ se obtiene

$$f_X(x_0) \leq \lim_{h \rightarrow 0} \frac{P_X([x_0 - h; x_0 + h])}{2h} \leq f_X(x_0),$$

de donde se deduce el Teorema. \square

Teorema 3.5 Sea f_X una función de densidad continua en x_0 y F_X la distribución asociada. Entonces F_X es derivable en x_0 y

$$F'_X(x_0) = f_X(x_0).$$

Demostración. Se deduce de la anterior. \square

Comentario vinculado a la teoría de la medida.

En este párrafo el signo \int corresponde a la integral de Lebesgue. Más generalmente se definen distribuciones absolutamente continuas utilizando funciones Borel medibles. Sea $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ una función Borel medible tal que

$$\int_{-\infty}^{\infty} f(t) dt = 1. \quad (3.9)$$

Entonces se puede definir una función de distribución absolutamente continua por

$$F(x) = \int_{-\infty}^x f(t) dt, \quad (3.10)$$

Se puede demostrar que la función F definida por (3.10) cumple las cuatro propiedades del Teorema 2.5 y es continua y derivable en casi todo punto con derivada $f(x)$. Además si P es la correspondiente probabilidad sobre \mathbb{R} asociada a F y garantizada por el Teorema de Extensión, dado cualquier boreliano B se tendrá

$$P(B) = \int_B f(t) dt = \int_{-\infty}^{\infty} I_B(t)f(t) dt,$$

donde $I_B(t)$ es la función indicadora del conjunto B .

3.4. Ejemplos de distribuciones continuas.

3.4.1. Distribución uniforme en un intervalo.

Consideremos dos números reales $a < b$. Luego la distribución uniforme, denotada por $\mathcal{U}(a, b)$, tiene como densidad

$$f_X(x) = \begin{cases} k & \text{si } x \in [a, b] \\ 0 & \text{si } x \notin [a, b]. \end{cases}$$

con $k = \frac{1}{b-a} > 0$. Claramente

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_a^b k dx = \frac{k}{b-a} = 1.$$

Ejercicio. Mostrar que la función distribución de $\mathcal{U}(a, b)$ es

$$F_X(x) \begin{cases} 0 & \text{si } x \in (-\infty, a) \\ \frac{x-a}{b-a} & \text{si } x \in [a; b) \\ 1 & \text{si } x \in (b, \infty). \end{cases}$$

Ejercicio. Mostrar que no existe ninguna distribución uniforme sobre toda la recta.

En particular consideremos la distribución uniforme $\mathcal{U}(0, 1)$ que tiene como densidad

$$f_X(x) = \begin{cases} 1 & \text{si } x \in [0; 1] \\ 0 & \text{si } x \notin [0; 1]. \end{cases}$$

La función de distribución es en este caso

$$F_X(x) = \begin{cases} 0 & \text{si } x \in (-\infty, 0] \\ x & \text{si } x \in (0, 1] \\ 1 & \text{si } x \in (1, \infty). \end{cases} \quad (3.11)$$

Observaciones.

1. Es claro que (3.11) es cierta puesto que si $x \in (0, 1)$

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x f_X(t) dt \\ &= \int_{-\infty}^0 f_X(t) dt + \int_0^x f_X(t) dt \\ &= 0 + \int_0^x 1 dt \\ &= x. \end{aligned}$$

2. Sea $I = (c, d) \subset (0, 1)$ ¿Cuál es la probabilidad de que $X \in (c, d)$?

$$P_X([c < X < d]) = F_X(d) - F_X(c) = d - c.$$

Es decir, la probabilidad que esta distribución asigna a cada intervalo contenido en $[0, 1]$ es su longitud.

3. Pueden generarse distribuciones uniformes de muchas maneras diferentes. Por ejemplo podemos elegir dos números A_1, A_2 de ocho dígitos, y definir A_3 por los últimos ocho dígitos de A_1A_2 . En general si ya hemos definido A_1, A_2, \dots, A_k como enteros de ocho dígitos, podemos definir recursivamente A_{k+1} como los últimos ocho dígitos de $A_{k-1}A_k$. Este proceso lo podemos continuar hasta obtener A_n para un n dado. Luego generamos n números con distribución $\mathcal{U}(0, 1)$ por

$$U_i = A_i 10^{-8}, \quad 1 \leq i \leq n.$$

Estos números no serán aleatorios. Sin embargo se comportarán como si fuesen variables aleatorias independientes con distribución $\mathcal{U}(0, 1)$. En particular, dados a y b tales que $0 < a < b < 1$, se tendrá que si n es grande

$$\frac{\#\{i : 1 \leq i \leq n, a < U_i < b\}}{n}$$

será aproximadamente $b - a$. Es decir la frecuencia con la cual los U_i están en un intervalo (a, b) es aproximadamente la probabilidad que la distribución $\mathcal{U}(0, 1)$ asigna a ese intervalo.

3.4.2. Generación de distribuciones a partir de la distribución uniforme en $[0,1]$

Vamos a mostrar cómo a partir de una variable aleatoria con distribución $\mathcal{U}(0,1)$ se puede generar cualquier otra variable con cualquier función de distribución.

Para esto en primer lugar necesitamos algunas definiciones. Sabemos que una función de distribución no tiene por qué ser continua y mucho menos biyectiva, de manera que en general su inversa no existe. Pero podemos definir una función que tendrá propiedades análogas.

Sea $F : \mathbb{R} \rightarrow [0,1]$ una función que cumple con las cuatro propiedades del Teorema 2.5 que caracterizan una función de distribución y consideremos $y \in (0,1)$.

Definimos

$$A_y = \{x \in \mathbb{R} : F(x) \geq y\}.$$

Observaciones.

1. Puede ocurrir que exista una preimagen vía F del punto $y : F^{-1}(y) \neq \emptyset$. Si F es continua por Bolzano podemos asegurar que asume todos los valores intermedios entre el 0 y el 1 y en consecuencia en algún punto x asumirá el valor y .
2. Puede ocurrir también que no exista la preimagen. Por ejemplo si F no es continua para algunos valores de y ocurrirá que $F^{-1}(y) = \emptyset$.
3. Puede ocurrir que existan infinitas preimágenes. Basta con tomar una función con las propiedades de función de distribución que sea constante en un intervalo. Para y igual a ese valor hay infinitas preimágenes.

Ejercicio. Dar un ejemplo de cada una de las situaciones y dibujar el gráfico correspondiente.

Teorema 3.6 *Existe el ínfimo del conjunto A_y .*

Demostración. Basta probar que $A_y \neq \emptyset$ y está acotado inferiormente. Comencemos probando que $A_y \neq \emptyset$. Sabemos que F satisface la propiedad (2) del Teorema 2.5 y por lo tanto

$$\lim_{n \rightarrow \infty} F(n) = 1.$$

Como $0 < y < 1$ existe $n_0 \in \mathbb{N}$ tal que

$$F(n_0) \geq y,$$

de manera que $n_0 \in A_y$. Ahora probaremos que A_y esta acotado inferiormente. Por la propiedad (3) del Teorema 2.5 se tiene que,

$$\lim_{n \rightarrow \infty} F(-n) = 0.$$

Como $y > 0$ entonces existe $n_0 \in \mathbb{N}$ tal que

$$F(-n_0) < y. \quad (3.12)$$

Ahora bien si $x \in A_y$ no puede ser que $-n_0 > x$ puesto que por monotonia (Propiedad (1) del Teorema 2.5) se cumpliría

$$F(-n_0) \geq F(x) \geq y,$$

en contradicción con (3.12). En definitiva se tiene que si $x \in A_y$, entonces $-n_0 \leq x$, y por lo tanto A_y esta acotado inferiormente. \square

En virtud de la existencia y unicidad del ínfimo podemos definir la siguiente función

Definición 3.6 *Dada*

$$F : \mathbb{R} \rightarrow [0, 1]$$

que satisface las propiedades de una función de distribución (Propiedades (1)-(4) del Teorema 2.5) se define $F^{-1} : (0, 1) \rightarrow \mathbb{R}$ por

$$F^{-1}(y) = \inf A_y.$$

Propiedades de la función F^{-1} .

Propiedad 3.4 (a) *Dada una función de distribución F , se tiene*

$$F(F^{-1}(y)) \geq y.$$

(b) *El ínfimo del conjunto A_y resulta ser el mínimo de A_y , es decir*

$$F^{-1}(y) = \min A_y.$$

Demostración. Bastará probar (a), ya que en ese caso $F^{-1}(y)$ pertenece al conjunto A_y . Por definición de ínfimo existe una sucesión $(x_n)_{n \in \mathbb{N}} \subset A_y$ decreciente que converge a $F^{-1}(y)$, es decir tal que

$$\lim_{n \rightarrow \infty} x_n = F^{-1}(y).$$

Por la propiedad de continuidad a derecha de F

$$\lim_{n \rightarrow \infty} F(x_n) = F(F^{-1}(y)). \quad (3.13)$$

Ahora, como para todo $n \in \mathbb{N}$ se tiene que $x_n \in A_y$ sabemos que

$$F(x_n) \geq y,$$

y luego por (3.13) resulta

$$F(F^{-1}(y)) \geq y, \quad (3.14)$$

por lo tanto (a) queda demostrado. Esto implica $F^{-1}(y) \in A_y$. Luego hemos mostrado (a) y por lo tanto también hemos demostrado (b). \square

Propiedad 3.5 *Si F es continua entonces*

$$F(F^{-1}(y)) = y.$$

Demostración. Sabemos que $F(F^{-1}(y)) \geq y$. Ahora supongamos que no se cumple la igualdad, esto es que

$$F(F^{-1}(y)) > y.$$

Veremos que esto contradice el carácter de ínfimo del elemento $F^{-1}(y)$. Tomemos un punto intermedio entre $F(F^{-1}(y))$ e y que llamaremos y^* . Entonces

$$y < y^* < F(F^{-1}(y)).$$

Por ser F continua, por el teorema de Bolzano se deduce que existe $x^* \in (0, 1)$ tal que

$$F(x^*) = y^*.$$

Luego reemplazando en la inecuación anterior se obtiene la desigualdad

$$y < F(x^*) < F(F^{-1}(y)).$$

Por un lado esto dice que $x^* \in A_y$ y por otro teniendo en cuenta la monotonía de F resulta

$$x^* < F^{-1}(y).$$

Esto contradice que $F^{-1}(y)$ sea el mínimo, absurdo. \square

Propiedad 3.6 *Dada una función de distribución F , se cumple que*

$$F^{-1}(F(x)) \leq x.$$

Demostración. Es claro que para todo x se tiene que $x \in A_{F(x)}$ puesto que $F(x) \leq F(x)$. Sabemos que $F^{-1}(F(x))$ es el mínimo de $A_{F(x)}$ y luego

$$a \in A_{F(x)} \text{ implica } F^{-1}(F(x)) \leq a.$$

En particular si tomamos $a = x \in A_{F(x)}$ se obtiene el resultado buscado. \square

Teorema 3.7 (Caracterización de A_y como semirecta) Sea F una función de distribución y tomemos $y \in (0, 1)$ fijo. Los conjuntos

$$A_y = \{x : F(x) \geq y\},$$

$$B_y = \{x : x \geq F^{-1}(y)\} = [F^{-1}(y), +\infty)$$

coinciden.

Demostración. Sabemos por la Propiedad 3.4 (b) que

$$F^{-1}(y) = \min A_y.$$

Por otro lado es fácil ver que si $x \in A_y$ y $x^* > x$, entonces también $x^* \in A_y$. Luego $A_y = [F^{-1}(y), \infty)$. \square

Ejercicio. Probar que F^{-1} es monótona no decreciente y por lo tanto medible.

Veremos ahora que dada cualquier función de distribución F , a partir de cualquier variable aleatoria con distribución $\mathcal{U}(0, 1)$, se puede generar otra variable aleatoria con función de distribución F .

Teorema 3.8 Sea U una variable aleatoria con distribución $\mathcal{U}(0, 1)$. Luego si F es una función de distribución (propiedades (1)-(4) del Teorema 2.5) se tiene que $X = F^{-1}(U)$ tiene función de distribución F

Demostración. Usando el Teorema 3.7 y el hecho de que $F_U(u) = u, 0 \leq u \leq 1$, se tiene

$$F_X(x) = P_X((-\infty, x]) = P(\{F^{-1}(U) \leq x\}) = P(\{U \leq F(x)\})$$

$$= F_U(F(x)) = F(x). \quad \square$$

Ejercicio. Sea X una variable con rango $R_X = \mathbb{N}_{\geq 0}$ (enteros no negativos) y sea $p_j = p_X(j), j \in \mathbb{N}_{\geq 0}$. Verificar que F_X^{-1} es de la forma

$$F_X^{-1}(y) = \begin{cases} 0 & \text{si } 0 < y \leq p_0 \\ i & \text{si } \sum_{j=0}^{i-1} p_j < y \leq \sum_{j=0}^i p_j, i \geq 1. \end{cases}$$

Comprobar que el resultado anterior vale en este caso.

El siguiente teorema de demostración inmediata es muy importante.

Teorema 3.9 Sean X y X^* dos variables aleatorias tales que $F_X = F_{X^*}$. Consideremos una función g medible y consideremos las variables aleatorias obtenidas componiendo

$$Z = g(X); Z^* = g(X^*).$$

Entonces

$$P_Z = P_{Z^*}.$$

Demostración. Sea $B \in \mathcal{B}$ y probemos que

$$P_Z(B) = P_{Z^*}(B).$$

Sabemos que

$$\begin{aligned} P_Z(B) &= P(Z^{-1}(B)) \\ &= P(X^{-1}(g^{-1}(B))) \\ &= P_X(g^{-1}(B)). \end{aligned}$$

Por el Corolario 2.1 del Teorema de Extensión se tiene que $P_X(g^{-1}(B)) = P_{X^*}(g^{-1}(B))$ y luego

$$\begin{aligned} P_Z(B) &= P_{X^*}(g^{-1}(B)) \\ &= P(X^{*-1}(g^{-1}(B))) \\ &= P(Z^{*-1}(B)) \\ &= P_{Z^*}(B). \quad \square \end{aligned}$$

El siguiente resultado vale para funciones de distribución continuas.

Teorema 3.10 *Si X es una variable aleatoria con distribución F_X continua y consideramos la variable aleatoria $Y = F_X(X)$ entonces Y tiene distribución $\mathcal{U}(0,1)$.*

Demostración. Consideremos una variable aleatoria U con distribución $\mathcal{U}(0,1)$ y sea $X^* = F_X^{-1}(U)$. Sabemos que X^* tiene distribución F_X . Luego por el Teorema 3.9 las variables

$$Y = F_X(X), \quad Y^* = F_X(X^*)$$

tienen la misma distribución. Pero

$$Y^* = F_X(X^*) = F_X(F_X^{-1}(U)),$$

y siendo F_X continua por Propiedad 3.5 se tiene $F_X(F_X^{-1}(U)) = U$. Luego Y^* tiene distribución $\mathcal{U}(0,1)$ y por lo tanto, de acuerdo al Teorema 3.9 también esa es la distribución de Y . \square

3.4.3. Distribución Normal $N(\mu, \sigma^2)$.

La distribución *normal* es tal vez la más importante y sin lugar a dudas la que se usa con mayor frecuencia. A veces este uso se hace de manera inadecuada sin verificar los supuestos que la identifican. Veremos más adelante la importancia de esta distribución. Adelantamos sin embargo, informalmente

que si $\{Y_n\}_{n \in \mathbb{N}}$ es una sucesión de variables a independientes tales que ninguna de ellas prevalezca sobre las otras, entonces la variable aleatoria

$$S_n = \sum_{j=1}^n Y_j$$

es aproximadamente normal para n suficientemente grande. Esta distribución tiene mucha aplicación en la teoría de errores, donde se supone que el error total de medición es la suma de errores que obedecen a diferentes causas. La distribución normal depende de dos parámetros $\mu \in \mathbb{R}$ y $\sigma^2 \in \mathbb{R}_{>0}$.

En este capítulo solo veremos la distribución normal correspondiente a $\mu = 0$ y $\sigma^2 = 1$. En este caso la función de densidad es

$$f_X(x) = K \exp\left(\frac{-x^2}{2}\right),$$

donde K es una constante y $\exp(x)$ es la función exponencial e^x . Calcularemos la constante K de forma tal que

$$1 = \int_{-\infty}^{+\infty} K \exp\left(\frac{-x^2}{2}\right) dx,$$

y por lo tanto

$$K = \frac{1}{\int_{-\infty}^{+\infty} \exp\left(\frac{-x^2}{2}\right) dx}.$$

Sea

$$I = \int_{-\infty}^{+\infty} \exp\left(\frac{-x^2}{2}\right) dx.$$

Para el cálculo de esta integral podemos usar o bien residuos (teoría de análisis complejo) o bien calcular I^2 como integral doble a través de un cambio de variable a coordenadas polares. Optamos por la segunda forma

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{+\infty} \exp\left(\frac{-x^2}{2}\right) dx \right) \left(\int_{-\infty}^{+\infty} \exp\left(\frac{-y^2}{2}\right) dy \right) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\left(\frac{-x^2}{2}\right) \exp\left(\frac{-y^2}{2}\right) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\left(\frac{-(x^2 + y^2)}{2}\right) dx dy. \end{aligned}$$

Ahora hacemos el cambio de variable

$$\begin{aligned} x(\rho, \phi) &= x = \rho \cos(\phi) \\ y(\rho, \phi) &= y = \rho \sin(\phi) \end{aligned}$$

Claramente se tiene

$$x^2 + y^2 = \rho^2$$

La transformación del cambio de variable $T(\rho, \phi) = (x(\rho, \phi), y(\rho, \phi)) = (\rho \cos(\phi), \rho \sin(\phi))$ $\rho \geq 0$, $0 \leq \phi < 2\pi$ tiene matriz diferencial

$$DT(\rho, \phi) = \begin{pmatrix} x_\rho & x_\phi \\ y_\rho & y_\phi \end{pmatrix} = \begin{pmatrix} \cos(\phi) & -\rho \sin(\phi) \\ \sin(\phi) & \rho \cos(\phi) \end{pmatrix}.$$

Entonces su jacobiano

$$\begin{aligned} J(\rho, \phi) &= \det(DT(\rho, \phi)) = \det \begin{pmatrix} \cos(\phi) & -\rho \sin(\phi) \\ \sin(\phi) & \rho \cos(\phi) \end{pmatrix} \\ &= \rho \cos^2(\phi) + \rho \sin^2(\phi) = \rho. \end{aligned}$$

En definitiva $|J(\rho, \phi)| = \rho$ y aplicando la fórmula de cambio de variables en integrales múltiples resulta

$$\begin{aligned} I^2 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\left(\frac{-(x^2 + y^2)}{2}\right) dx dy = \\ &= \int_0^{+\infty} \int_0^{2\pi} \exp\left(\frac{-\rho^2}{2}\right) \rho d\phi d\rho = \\ &= 2\pi \int_0^{+\infty} \exp\left(\frac{-\rho^2}{2}\right) \rho d\rho = 2\pi \int_0^{+\infty} \exp\left(\frac{-\rho^2}{2}\right) \rho d\rho. \end{aligned}$$

Haciendo el cambio de variable

$$\begin{aligned} u &= \frac{\rho^2}{2}, \\ du &= \rho d\rho \end{aligned}$$

se obtiene

$$\begin{aligned} I^2 &= 2\pi \int_0^{+\infty} \exp(-u) du \\ &= 2\pi \left(-\exp(-u) \Big|_0^{+\infty}\right) \\ &= 2\pi, \end{aligned}$$

y por lo tanto

$$I = \sqrt{2\pi}$$

Luego

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right).$$

3.4.4. Distribución Exponencial.

Esta distribución depende de un parámetro λ que puede tomar cualquier valor real positivo. Su función de densidad es

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0. \end{cases}$$

Haciendo la transformación $y = \lambda x$, $dy = \lambda dx$ se obtiene

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \lambda \int_0^{\infty} e^{-\lambda x} dx = \int_0^{\infty} e^{-y} dy \\ &= [-e^{-y}] \Big|_0^{\infty} = 0 + 1 = 1. \end{aligned}$$

Se deja como ejercicio verificar que la correspondiente función de distribución es

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0. \end{cases} \quad (3.15)$$

La distribución exponencial con parámetro λ será denotada por $\mathcal{E}(\lambda)$.

Esta distribución aparece generalmente cuando se trata de estudiar la durabilidad de un mecanismo bajo el supuesto de que el sistema no se desgasta a lo largo del tiempo. Como ejemplo suele citarse a veces la duración de una lámpara eléctrica. Sin embargo en este caso existe un cierto desgaste propio de la lámpara y su distribución no es exactamente exponencial. Esta distribución es más adecuada para modelar la duración de los mecanismos electrónicos, ya que estos no tienen prácticamente desgaste.

Para precisar el concepto de desgaste decimos que la distribución de X *no tiene desgaste* cuando dado $a > 0$ y $b > 0$ se tiene

$$P(X \geq a + b | X \geq a) = P(X \geq b).$$

Esto significa que la probabilidad de que llegue a durar hasta el tiempo $a + b$, dado que ha llegado hasta el tiempo a , es igual a la probabilidad de que haya durado hasta el tiempo b . Es decir el proceso “no tiene memoria del tiempo que estuvo funcionando” (no recuerda qué tan viejo es) y por tanto, mientras funciona lo hace como si fuese nuevo.

Decimos por el contrario que hay desgaste si

$$P(X \geq a + b | X \geq a)$$

es una función decreciente de a .

Vamos a mostrar que la propiedad de falta de desgaste caracteriza a la distribución exponencial. Esto significa que las únicas distribuciones continuas y no negativas que tienen la propiedad de “falta de desgaste” son las exponenciales.

Como $\{X \geq a + b\} \cap \{X \geq a\} = \{X \geq a + b\}$ resulta que

$$P(X \geq a + b | X \geq a) = \frac{P(\{X \geq a + b\} \cap \{X \geq a\})}{P(X \geq a)} = \frac{P(\{X \geq a + b\})}{P(X \geq a)}.$$

Por lo tanto la propiedad de “falta de desgaste” se puede escribir como

$$\frac{P(X \geq a + b)}{P(X \geq a)} = P(X \geq b),$$

o equivalentemente

$$P(X \geq a + b) = P(X \geq b) P(X \geq a). \quad (3.16)$$

Si X tiene distribución continua de $P(X \leq a) = F_X(a)$ resulta

$$1 - F_X(a) = P(X > a) = P(X \geq a).$$

Entonces definimos

$$G_X(a) = 1 - F_X(a),$$

y como la propiedad de “falta de memoria” es equivalente (3.16), esta se puede escribir también como

$$G_X(a + b) = G_X(a) G_X(b) \quad (3.17)$$

para todo $a \geq 0, b \geq 0$.

En el caso en que X tiene distribución exponencial por (3.15) se tiene

$$G_X(x) = e^{-\lambda x}$$

para todo $x \geq 0$. El siguiente teorema muestra que la propiedad de falta de memoria caracteriza a las distribuciones exponenciales.

Teorema 3.11 *Sea X una variable aleatoria continua con valores no negativos. Luego la propiedad de falta de memoria dada por (3.17) se cumple si y sólo si $G_X(x) = e^{-\lambda x}$ es decir si X tiene distribución exponencial.*

Demostración. Supongamos primero que $G_X(x) = e^{-\lambda x}$. Probaremos que (3.17) se cumple. En efecto

$$G_X(a + b) = e^{-\lambda(a+b)} = e^{(-\lambda a) + (-\lambda b)} = e^{-\lambda a} e^{-\lambda b} = G_X(a) G_X(b).$$

Supongamos ahora que (3.17) se cumple. Probaremos que $G_X(x) = e^{-\lambda x}$ para algún $\lambda > 0$. En primer lugar veamos que para todo n , dados $a_1 \geq 0, \dots, a_n \geq 0$ entonces

$$G_X\left(\sum_{i=1}^n a_i\right) = \prod_{i=1}^n G_X(a_i).$$

Probaremos esta proposición por inducción. Claramente vale para $n = 2$ por hipótesis.

Supongamos que vale para n y probemos que vale para $n + 1$.

$$\begin{aligned} G_X \left(\sum_{i=1}^{n+1} a_i \right) &= G_X \left(\sum_{i=1}^n a_i + a_{n+1} \right) \\ &= G_X \left(\sum_{i=1}^n a_i \right) G_X (a_{n+1}) \\ &= \left(\prod_{i=1}^n G_X (a_i) \right) G_X (a_{n+1}) \\ &= \prod_{i=1}^{n+1} G_X (a_i). \end{aligned}$$

Ahora probemos que para todo $a \geq 0$ vale que

$$G_X (a) = [G_X (1)]^a.$$

La estrategia es primero probarlo para cuando a es un entero no negativo, luego cuando es un racional no negativo y por último cuando es un número real no negativo. Sea $n \in \mathbb{N}$ entonces

$$\begin{aligned} G_X (n) &= G_X \left(\underbrace{1 + 1 + \dots + 1}_{n \text{ sumandos}} \right) \\ &= [G_X (1)]^n. \end{aligned}$$

Ahora sea $a = \frac{m}{n} \in \mathbb{Q}$ el conjunto de los números racionales. Entonces

$$\begin{aligned} G_X (m) &= G_X \left(n \frac{m}{n} \right) \\ &= G_X \left(\underbrace{\frac{m}{n} + \dots + \frac{m}{n}}_{n \text{ sumandos}} \right) \\ &= G_X \left(\frac{m}{n} \right)^n. \end{aligned}$$

Entonces

$$\begin{aligned} G_X \left(\frac{m}{n} \right) &= [G_X (m)]^{\frac{1}{n}} \\ &= [(G_X (1))^m]^{\frac{1}{n}} \\ &= [G_X (1)]^{\frac{m}{n}}. \end{aligned}$$

Por último consideremos $a \in \mathbb{R}_{\geq 0}$. Elijamos una sucesión $(r_n)_{n \in \mathbb{N}} \subset \mathbb{Q}$ tal que $r_n \rightarrow a$. Siendo G_X continua resulta

$$\begin{aligned} G_X(a) &= \lim_{n \rightarrow \infty} G_X(r_n) \\ &= \lim_{n \rightarrow \infty} (G_X(1))^{r_n} \\ &= (G_X(1))^{\lim_{n \rightarrow \infty} r_n} \\ &= [G_X(1)]^a. \end{aligned} \tag{3.18}$$

Veamos que $0 < G_X(1) < 1$. Supongamos que $G_X(1) = 0$. Luego por (3.18) $G_X(a) = 0$ para todo $a \geq 0$. En particular $G_X(0) = 0$ y luego $F_X(0) = 1$. Esto implica que $P(X = 0) = 1$ y luego X es discreta. Supongamos ahora que $G_X(1) = 1$. Luego por (3.18) tenemos que para todo $a \geq 0$ se tiene $G_X(a) = 1$. Luego para todo $a \geq 0$ resulta $F_X(a) = 0$ y entonces $\lim_{x \rightarrow \infty} F_X(x) = 0$, lo cual es un absurdo, ya que este límite es 1. Luego podemos definir

$$\lambda = -\log(G_X(1)),$$

de manera que

$$G_X(1) = e^{-\lambda}$$

Luego, usando (3.18), podemos escribir

$$G_X(a) = [G_X(1)]^a = e^{-\lambda a},$$

y el teorema queda probado. \square

3.5. Variables aleatorias mixtas.

Además de las variables discretas y absolutamente continuas existen otros tipos de variables. Un estudio exhaustivo de los tipos de variables aleatorias requiere algunos conocimientos de la teoría de la medida. Aquí introduciremos las variables mixtas cuya función distribución es una combinación convexa de funciones de una distribución discreta y otra absolutamente continua.

Definición 3.7 *Decimos que F es una función de distribución mixta si es una combinación convexa de una distribución absolutamente continua y otra discreta. Más precisamente, si existen δ , $0 < \delta < 1$, F_1 función de distribución absolutamente continua, F_2 función de distribución discreta tal que*

$$F = (1 - \delta) F_1 + \delta F_2. \tag{3.19}$$

Teorema 3.12 *Si F está dada por (3.19) se tiene que*

(a) F es una función de distribución.

(b) F no corresponde a la función de distribución de una variable absolutamente continua ni a una discreta.

Demostración.

(a) Por el Corolario 2.2 de la página 39 basta probar que F satisface las Propiedades 1-4 del Teorema 2.5. Probemos primero que F es monótona no decreciente. Sean $x < x'$. Luego como F_1 y F_2 son monótonas no decrecientes se tendrá $F_1(x) \leq F_1(x')$ y como $1 - \delta > 0$ resulta

$$(1 - \delta)F_1(x) \leq (1 - \delta)F_1(x'). \quad (3.20)$$

Del mismo se tiene que

$$\delta F_2(x) \leq \delta F_2(x'). \quad (3.21)$$

Sumando miembro a miembro (3.20) y (3.21) resulta que $F(x) \leq F(x')$.

Multiplicando por una constante se conserva la propiedad de que una función es continua a derecha y sumando funciones continuas a derecha se obtiene otra función continua a derecha. Esto prueba que F es continua a derecha.

Por otro lado, tenemos que

$$\begin{aligned} \lim_{x \rightarrow +\infty} F(x) &= \lim_{x \rightarrow +\infty} ((1 - \delta)F_1 + \delta F_2)(x) \\ &= (1 - \delta) \lim_{x \rightarrow +\infty} F_1(x) + \delta \lim_{x \rightarrow +\infty} F_2(x) \\ &= (1 - \delta) + \delta = 1. \end{aligned}$$

Finalmente, también vale que:

$$\begin{aligned} \lim_{x \rightarrow -\infty} F(x) &= \lim_{x \rightarrow -\infty} ((1 - \delta)F_1 + \delta F_2)(x) \\ &= (1 - \delta) \lim_{x \rightarrow -\infty} F_1(x) + \delta \lim_{x \rightarrow +\infty} F_2(x) \\ &= 0. \end{aligned}$$

Por lo tanto (a) queda probado.

(b) Veamos ahora que F no corresponde a la función de de distribución de una variable absolutamente continua o discreta. Sean P_i , las probabilidades inducidas por las distribuciones F_i , $i = 1, 2$. Luego si P es la probabilidad asociada a F , usando el Teorema de Extensión de la 39 se puede probar que

$$P(B) = (1 - \delta)P_1(B) + \delta P_2(B) \quad \forall B \in \mathcal{B}_1.$$

Esta comprobación se deja como ejercicio. Sea R_2 el rango de una variable con distribución F_2 . Por lo tanto R_2 es numerable y $P_2(R_2) = 1$. Luego

$$P(R_2) = (1 - \delta) P_1(R_2) + \delta P_2(R_2) \geq \delta P_2(R_2) = \delta > 0$$

Por lo que se deduce que F no corresponde a una distribución absolutamente continua, ya que éstas asignan probabilidad 0 a todo conjunto numerable.

Para ver que no es discreta veamos que sobre un conjunto numerable arbitrario su probabilidad es menor que 1. Sea A un conjunto numerable, luego, teniendo en cuenta que F_1 es absolutamente continua resulta que $P_1(A) = 0$. Luego

$$\begin{aligned} P(A) &= (1 - \delta) P_1(A) + \delta P_2(A) \\ &= \delta P_2(A) \leq \delta < 1. \end{aligned}$$

Como esto ocurre para todo A arbitrario, F no puede ser discreta. \square

Ejemplo 3.1 Sea $U \sim \mathcal{U}[0, 1]$ y consideremos $V = \min(U, \frac{1}{2})$. Entonces

$$F_V(u) = \begin{cases} u & \text{si } u < \frac{1}{2} \\ 1 & \text{si } u \geq \frac{1}{2} \end{cases}$$

Claramente $P(V = 1/2) = P(1/2 \leq U \leq 1) = 1/2$ de manera que V no es absolutamente continua. Tampoco es discreta. Es fácil ver que

$$F = \frac{1}{2}F_1 + \frac{1}{2}F_2$$

donde F_1 es la distribución de una $\mathcal{U}[0, 1/2)$ y F_2 la distribución de una variable discreta que asigna probabilidad 1 a $x = \frac{1}{2}$.

Veremos cómo se puede generar una variable con la distribución mixta (3.19).

Teorema 3.13 Consideremos variables aleatorias independientes X_1 con distribución F_1 , X_2 con distribución F_2 y U que toma valores 0 y 1 con probabilidades $1 - \delta$ y δ respectivamente. Definimos la variable

$$X = \begin{cases} X_1 & \text{si } U = 0 \\ X_2 & \text{si } U = 1 \end{cases}$$

Luego $F_X = (1 - \delta)F_1 + \delta F_2$.

Demostración. Teniendo en cuenta la independencia de las variables resulta que

$$\begin{aligned} F_X(x) &= P_X((-\infty, x]) \\ &= P(\{X \leq x\}) \\ &= P\left(\left(\{X_1 \leq x\} \cap \{U = 0\}\right) \cup \left(\{X_2 \leq x\} \cap \{U = 1\}\right)\right) \\ &= P(\{X_1 \leq x\} \cap \{U = 0\}) + P(\{X_2 \leq x\} \cap \{U = 0\}) \\ &= P(X_1 \leq x)P(U = 0) + P(X_2 \leq x)P(U = 1) \\ &= (1 - \delta)P(X_1 \leq x) + \delta P(X_2 \leq x) \\ &= (1 - \delta)F_1(x) + \delta F_2(x). \quad \square \end{aligned}$$

Capítulo 4

Vectores aleatorios.

4.1. Definición de vector aleatorio.

En muchos casos interesa estudiar simultáneamente más de una característica del resultado de un experimento aleatorio. Supongamos que el experimento consiste en elegir al azar alumnos de un determinado grado, y que estamos interesados en estudiar “el perfil biológico” de esos alumnos. Podríamos considerar que el perfil se compone de la talla, el peso, presión sanguínea, frecuencia cardíaca y capacidad respiratoria. Por lo tanto interesarían cinco variables aleatorias que deberían estudiarse simultáneamente. Esto motiva la siguiente definición de un vector aleatorio.

Definición 4.1 Sea (Ω, \mathcal{A}, P) un espacio de probabilidad. Se dice que $\mathbf{X} = (X_1, X_2, \dots, X_k)$ es un vector aleatorio de dimensión k si para cada $j = 1, 2, \dots, k$ se tiene que $X_j : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria.

Obsérvese que si $\mathbf{X} = (X_1, \dots, X_k)$ es un vector aleatorio de dimensión k , entonces también puede ser interpretado como una función $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k$. En efecto dado $\omega \in \Omega$, el correspondiente valor de la función es $\mathbf{X}(\omega) = (X_1(\omega), \dots, X_k(\omega)) \in \mathbb{R}^k$.

Teorema 4.1 Para todo $\mathbf{x} = (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ se tendrá

$$\mathbf{X}^{-1}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]) \in \mathcal{A}.$$

Demostración. Sea $B = (-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]$. Entonces

$$\begin{aligned} \mathbf{X}^{-1}(B) &= \{\omega \in \Omega : \mathbf{X}(\omega) \in B\} \\ &= \bigcap_{i=1}^k \{\omega \in \Omega : X_i(\omega) \in (-\infty, x_i]\} = \\ &= \bigcap_{i=1}^k X_i^{-1}((-\infty, x_i]). \end{aligned}$$

Luego como por definición de variable aleatoria para todo i se tiene que $X_i^{-1}((-\infty, x_i]) \in \mathcal{A}$ y \mathcal{A} es una σ -álgebra se concluye que $\mathbf{X}^{-1}(B) \in \mathcal{A}$. \square

Recordemos que \mathcal{B}^k denota la σ -álgebra generada por los conjuntos de \mathbb{R}^k de la forma

$$A_{x_1, x_2, \dots, x_k} = (-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_k]$$

En \mathbb{R}^2 es fácil verificar gráficamente que los conjuntos de la forma

$$(a_1, b_1] \times (a_2, b_2] \in \mathcal{B}^2$$

ya que se pueden escribir de la siguiente forma

$$(a_1, b_1] \times (a_2, b_2] = A_{b_1, b_2} - A_{a_1, b_2} - (A_{b_1, a_2} - A_{a_1, a_2}) \quad (4.1)$$

y que diferencias de conjuntos de una σ -álgebra son conjuntos de la σ -álgebra.

Va a ser útil observar que

$$A_{a_1, b_2} \subset A_{b_1, b_2} \quad (4.2)$$

$$A_{a_1, a_2} \subset A_{b_1, a_2} \quad (4.3)$$

y

$$(A_{b_1, a_2} - A_{a_1, a_2}) \subset A_{b_1, b_2} - A_{a_1, b_2}. \quad (4.4)$$

Ejercicio. Probar el siguiente teorema.

Teorema 4.2 *Sea \mathbf{X} un vector aleatorio de dimensión k . Entonces si $B \in \mathcal{B}^k$ se tiene que $\mathbf{X}^{-1}(B) \in \mathcal{A}$.*

4.2. Espacio de probabilidad inducido.

Definición 4.2 *Dado el espacio de probabilidad (Ω, \mathcal{A}, P) y un vector aleatorio $\mathbf{X} = (X_1, \dots, X_k)$ se puede definir un nuevo espacio de probabilidad $(\mathbb{R}^k, \mathcal{B}^k, P_{\mathbf{X}})$ donde dado $B \in \mathcal{B}^k$ se define*

$$P_{\mathbf{X}}(B) = P(\mathbf{X}^{-1}(B)).$$

Ejercicio. Probar el siguiente teorema.

Teorema 4.3 *$P_{\mathbf{X}}$ es una función de probabilidad sobre $(\mathbb{R}^k, \mathcal{B}^k)$.*

La demostración es similar a la correspondiente a P_X donde X es una variable aleatoria. La probabilidad $P_{\mathbf{X}}$ se denomina probabilidad inducida por el vector \mathbf{X} o distribución de \mathbf{X} .

4.3. Función de distribución conjunta de un vector aleatorio.

Definición 4.3 Dado un vector aleatorio $\mathbf{X} = (X_1, \dots, X_k)$, se define la función de distribución conjunta del vector \mathbf{X} como la función $F_{\mathbf{X}} : \mathbb{R}^k \rightarrow [0; 1]$ dada por

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2, \dots, x_k) &= P_{\mathbf{X}}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]) = \\ &= P\left(\bigcap_{i=1}^k \{\omega : X_i(\omega) \leq x_i\}\right). \end{aligned}$$

Propiedades de $F_{\mathbf{X}}$.

Propiedad 4.1 $F_{\mathbf{X}}$ es monótona no decreciente en cada componente.

Demostración. Si $x_i < x'_i$ entonces

$$A_{x_1, \dots, x_i, \dots, x_n} \subset A_{x_1, \dots, x'_i, \dots, x_n},$$

de manera que

$$F_{\mathbf{X}}((x_1, \dots, x_i, \dots, x_n)) \leq F_{\mathbf{X}}((x_1, \dots, x'_i, \dots, x_n)). \quad \square$$

Propiedad 4.2 Se tiene que

$$\lim_{x_1 \rightarrow \infty, \dots, x_k \rightarrow \infty} F_{\mathbf{X}}(x_1, x_2, \dots, x_k) = 1.$$

Demostración. Sean sucesiones crecientes

$$\{x_{1i}\}_{i \in \mathbb{N}} \uparrow \infty, \{x_{2i}\}_{i \in \mathbb{N}} \uparrow \infty, \dots, \{x_{ki}\}_{i \in \mathbb{N}} \uparrow \infty.$$

Queremos probar que

$$\lim_{i \rightarrow +\infty} F_{\mathbf{X}}(x_{1i}, x_{2i}, \dots, x_{ki}) = 1.$$

Ahora bien la sucesión de conjuntos

$$C_i = (-\infty, x_{1i}] \times (-\infty, x_{2i}] \times \dots \times (-\infty, x_{ki}] \quad (4.5)$$

es monótona no decreciente. Por otro lado

$$\bigcup_{i \in \mathbb{N}} C_i = \mathbb{R}^k,$$

y en consecuencia

$$\begin{aligned} \lim_{i \rightarrow +\infty} F_{\mathbf{X}}(x_{1i}, x_{2i}, \dots, x_{ki}) &= \lim_{i \rightarrow \infty} P_{\mathbf{X}}((-\infty, x_{1i}] \times (-\infty, x_{2i}] \times \cdots \times (-\infty, x_{ki})) = \\ &= P_{\mathbf{X}}\left(\bigcup_{i \in \mathbb{N}} C_i\right) = P_{\mathbf{X}}(\mathbb{R}^k) = 1. \quad \square \end{aligned}$$

Propiedad 4.3 Para todo i , $1 \leq i \leq k$, se tiene que

$$\lim_{x_i \rightarrow -\infty} F_{\mathbf{X}}(x_1, x_2, \dots, x_i, \dots, x_k) = 0.$$

Demostración. Sin pérdida de generalidad lo mostraremos para $i = 1$. Para este caso consideremos una sucesión monótona no creciente tal que $\{y_j\}_{j \in \mathbb{N}} \downarrow -\infty$.

Entonces si definimos $\{C_j\}_{j \in \mathbb{N}}$ por

$$C_j = (-\infty, y_j] \times (-\infty, x_2] \times \cdots \times (-\infty, x_k] \quad (4.6)$$

se tiene que $C_{j+1} \subset C_j$ para todo $j \in \mathbb{N}$, y además

$$\bigcap_{j \in \mathbb{N}} C_j = \emptyset.$$

Por lo tanto

$$\begin{aligned} \lim_{j \rightarrow -\infty} F_{\mathbf{X}}(y_j, x_2, \dots, x_k) &= \lim_{j \rightarrow \infty} P_{\mathbf{X}}((-\infty, y_j] \times (-\infty, x_2] \times \cdots \times (-\infty, x_k]) = \\ &= P_{\mathbf{X}}\left(\bigcap_{j \in \mathbb{N}} C_j\right) \\ &= P_{\mathbf{X}}(\emptyset) \\ &= 0. \quad \square \end{aligned}$$

Propiedad 4.4 $F_{\mathbf{X}}$ es continua a derecha.

Demostración. Sea $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ y consideremos sucesiones monótonas decrecientes tales que

$$\{x_{1i}\}_{i \in \mathbb{N}} \downarrow x_1; \{x_{2i}\}_{i \in \mathbb{N}} \downarrow x_2; \dots; \{x_{ki}\}_{i \in \mathbb{N}} \downarrow x_k$$

Consideremos los conjuntos

$$C_i = (-\infty, x_{1i}] \times (-\infty, x_{2i}] \times \cdots \times (-\infty, x_{ki}].$$

Entonces

$$C_{i+1} \subset C_i$$

y

$$\bigcap_{i \in \mathbb{N}} C_i = A_{x_1, \dots, x_k}.$$

Luego

$$\begin{aligned} \lim_{i \rightarrow \infty} F_{\mathbf{X}}(x_{1i}, x_{2i}, \dots, x_{ki}) &= \lim_{i \rightarrow \infty} P(C_i) \\ &= P(A_{x_1, \dots, x_k}) \\ &= F_{\mathbf{X}}(x_1, x_2, \dots, x_k). \quad \square \end{aligned}$$

Las Propiedades 4.1, 4.2, 4.3 y 4.4 no caracterizan a una función de distribución de un vector aleatorio como ocurría para el caso de la función de distribución de una variable aleatoria.

Para fijar ideas de por qué sucede esto, pensemos en \mathbb{R}^2 . Sea entonces un vector aleatorio en \mathbb{R}^2 $\mathbf{X} = (X_1, X_2)$ y $F_{\mathbf{X}}$ su función de distribución conjunta. Sea $A_{x_1 x_2} = (-\infty, x_1] \times (-\infty, x_2]$ y $C = (a_1, b_1] \times (a_2, b_2]$.

El rectángulo C puede ser escrito de la siguiente manera

$$C = (A_{b_1 b_2} - A_{a_1 b_2}) - (A_{b_1 a_2} - A_{a_1 a_2}).$$

Teniendo en cuenta las inclusiones

$$A_{a_1 a_2} \subset A_{b_1 a_2}, \quad (4.7)$$

$$A_{a_1 b_2} \subset A_{b_1 b_2} \quad (4.8)$$

y

$$(A_{b_1 a_2} - A_{a_1 a_2}) \subset (A_{b_1 b_2} - A_{a_1 b_2}), \quad (4.9)$$

resulta que

$$\begin{aligned} P_{\mathbf{X}}(C) &= P_{\mathbf{X}}(A_{b_1 b_2} - A_{a_1 b_2}) - P_{\mathbf{X}}(A_{b_1 a_2} - A_{a_1 a_2}) \\ &= P_{\mathbf{X}}(A_{b_1 b_2}) - P_{\mathbf{X}}(A_{a_1 b_2}) - P_{\mathbf{X}}(A_{b_1 a_2}) + P_{\mathbf{X}}(A_{a_1 a_2}). \end{aligned}$$

Como $P_{\mathbf{X}}(A_{x_1 x_2}) = F_{\mathbf{X}}(x_1, x_2)$, resulta

$$P_{\mathbf{X}}(C) = F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(b_1, a_2) + F_{\mathbf{X}}(a_1, a_2).$$

Observaciones.

1. Para verificar las inclusiones (4.7), (4.8) y (4.9), se sugiere hacer un dibujo.

2. Esto muestra que la probabilidad de el rectángulo C se determina por el valor de $F_{\mathbf{X}}$ sobre los vértices: es la suma de los valores sobre los vértices de la diagonal principal menos la suma de los valores sobre los vértices de la otra diagonal.
3. Luego dada una función de distribución F_X para todo $a_1 < b_1$ y $a_2 < b_2$ se debería cumplir

$$F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(b_1, a_2) + F_{\mathbf{X}}(a_1, a_2) \geq 0. \quad (4.10)$$

4. Veamos que esta propiedad no se deduce de las propiedades P1, P2, P3 y P4. Para ello damos un ejemplo de una función que satisface P1, P2, P3 y P4 pero no (4.10). Sea $F : \mathbb{R}^2 \rightarrow [0, 1]$ definida por

$$F(x_1, x_2) = \begin{cases} 1 & \text{si } x_1 + x_2 \geq 1, x_1 \geq 0, x_2 \geq 0 \\ 0 & \text{si en otra parte.} \end{cases}$$

Es fácil verificar que esta función es (i) monótona no decreciente en cada variable, (ii)

$$\lim_{x_1 \rightarrow \infty, x_2 \rightarrow \infty} F(x_1, x_2) = 1,$$

(iii)

$$\lim_{x_i \rightarrow -\infty} F(x_1, x_2) = 0 \text{ para cualquier } i = 1, 2,$$

y (iv) es continua a derecha. Pero si consideramos el rectángulo $C = (0, 1] \times (0, 1]$ entonces si F es una función de distribución deberíamos tener

$$P(C) = F(1, 1) + F(0, 0) - (F(0, 1) + F(1, 0)) = 1 - 2 = -1.$$

Esto muestra que F no puede ser la función de distribución de ningún vector aleatorio en \mathbb{R}^2 .

Para estudiar las propiedades faltantes vamos a necesitar la siguiente definición.

Definición 4.4 Sea F una función de k variables. Si $a_i < b_i$ se define el operador diferencia en la variable i por

$$\Delta_i(a, b)F = F(x_1, x_2, \dots, x_{i-1}, b, x_{i+1}, \dots, x_k) - F(x_1, x_2, \dots, x_{i-1}, a, x_{i+1}, \dots, x_k).$$

Estos operadores se pueden aplicar en forma sucesiva. Por ejemplo

$$\begin{aligned}
& \Delta_j (a_j, b_j) \Delta_i (a_i, b_i) F \\
&= \Delta_j (a_j, b_j) (F(x_1, \dots, x_{i-1}, b_i, x_{i+1}, \dots, x_k) \\
&\quad - F(x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots, x_k)) \\
&= \Delta_j (a_j, b_j) F(x_1, x_2, \dots, x_{i-1}, b_i, x_{j+1}, \dots, x_k) \\
&\quad - \Delta_j (a_j, b_j) F(x_1, x_2, \dots, x_{i-1}, a_i, x_{j+1}, \dots, x_k) \\
&= (F(x_1, \dots, x_{i-1}, b_i, x_{i+1}, \dots, x_{j-1}, b_j, x_{j+1}, \dots, x_k) \\
&\quad - F(x_1, \dots, x_{i-1}, b_i, x_{i+1}, \dots, x_{j-1}, a_j, x_{j+1}, \dots, x_k)) \\
&\quad - (F(x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots, x_{j-1}, b_j, x_{j+1}, \dots, x_k) \\
&\quad - F(x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots, x_{j-1}, a_j, x_{j+1}, \dots, x_k)).
\end{aligned}$$

Es fácil ver que estos operadores conmutan, es decir

$$\Delta_j (a_j, b_j) \Delta_i (a_i, b_i) F = \Delta_i (a_i, b_i) \Delta_j (a_j, b_j) F$$

Más generalmente, si $a_1 < b_1$, $a_2 < b_2$, \dots , $a_k < b_k$ podemos considerar la diferencia sucesiva

$$\Delta_1 (a_1, b_1) \cdots \Delta_{k-1} (a_{k-1}, b_{k-1}) \Delta_k (a_k, b_k).$$

Observación. Podemos expresar la propiedad (4.10) en términos del operador diferencia como

$$\begin{aligned}
P_{\mathbf{X}}((a_1, b_1] \times (a_2, b_2]) &= (F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(a_1, b_2)) - (F_{\mathbf{X}}(b_1, a_2) - F_{\mathbf{X}}(a_1, a_2)) \\
&= \Delta_1 (b_1, a_1) F_{\mathbf{X}}(x_1, b_2) - \Delta_1 (b_1, a_1) F_{\mathbf{X}}(x_1, a_2) \\
&= \Delta_2 (b_2, a_2) \Delta_1 (b_1, a_1) F_{\mathbf{X}}(x_1, x_2) \geq 0
\end{aligned}$$

En general se puede probar el siguiente Teorema

Teorema 4.4 *Sea $F_{\mathbf{X}}$ la función de distribución conjunta del vector aleatorio $\mathbf{X} = (X_1, \dots, X_k)$ y sean $a_1 < b_1$, $a_2 < b_2, \dots, a_k < b_k$. Entonces se tiene que*

$$\begin{aligned}
& P_{\mathbf{X}}((a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_k, b_k]) \\
&= \Delta_1 (b_1, a_1) \cdots \Delta_{k-1} (b_{k-1}, a_{k-1}) \Delta_k (b_k, a_k) F_{\mathbf{X}}(x_1, x_2, \dots, x_k) \geq 0.
\end{aligned}$$

Demostración. Para probar el teorema, consideremos para cada h , $0 \leq h \leq k$ los conjuntos de la forma

$$C_h = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_h, b_h] \times (-\infty, x_{h+1}] \times \cdots \times (-\infty, x_k].$$

Se prueba por inducción que para todo $h \leq k$

$$P_{\mathbf{X}}(C_h) = \Delta_1 (b_1, a_1) \cdots \Delta_{h-1} (b_{h-1}, a_{h-1}) \Delta_h (b_h, a_h) F(x_1, x_2, \dots, x_h, x_{h+1}, \dots, x_k). \quad (4.11)$$

Probaremos primero (4.11) para $h = 1$. Sea

$$C_1 = (a_1, b_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_k].$$

Luego

$$C_1 = (-\infty, b_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_k] - (-\infty, a_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_k],$$

y como el segundo conjunto está incluido en el primero, se tiene

$$\begin{aligned} P_{\mathbf{X}}(C_1) &= P_{\mathbf{X}}((-\infty, b_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_k] - (-\infty, a_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_k]) \\ &= F_{\mathbf{X}}(b_1, x_2, \dots, x_k) - F_{\mathbf{X}}(a_1, x_2, \dots, x_k) \\ &= \Delta_1(b_1, a_1) F(x_1, x_2, \dots, x_k). \end{aligned}$$

Supongamos ahora que (4.11) vale para $h = i < k$. Probaremos que también vale para $h = i + 1$. Sea

$$C_{i+1} = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_{i+1}, b_{i+1}] \times (-\infty, x_{i+2}] \times \cdots \times (-\infty, x_k].$$

Claramente $C_{i+1} = C_i^{(2)} - C_i^{(1)}$, donde

$$C_i^{(1)} = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_i, b_i] \times (-\infty, a_{i+1}] \times (-\infty, x_{i+2}] \times \cdots \times (-\infty, x_k]$$

$$\text{y } C_i^{(2)} = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_i, b_i] \times (-\infty, b_{i+1}] \times (-\infty, x_{i+2}] \times \cdots \times (-\infty, x_k].$$

Como además se tiene $C_i^{(1)} \subset C_i^{(2)}$, se tendrá

$$P_{\mathbf{X}}(C_{i+1}) = P_{\mathbf{X}}(C_i^{(2)}) - P_{\mathbf{X}}(C_i^{(1)}).$$

Como (4.11) vale para $h = i$ tendremos

$$\begin{aligned} P_{\mathbf{X}}(C_{i+1}) &= \Delta_1(b_1, a_1) \dots \Delta_i(b_i, a_i) F(x_1, x_2, \dots, x_i, b_{i+1}, x_{i+2}, \dots, x_k) \\ &\quad - \Delta_1(b_1, a_1) \dots \Delta_i(b_i, a_i) F(x_1, x_2, \dots, x_i, a_{i+1}, x_{i+2}, \dots, x_k). \end{aligned}$$

Luego (4.11) vale para $h = i + 1$. Esto muestra que (4.11) vale para todo $h \leq k$. Haciendo $h = k$ se obtiene el Teorema. \square

Luego podemos enunciar una propiedad adicional que satisface una función de distribución conjunta

Propiedad 4.5 Si $F_{\mathbf{X}}$ es la función de distribución conjunta del vector aleatorio $\mathbf{X} = (X_1, \dots, X_k)$ para todo $a_1 < b_1, \dots, a_k < b_k$ se debe cumplir que

$$\Delta_1(b_1, a_1) \dots \Delta_{k-1}(b_{k-1}, a_{k-1}) \Delta_k(b_k, a_k) F_{\mathbf{X}}(x_1, x_2, \dots, x_k) \geq 0.$$

El siguiente Teorema generaliza para vectores aleatorios el Teorema de Extensión para variables aleatorias.

Teorema 4.5 Sea $F : \mathbb{R}^k \rightarrow [0, 1]$ una función que satisface las propiedades 4.1, 4.2, 4.3, 4.4 y 4.5. Luego existe una única función de probabilidad $P : \mathcal{B}^k \rightarrow [0, 1]$, tal que para todo $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ se cumple

$$P((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]) = F(x_1, x_2, \dots, x_k).$$

Demostración. No se dará la demostración en este curso. Utiliza argumentos de la Teoría de la Medida. \square

Corolario 4.1 Sean $\mathbf{X} = (X_1, X_2, \dots, X_k)$ y $\mathbf{X}^* = (X_1^*, X_2^*, \dots, X_k^*)$ dos vectores aleatorios. Supongamos que para todo x_1, x_2, \dots, x_k se tiene que

$$F_{\mathbf{X}}(x_1, \dots, x_k) = F_{\mathbf{X}^*}(x_1, \dots, x_k).$$

Luego también se cumple que para todo $B \in \mathcal{B}^k$

$$P_{\mathbf{X}}(B) = P_{\mathbf{X}^*}(B).$$

Demostración. Basta con observar que para todo $(x_1, \dots, x_k) \in \mathbb{R}^k$

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2, \dots, x_k) &= F_{\mathbf{X}^*}(x_1, x_2, \dots, x_k) \\ &= P_{\mathbf{X}^*}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]). \end{aligned}$$

Por lo tanto como $P_{\mathbf{X}}$ y $P_{\mathbf{X}^*}$ son extensiones de $F_{\mathbf{X}}$ deben coincidir por unicidad de la extensión. \square

Corolario 4.2 Si F satisface propiedades 4.1, 4.2, 4.3, 4.4 y 4.5. entonces existe un vector aleatorio $\mathbf{X} = (X_1, \dots, X_k)$ tal que

$$F_{\mathbf{X}} = F.$$

Demostración. Sea $(\mathbb{R}^k, \mathcal{B}^k, P_F)$ el espacio de probabilidad tal que P_F es la extensión de F . Luego para todo $(x_1, \dots, x_k) \in \mathbb{R}^k$

$$F(x_1, x_2, \dots, x_k) = P_F((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]).$$

Definimos el vector aleatorio $\mathbf{X} = (X_1, \dots, X_i, \dots, X_k)$ de forma tal que X_i sea la proyección sobre la coordenada i -ésima. Es decir $X_i : \mathbb{R}^k \rightarrow \mathbb{R}$ está definida por

$$X_i(x_1, x_2, \dots, x_k) = x_i$$

Observemos que para todo i , $1 \leq i \leq k$ se tiene que

$$X_i^{-1}((-\infty, x_i]) = \mathbb{R} \times \dots \times \mathbb{R} \times (-\infty, x_i] \times \mathbb{R} \times \dots \times \mathbb{R},$$

y que

$$\begin{aligned}
& F_{\mathbf{X}}(x_1, x_2, \dots, x_k) \\
&= P_{\mathbf{X}}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]) \\
&= P_F(\mathbf{X}^{-1}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k])) \\
&= P_F\left(\bigcap_{i=1}^k X_i^{-1}((-\infty, x_i])\right) \\
&= P_F((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]) \\
&= F(x_1, x_2, \dots, x_k). \quad \square
\end{aligned}$$

4.4. Algunas propiedades de vectores aleatorios.

Sea un vector $\mathbf{X} = (X_1, \dots, X_k)$ con función de distribución $F_{\mathbf{X}}$. El siguiente teorema muestra como se obtiene la función de distribución del vector formado con un subconjunto de componentes $\tilde{\mathbf{X}} = (X_{i_1}, X_{i_2}, \dots, X_{i_h})$ para cualquier subconjunto de índices $1 \leq i_1 < i_2 < \dots < i_h \leq k$.

Teorema 4.6 *Sea $\mathbf{X} = (X_1, \dots, X_k)$ un vector aleatorio de dimensión k . Sea $A = \{i_1, \dots, i_h\} \subset \{1, 2, \dots, k\}$ y $B = \{i : 1 \leq i \leq k, i \notin A\} = \{j_1, \dots, j_r\}$. Entonces, si $\tilde{\mathbf{X}} = (X_{i_1}, X_{i_2}, \dots, X_{i_h})$, se tiene*

$$F_{\tilde{\mathbf{X}}}(x_{i_1}, \dots, x_{i_h}) = \lim_{x_{j_1} \rightarrow \infty, \dots, x_{j_r} \rightarrow \infty} F_{\mathbf{X}}(x_1, x_2, \dots, x_k).$$

Demostración. Para facilitar la notación supongamos que $A = \{1, 2, \dots, h\}$ y luego $B = \{h+1, \dots, k\}$. Sean $\{y_{h+1,j}\}_{j \in \mathbb{N}}, \dots, \{y_{k,j}\}_{j \in \mathbb{N}}$, sucesiones crecientes tendiendo a ∞ . Luego bastará probar que

$$\lim_{j \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_h, y_{h+1,j}, \dots, y_{k,j}) = F_{\tilde{\mathbf{X}}}(x_1, \dots, x_h). \quad (4.12)$$

Consideremos la sucesión de eventos

$$C_j = (-\infty, x_1] \times \dots \times (-\infty, x_h] \times (-\infty, y_{h+1,j}] \times \dots \times (-\infty, y_{k,j}]$$

es creciente y

$$\bigcup_{j=1}^{\infty} C_j = (-\infty, x_1] \times \dots \times (-\infty, x_h] \times \mathbb{R} \times \dots \times \mathbb{R}.$$

Luego

$$\begin{aligned}
F_{\bar{\mathbf{X}}}(x_1, \dots, x_h) &= P_{\bar{\mathbf{X}}}((-\infty, x_1] \times \dots \times (-\infty, x_h]) \\
&= P\left(\bigcap_{i=1}^h \{\omega : X_i(\omega) \leq x_i\}\right) \\
&= P\left(\left(\bigcap_{i=1}^h \{\omega : X_i(\omega) \leq x_i\}\right) \cap \left(\bigcap_{i=h+1}^k \{\omega : X_i(\omega) \in \mathbb{R}\}\right)\right) \\
&= P_{\mathbf{X}}((-\infty, x_1] \times \dots \times (-\infty, x_h] \times \mathbb{R} \times \dots \times \mathbb{R}) \\
&= \lim_{j \rightarrow \infty} P_{\mathbf{X}}(C_j) \\
&= \lim_{j \rightarrow \infty} P_{\mathbf{X}}((-\infty, x_1] \times \dots \times (-\infty, x_h] \times (-\infty, y_{h+1,j}] \times \dots \times (-\infty, y_{k,j}]) \\
&= \lim_{j \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_h, y_{h+1,j}, \dots, y_{k,j}).
\end{aligned}$$

y luego (4.12) vale. \square

Definición 4.5 Diremos que $g : \mathbb{R}^k \rightarrow \mathbb{R}$ es medible Borel si para todo $x \in \mathbb{R}$ se tiene que $g^{-1}((-\infty, x]) \in \mathcal{B}^k$.

Observación. Una función medible Borel puede interpretarse como una variable aleatoria en el espacio $(\mathbb{R}^k, \mathcal{B}^k)$. Como en este curso solo consideramos funciones medibles Borel, se las llamará simplemente funciones medibles

En particular se tendrá

Teorema 4.7 Si $g : \mathbb{R}^k \rightarrow \mathbb{R}$ es continua entonces g es medible.

Demostración. Siendo $(-\infty, x]$ cerrado se tiene que $g^{-1}((-\infty, x]) \in \mathcal{B}^k$ y por lo tanto es medible. \square

Ejercicio. Probar el siguiente teorema.

Teorema 4.8 Sea $\mathbf{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio sobre un espacio de probabilidad (Ω, \mathcal{A}, P) y $g : \mathbb{R}^k \rightarrow \mathbb{R}$ una función medible. Entonces $Y = g(\mathbf{X}) : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria.

Ahora podemos probar lo siguiente.

Teorema 4.9 Si X e Y son variables aleatorias, entonces

(i) $Z = X + Y$ es una variable aleatoria.

(ii) $Z = XY$ es una variable aleatoria.

(iii) Si $P(Y = 0) = 0$ entonces $Z = X/Y$ es una variable aleatoria.

Demostración. Se trata de escribir a Z como imagen de X e Y usando una función g medible.

(i) Definimos $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, $g(x, y) = x + y$. Como g es continua es medible. Luego si tomamos $\mathbf{W} = (X, Y)$ se tiene que $Z = g(\mathbf{W}) = X + Y$ es una variable aleatoria.

(ii) y (iii) La demostración de (ii) y (iii) se deja como ejercicio. \square

Definición 4.6 Sea $g : \mathbb{R}^k \rightarrow \mathbb{R}^h$, es decir $g = (g_1, g_2, \dots, g_h)$ tal que para cada $j = 1, 2, \dots, h$, $g_j : \mathbb{R}^k \rightarrow \mathbb{R}$.

Diremos que g es medible sii g_j es medible para cada $j = 1, 2, \dots, h$.

Teorema 4.10 Sea $\mathbf{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio y $g : \mathbb{R}^k \rightarrow \mathbb{R}^j$ una función medible. Entonces $\mathbf{Z} = g(\mathbf{X})$ es un vector aleatorio de dimensión j .

Demostración. Se deja como ejercicio. \square

4.5. Independencia de variables aleatorias.

4.5.1. Algunas consideraciones heurísticas.

Hemos visto con anterioridad lo que significaba la independencia de eventos. Brevemente recordemos que una familia de eventos es independiente si la ocurrencia de algunos de ellos no incide sobre la probabilidad de ocurrencia del otro. Más precisamente, un conjunto de eventos A_1, A_2, \dots, A_k son independientes si para toda elección $1 \leq i_1 < i_2 < \dots < i_h \leq k$

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_h}) = \prod_{j=1}^h P(A_{i_j}).$$

Ahora queremos definir la independencia de un conjunto de variables aleatorias. Queremos dar respuesta a la pregunta ¿en qué medida la información referida a una variable aleatoria X incide en el conocimiento de los valores de la variable aleatoria Y . Por ejemplo ¿la “inflación” y la “emisión monetaria” son independientes? ¿El peso de un individuo y su presión sanguínea son independientes? Para definir el concepto de independencia de variables aleatorias utilizaremos la noción de independencia de eventos.

Definición 4.7 Sean X_1, X_2, \dots, X_k variables aleatorias, definidas sobre un mismo espacio de probabilidad (Ω, \mathcal{A}, P) . Diremos que dichas variables son independientes si cualquiera sean los conjuntos $B_1, B_2, \dots, B_k \in \mathcal{B}$ (Borelianos en \mathbb{R}), los eventos $X_j^{-1}(B_j)$, $j = 1, 2, \dots, k$ son independientes.

Los dos siguientes teoremas dan caracterizaciones de la propiedad de independencia de un conjunto de variables aleatorias.

Teorema 4.11 Las variables aleatorias X_1, \dots, X_k son independientes si y sólo si para toda elección de conjuntos borelianos B_1, B_2, \dots, B_k vale que

$$P\left(\bigcap_{j=1}^k X_j^{-1}(B_j)\right) = \prod_{j=1}^k P\left(X_j^{-1}(B_j)\right). \quad (4.13)$$

Demostración. Primero mostraremos que (4.13) es una condición necesaria. En efecto, si X_1, \dots, X_k son independientes, (4.13) debe cumplirse por definición de independencia de eventos. Ahora probaremos la suficiencia de (4.13). Debemos probar que (4.13) implica para cualquier subconjunto de índices $i_1 < i_2 < \dots < i_h$, $h \leq k$ que

$$P\left(\bigcap_{j=1}^h X_{i_j}^{-1}(B_{i_j})\right) = \prod_{j=1}^h P\left(X_{i_j}^{-1}(B_{i_j})\right).$$

Consideremos los conjuntos C_i , $1 \leq i \leq k$, definidos de la siguiente manera

$$C_i = \begin{cases} B_i & \text{si } i \text{ coincide con algún } i_j \\ \mathbb{R} & \text{en caso contrario.} \end{cases}$$

Entonces dado que $X_i^{-1}(\mathbb{R}) = \Omega$ y $P(\Omega) = 1$, se tiene que

$$\begin{aligned} P\left(\bigcap_{j=1}^h X_{i_j}^{-1}(B_{i_j})\right) &= P\left(\bigcap_{i=1}^k X_i^{-1}(C_i)\right) \\ &= \prod_{j=1}^k P\left(X_i^{-1}(C_i)\right) \\ &= \prod_{j=1}^h P\left(X_{i_j}^{-1}(B_{i_j})\right). \quad \square \end{aligned}$$

Ahora escribiremos la misma proposición de otra manera

Teorema 4.12 *Las variables aleatorias X_1, \dots, X_k son independientes si y sólo si para toda colección de borelianos B_1, B_2, \dots, B_k vale que*

$$P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_k) = \prod_{j=1}^k P_{X_j}(B_j),$$

donde $\mathbf{X} = (X_1, X_2, \dots, X_k)$.

Demostración. Como $P_{X_j}(B_j) = P(X_j^{-1}(B_j))$ por el Teorema 4.11 bastará mostrar que

$$P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_k) = P\left(\bigcap_{j=1}^k X_j^{-1}(B_j)\right).$$

Para eso observamos que

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_k) &= P(\mathbf{X}^{-1}(B_1 \times B_2 \times \cdots \times B_k)) \\ &= P_{\mathbf{X}}(\{\omega : \mathbf{X}(\omega) \in B_1 \times B_2 \times \cdots \times B_k\}) \\ &= P_{\mathbf{X}}(\{\omega : (X_1(\omega), X_2(\omega), \dots, X_k(\omega)) \in B_1 \times B_2 \times \cdots \times B_k\}) \\ &= P\left(\bigcap_{j=1}^k \{\omega : X_j(\omega) \in B_j\}\right) \\ &= P\left(\bigcap_{j=1}^k X_j^{-1}(B_j)\right). \quad \square \end{aligned}$$

El siguiente teorema, da una condición necesaria y suficiente para la independencia de un conjunto de variables que es más simple de verificar.

Teorema 4.13 *Una condición necesaria y suficiente para que las variables aleatorias X_1, X_2, \dots, X_k sean independientes es que para todo $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ se cumpla que*

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_k) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_k}(x_k), \quad (4.14)$$

donde $\mathbf{X} = (X_1, X_2, \dots, X_k)$.

Demostración.

Para ver que (4.14) es una condición necesaria para la independencia de X_1, \dots, X_k , basta aplicar el Teorema 4.12 a los conjuntos

$$B_1 = (-\infty, x_1], B_2 = (-\infty, x_2], \dots, B_k = (-\infty, x_k].$$

Probaremos ahora la suficiencia. Consideremos los conjuntos del tipo

$$B_1 \times B_2 \times \cdots \times B_r \times (-\infty, x_{r+1}] \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k],$$

donde $B_1 \times B_2 \times \cdots \times B_r$ son borelianos en \mathbb{R} . Probaremos por inducción sobre r que vale la siguiente propiedad que llamamos A_r :

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times (-\infty, x_{r+1}] \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k]) \\ = P_{X_1}(B_1) \cdots P_{X_r}(B_r) P_{X_{r+1}}((-\infty, x_{r+1}]) \cdots P_{X_k}((-\infty, x_k]). \end{aligned} \quad (4.15)$$

Para $r = 0$, la condición (4.15) vale por hipótesis, puesto que se reduce a un producto de semirectas. Supongamos que vale para r y probemos que vale para $r + 1$. En primer lugar probemos que si (4.15) vale para r , también vale reemplazando $(-\infty, x_{r+1}]$ por \mathbb{R} , esto es

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times \mathbb{R} \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k]) \\ = P_{X_1}(B_1) \cdots P_{X_r}(B_r) P_{X_{r+1}}(\mathbb{R}) \cdots P_{X_k}((-\infty, x_k]) = \\ = P_{X_1}(B_1) \cdots P_{X_r}(B_r) P_{X_{r+2}}((-\infty, x_{r+2}]) \cdots P_{X_k}((-\infty, x_k]). \end{aligned} \quad (4.16)$$

Para mostrar esto podemos considerar una sucesión creciente de semirectas $C_n = (-\infty, n]$. Luego

$$\mathbb{R} = \bigcup_{n=1}^{\infty} C_n$$

y la sucesión $\{B_1 \times B_2 \times \cdots \times B_r \times C_n \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k]\}$, $n = 1, 2, \dots$ es monótona no decreciente en \mathbb{R}^k y vale

$$\begin{aligned} \bigcup_{n \in \mathbb{N}} B_1 \times B_2 \times \cdots \times B_r \times C_n \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k] \\ = B_1 \times B_2 \times \cdots \times B_r \times \mathbb{R} \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k] \end{aligned}$$

Luego usando que vale A_r tenemos que

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times \mathbb{R} \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k]) \\ = \lim_{n \rightarrow \infty} P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times C_n \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k]) \\ = \lim_{n \rightarrow \infty} P_{\mathbf{X}}(B_1) P_{\mathbf{X}}(B_2) \cdots P_{\mathbf{X}}(B_r) P_{\mathbf{X}}(C_n) P_{\mathbf{X}}((-\infty, x_{r+2}]) \cdots P_{\mathbf{X}}((-\infty, x_k]) \\ = P_{\mathbf{X}}(B_1) P_{\mathbf{X}}(B_2) \cdots P_{\mathbf{X}}(B_r) P_{\mathbf{X}}(\mathbb{R}) P_{\mathbf{X}}((-\infty, x_{r+2}]) \cdots P_{\mathbf{X}}((-\infty, x_k]), \end{aligned}$$

que es lo que queríamos probar.

Ahora probaremos A_{r+1} . Es decir debemos probar que dados borelianos B_1, \dots, B_{r+1} y reales x_{r+2}, \dots, x_k se tiene

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times B_{r+1} \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k]) \\ = P_{X_1}(B_1) \cdots P_{X_r}(B_r) P_{X_{r+1}}(B_{r+1}) \cdots P_{X_k}((-\infty, x_k]). \end{aligned} \quad (4.17)$$

Consideremos el conjunto

$$A = B_1 \times B_2 \times \cdots \times B_r \times \mathbb{R} \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k],$$

y distinguimos dos casos: (a) $P_{\mathbf{X}}(A) = 0$, (b) $P_{\mathbf{X}}(A) > 0$.

Consideremos primero el caso (a). Por (4.16)

$$\begin{aligned} 0 = P_{\mathbf{X}}(A) &= P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times \mathbb{R} \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k]) \\ &= P_{X_1}(B_1) \cdots P_{X_r}(B_r) P_{X_{r+1}}(\mathbb{R}) \cdots P_{X_k}((-\infty, x_k]) \end{aligned}$$

se tiene que

$$P_{\mathbf{X}}(B_i) = 0 \text{ para alg\u00fan } 1 \leq i \leq r$$

o bien

$$P_{X_i}((-\infty, x_i]) = 0 \text{ para alg\u00fan } r+2 \leq i \leq k.$$

En cualquiera de los dos casos el miembro derecho de (4.17) es 0.

Supongamos que $P_{\mathbf{X}}(B_i) = 0$ podemos suponer que $i = 1$, para fijar ideas. Entonces teniendo en cuenta que

$$B_1 \times B_2 \times \cdots \times B_r \times B_{r+1} \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k] \subset B_1 \times \mathbb{R} \times \cdots \times \mathbb{R},$$

obtenemos que

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times B_{r+1} \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k]) \\ \leq P_{\mathbf{X}}(B_1 \times \mathbb{R} \times \cdots \times \mathbb{R}) = P_{X_1}(B_1) = 0, \end{aligned}$$

y luego el miembro izquierdo de (4.17) tambi\u00e9n es 0 y la igualdad se cumple. Ahora si $P_{X_i}((-\infty, x_i]) = 0$, podemos suponer que $i = k$ y proceder de manera an\u00e1loga. Luego (4.17) vale para el caso (a).

Consideremos el caso (b), es decir que $P_{\mathbf{X}}(A) > 0$. Definimos un nuevo espacio de probabilidades $(\mathbb{R}, \mathcal{B}, P^*)$ de la siguiente manera: Para todo $B \in \mathcal{B}$ definimos

$$P^*(B) = \frac{P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times B \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k])}{P_{\mathbf{X}}(A)}.$$

Obs\u00e9rvese que los borelianos B_1, B_2, \dots, B_r y los reales x_{r+2}, \dots, x_k permanecen fijos cuando se cambia B . Veamos en primer lugar que efectivamente $P^* : \mathcal{B} \rightarrow [0, 1]$ es una probabilidad.

(i) Claramente

$$P^*(\mathbb{R}) = \frac{P_{\mathbf{X}}(A)}{P_{\mathbf{X}}(A)} = 1.$$

(ii) Supongamos que $(C_n)_{n \geq 1} \subset \mathcal{B}$ es una sucesión de borelianos disjuntos dos a dos. Entonces

$$\begin{aligned}
& P^* \left(\bigcup_{n \in \mathbb{N}} C_n \right) \\
&= \frac{P_{\mathbf{X}} \left(B_1 \times B_2 \times \cdots \times B_r \times \bigcup_{n \in \mathbb{N}} C_n \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k] \right)}{P_{\mathbf{X}}(A)} \\
&= \frac{P_{\mathbf{X}} \left(\bigcup_{n \in \mathbb{N}} (B_1 \times B_2 \times \cdots \times B_r \times C_n \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k]) \right)}{P_{\mathbf{X}}(A)} \\
&= \frac{\sum_{n=1}^{\infty} P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times C_n \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k])}{P_{\mathbf{X}}(A)} \\
&= \sum_{n=1}^{\infty} \frac{P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times C_n \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k])}{P_{\mathbf{X}}(A)} \\
&= \sum_{n=1}^{\infty} P^*(C_n).
\end{aligned}$$

Esto prueba que P^* es una probabilidad.

Observemos que en la deducción anterior se usó, además de que P es una probabilidad, una propiedad de la teoría de conjuntos, fácil de probar:

$$\begin{aligned}
& B_1 \times B_2 \times \cdots \times B_r \times \bigcup_{n \in \mathbb{N}} C_n \times (-\infty, x_{r+2}], \times \cdots \times (-\infty, x_k] \\
&= \bigcup_{n \in \mathbb{N}} (B_1 \times B_2 \times \cdots \times B_r \times C_n \times (-\infty, x_{r+2}], \times \cdots \times (-\infty, x_k]).
\end{aligned}$$

Ahora calcularemos el valor de P^* sobre una semirecta. Dado que A_r es válida (hipótesis inductiva), si $x \in \mathbb{R}$ se tiene

$$\begin{aligned}
& P^*((-\infty, x]) \\
&= \frac{P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times (-\infty, x] \times (-\infty, x_{r+2}], \times \cdots \times (-\infty, x_k])}{P_{\mathbf{X}}(A)} \\
&= \frac{P_{X_1}(B_1) \cdots P_{X_r}(B_r) P_{X_{r+1}}((-\infty, x]) \cdots P_{X_k}((-\infty, x_k])}{P_{X_1}(B_1) \cdots P_{X_r}(B_r) P_{X_{r+1}}(\mathbb{R}) \cdots P_{X_k}((-\infty, x_k])} \\
&= P_{X_{r+1}}((-\infty, x]).
\end{aligned}$$

Entonces por la unicidad de la extensión como $P_{X_{r+1}}$ y P^* coinciden en las semirectas $(-\infty, x]$ se tendrá por el Teorema de Extensión que para todo $B \in \mathcal{B}$,

$$P^*(B) = P_{X_{r+1}}(B).$$

En particular

$$P^*(B_{r+1}) = P_{X_{r+1}}(B_{r+1}),$$

y luego

$$P_{X_{r+1}}(B_{r+1}) = \frac{P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times B_{r+1} \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k])}{P_{X_1}(B_1) \cdots P_{X_r}(B_r) P_{X_{r+1}}(\mathbb{R}) \cdots P_{X_k}((-\infty, x_k])}.$$

Despejando de la ecuación anterior y usando que $P_{X_{r+1}}(\mathbb{R}) = 1$ obtenemos

$$\begin{aligned} & P_{\mathbf{X}}(B_1 \times B_2 \times \cdots \times B_r \times B_{r+1} \times (-\infty, x_{r+2}] \times \cdots \times (-\infty, x_k]) \\ &= P_{X_{r+1}}(B_{r+1}) P_{X_1}(B_1) \cdots P_{X_r}(B_r) P_{X_{r+2}}(B_{r+2}) \cdots P_{X_k}((-\infty, x_k]) \\ &= P_{X_1}(B_1) \cdots P_{X_r}(B_r) P_{X_{r+1}}(B_{r+1}) \cdots P_{X_k}((-\infty, x_k]), \end{aligned}$$

y luego también vale A_{r+1} . \square

4.5.2. Conservación de la independencia por transformaciones.

El siguiente teorema prueba que la independencia se conserva por transformaciones.

Teorema 4.14 *Sea (Ω, \mathcal{A}, P) un espacio de probabilidad sean X_1, X_2, \dots, X_h variables aleatorias independientes. Si $g_j : \mathbb{R} \rightarrow \mathbb{R}$, $j = 1, 2, \dots, h$ son funciones medibles entonces $Y_1 = g_1(X_1), Y_2 = g_2(X_2), \dots, Y_h = g_h(X_h)$ también son variables aleatorias independientes.*

Demostración. Aplicamos la definición de independencia. Dados B_1, B_2, \dots, B_h borelianos arbitrarios queremos probar que los conjuntos

$$Y_1^{-1}(B_1), Y_2^{-1}(B_2), \dots, Y_h^{-1}(B_h)$$

son eventos independientes. Ahora bien para cada $j = 1, 2, \dots, h$ se tiene

$$Y_j^{-1}(B_j) = X_j^{-1}(g_j^{-1}(B_j)) = X_j^{-1}(C_j),$$

donde $C_j = g_j^{-1}(B_j)$. Como los C_j , $j = 1, 2, \dots, h$ son borelianos, la independencia de las variables X_j implica que los eventos $X_j^{-1}(C_j)$ son independientes. Luego las variables Y_1, \dots, Y_h son independientes. \square

4.5.3. Independencia de vectores aleatorios.

Definición 4.8 *Definición.* *Sea (Ω, \mathcal{A}, P) un espacio de probabilidad. Sean $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_h$ vectores aleatorios de dimensiones k_1, k_2, \dots, k_h respectivamente, esto es*

$$\mathbf{X}_i : \Omega \rightarrow \mathbb{R}^{k_i}, i = 1, 2, \dots, h$$

son vectores aleatorios. Diremos que el sistema de vectores es independiente si dados $B_1 \in \mathcal{B}^{k_1}, B_2 \in \mathcal{B}^{k_2}, \dots, B_h \in \mathcal{B}^{k_h}$, borelianos arbitrarios en sus respectivos espacios, los conjuntos $\mathbf{X}_j^{-1}(B_j)$, $j = 1, 2, \dots, h$ son eventos independientes.

Las siguientes dos proposiciones dan condiciones necesarias y suficientes para que un conjunto de vectores aleatorios sean independientes. Las dos condiciones son análogas a las obtenidas para variables aleatorias.

Propiedad 4.6 Una condición necesaria y suficiente para que el conjunto de vectores $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_h$, donde \mathbf{X}_i es de dimensión k_i sean independientes es que para todo $B_1 \in \mathcal{B}^{k_1}, B_2 \in \mathcal{B}^{k_2}, \dots, B_h \in \mathcal{B}^{k_h}$ se cumpla

$$P_{\tilde{\mathbf{X}}}(B_1 \times B_2 \times \dots \times B_h) = P_{\mathbf{X}_1}(B_1) P_{\mathbf{X}_2}(B_2) \dots P_{\mathbf{X}_h}(B_h),$$

donde $\tilde{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_h)$.

Demostración. Análoga a la demostración de la proposición correspondiente para variables aleatorias. \square

Propiedad 4.7 Una condición necesaria y suficiente para que un conjunto de vectores $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_h$ sean independientes es que para todo $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_h) \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \times \dots \times \mathbb{R}^{k_h}$ se tenga

$$F_{\tilde{\mathbf{X}}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_h) = F_{\mathbf{X}_1}(\mathbf{x}_1) F_{\mathbf{X}_2}(\mathbf{x}_2) \dots F_{\mathbf{X}_h}(\mathbf{x}_h),$$

donde $\tilde{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_h)$.

Demostración. Análoga a la demostración de la proposición correspondiente para variables aleatorias. \square

Propiedad 4.8 Sean $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_h$ un sistema de vectores aleatorios de dimensiones k_1, k_2, \dots, k_h respectivamente. Sean g_1, g_2, \dots, g_h funciones medibles, $g_i : \mathbb{R}^{k_i} \rightarrow \mathbb{R}^{j_i}$, $i = 1, 2, \dots, h$. Entonces los vectores aleatorios $\mathbf{Y}_1 = g_1(\mathbf{X}_1)$, $\mathbf{Y}_2 = g_2(\mathbf{X}_2)$, \dots , $\mathbf{Y}_h = g_h(\mathbf{X}_h)$ son independientes.

Demostración. Análoga a la demostración de la proposición correspondiente para variables aleatorias. \square

Capítulo 5

Vectores aleatorios discretos y continuos.

Tal como ocurre con las variables aleatorias, existen distintos tipos de vectores aleatorios.

5.1. Vectores aleatorios discretos.

Definición 5.1 Sea $\mathbf{X} = (X_1, X_2, \dots, X_h)$ un vector aleatorio. Si dice que \mathbf{X} es discreto o bien que tiene distribución discreta sii para cada $i = 1, 2, \dots, h$, X_i es un variable aleatoria discreta.

Esto implica, de acuerdo a lo estudiado, que para cada $i = 1, 2, \dots, h$ existe un conjunto finito o infinito numerable R_{X_i} tal que $P_{X_i}(R_{X_i}) = 1$.

La Propiedad 5.2 que enunciaremos en breve muestra que el conjunto

$$R_{\mathbf{X}}^* = R_{X_1} \times \dots \times R_{X_h}$$

es finito o infinito numerable y que $P_{\mathbf{X}}(R^*) = 1$.

Necesitamos previamente demostrar la siguiente propiedad

Propiedad 5.1 Sean A_1, \dots, A_h una sucesión finita de eventos tal que para todo i , $1 \leq i \leq h$, tal que $P(A_i) = 1$. Entonces

$$P\left(\bigcap_{i=1}^h A_i\right) = 1.$$

Demostración. Basta probar que la probabilidad del complemento es cero. Eso se sigue inmediatamente dado que la probabilidad es subaditiva y $P(A_i^c) = 0$. En efecto, se tiene

$$0 \leq P\left(\left(\bigcap_{i=1}^h A_i\right)^c\right) = P\left(\bigcup_{i=1}^h A_i^c\right) \leq \sum_{i=1}^h P(A_i^c) = 0.$$

Luego

$$P\left(\left(\bigcap_{i=1}^h A_i\right)\right) = 1 - P\left(\left(\bigcap_{i=1}^h A_i\right)^c\right) = 1. \quad \square$$

Observación. La Propiedad 5.1 también vale para una sucesión numerable de eventos y su demostración es análoga.

Propiedad 5.2 Sea $\mathbf{X} = (X_1, X_2, \dots, X_h)$ un vector aleatorio. Entonces el conjunto

$$R_{\mathbf{X}}^* = R_{X_1} \times \cdots \times R_{X_h}$$

es finito o infinito numerable y

$$P_{\mathbf{X}}(R_{\mathbf{X}}^*) = 1.$$

Demostración. $R_{\mathbf{X}}^*$ es a lo sumo numerable, porque un producto cartesiano finito de conjuntos a lo sumo numerables es a lo sumo numerable. Además

$$\{\omega: \mathbf{X}(\omega) \in R_{X_1} \times \cdots \times R_{X_h}\} = \bigcap_{i=1}^h \{\omega: X_i(\omega) \in R_{X_i}\}.$$

Luego por la Propiedad 5.1

$$\begin{aligned} P_{\mathbf{X}}(R_{\mathbf{X}}^*) &= P_{\mathbf{X}}(R_{X_1} \times \cdots \times R_{X_h}) = P(\{\omega: \mathbf{X}(\omega) \in R_{X_1} \times \cdots \times R_{X_h}\}) \\ &= P\left(\bigcap_{i=1}^h \{\omega: X_i(\omega) \in R_{X_i}\}\right) = 1, \end{aligned}$$

ya que $P(\{\omega: X_i(\omega) \in R_{X_i}\}) = P_{X_i}(R_{X_i}) = 1. \quad \square$

De manera análoga a como lo hicimos para una sola variable se puede buscar el mínimo conjunto que tiene probabilidad 1. Este conjunto puede ser distinto de $R_{\mathbf{X}}^*$.

Ejemplo 5.1 Consideremos un vector aleatorio $\mathbf{X} = (X_1, X_2)$ que asume los valores $\{(0, 0), (1, 1)\}$ con la misma probabilidad 0,5. De esto se deduce que las variables aleatorias X_1, X_2 a su vez asumen los valores 0 y 1 con probabilidad 0,5 para ambos. Ahora bien

$$R_{\mathbf{X}}^* = R_{X_1} \times R_{X_2} = \{(0, 0), (1, 1), (0, 1), (1, 0)\}.$$

Se ve que el conjunto $R_{\mathbf{X}}^*$ puede ser reducido a $R_{\mathbf{X}} = \{(0, 0), (1, 1)\}$.

Más generalmente si \mathbf{X} es un vector discreto de dimensión k , podemos considerar el conjunto de los átomos de la probabilidad,

$$R_{\mathbf{X}} = \{\mathbf{x}: P_{\mathbf{X}}(\{\mathbf{x}\}) > 0\} \subset R_{X_1} \times \cdots \times R_{X_h}.$$

El siguiente Teorema, cuya demostración es análoga al Teorema 3.1 muestra que $R_{\mathbf{X}}$ es el mínimo conjunto de probabilidad 1.

Teorema 5.1 *Se tiene que $P_{\mathbf{X}}(R_{\mathbf{X}}) = 1$. Además si $B \in \mathcal{B}^k$ es tal que $P_{\mathbf{X}}(B) = 1$, entonces $R_{\mathbf{X}} \subset B$.*

5.1.1. Función de densidad de probabilidad conjunta.

Una vez obtenido el conjunto $R_{\mathbf{X}}$ donde se concentra la probabilidad de un vector aleatorio discreto, vamos a mostrar que de igual manera que en el caso de una variable aleatoria, podemos determinar una función definida ahora sobre \mathbb{R}^k que determina totalmente a $P_{\mathbf{X}}$.

Definición 5.2 *Sea $\mathbf{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio discreto. Se define la función densidad de probabilidad conjunta $p_{\mathbf{X}} : \mathbb{R}^k \rightarrow [0, 1]$, asociada al vector \mathbf{X} por*

$$p_{\mathbf{X}}(\mathbf{x}) = P_{\mathbf{X}}(\{\mathbf{x}\}).$$

Observación. De acuerdo a la definición de $R_{\mathbf{X}}$ se tendrá

$$p_{\mathbf{X}}(\mathbf{x}) = \begin{cases} > 0 & \text{si } \mathbf{x} \in R_{\mathbf{X}} \\ 0 & \text{si } \mathbf{x} \notin R_{\mathbf{X}}. \end{cases}$$

Como consecuencia de las anteriores observaciones y de manera análoga a como lo hemos hecho para una sola variable se tiene el siguiente teorema.

Teorema 5.2 *Para todo $B \in \mathcal{B}^k$ se tiene*

$$\begin{aligned} P_{\mathbf{X}}(B) &= \sum_{\mathbf{x} \in B \cap R_{\mathbf{X}}} p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{\mathbf{x} \in B \cap R_{\mathbf{X}}^*} p_{\mathbf{X}}(\mathbf{x}). \end{aligned}$$

Muchas veces es conveniente considerar el conjunto $R_{\mathbf{X}}^* = R_{X_1} \times R_{X_2} \times \dots \times R_{X_k}$ en vez de $R_{\mathbf{X}}$.

Teorema 5.3 *Sea $B = B_1 \times B_2 \times \dots \times B_k$, donde B_1, \dots, B_k son borelianos en \mathbb{R} . Entonces*

(a)

$$P_{\mathbf{X}}(B) = \sum_{x_k \in B_k \cap R_{X_k}} \sum_{x_{k-1} \in B_{k-1} \cap R_{X_{k-1}}} \dots \sum_{x_1 \in B_1 \cap R_{X_1}} p_{\mathbf{X}}(x_1, x_2, \dots, x_k).$$

(b)

$$\sum_{x_k \in R_{X_k}} \sum_{x_{k-1} \in R_{X_{k-1}}} \dots \sum_{x_1 \in R_{X_1}} p_{\mathbf{X}}(\mathbf{x}) = 1.$$

Demostración.

$$\begin{aligned}
P_{\mathbf{X}}(B) &= \sum_{\mathbf{x} \in B \cap R_{\mathbf{X}}} p_{\mathbf{X}}(\mathbf{x}) \\
&= \sum_{\mathbf{x} \in B \cap R_{\mathbf{X}}^*} p_{\mathbf{X}}(\mathbf{x}) \\
&= \sum_{\mathbf{x} \in B \cap (R_{X_1} \times R_{X_2} \times \cdots \times R_{X_k})} p_{\mathbf{X}}(\mathbf{x}) \\
&= \sum_{\mathbf{x} \in B_1 \cap R_{X_1} \times B_2 \cap R_{X_2} \times \cdots \times B_k \cap R_{X_k}} p_{\mathbf{X}}(\mathbf{x}) \\
&= \sum_{x_k \in B_k \cap R_{X_k}} \sum_{x_{k-1} \in B_{k-1} \cap R_{X_{k-1}}} \cdots \sum_{x_1 \in B_1 \cap R_{X_1}} p_{\mathbf{X}}(x_1, x_2, \dots, x_k).
\end{aligned}$$

Luego (a) vale. En particular si tomamos $B_i = \mathbb{R}$, luego $B = \mathbb{R}^k$ y

$$\begin{aligned}
1 &= P_{\mathbf{X}}(\mathbb{R}^k) \\
&= \sum_{\mathbf{x} \in R_{X_1} \times R_{X_2} \times \cdots \times R_{X_k}} p_{\mathbf{X}}(\mathbf{x}) \\
&= \sum_{x_k \in R_{X_k}} \sum_{x_{k-1} \in R_{X_{k-1}}} \cdots \sum_{x_1 \in R_{X_1}} p_{\mathbf{X}}(\mathbf{x}),
\end{aligned}$$

y luego (b) vale. \square

5.1.2. Caracterización de la función de densidad marginal asociada a un subconjunto de variables.

Se trata de determinar a partir de la función de densidad conjunta, la marginal asociada a un subconjunto arbitrario de variables. Para fijar ideas, consideremos un vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_h, X_{h+1}, \dots, X_k)$ y un subvector $\mathbf{X}^* = (X_1, X_2, \dots, X_h)$.

Propiedad 5.3 *La función de densidad marginal asociada al vector \mathbf{X}^* viene dada por la fórmula*

$$p_{\mathbf{X}^*}(\mathbf{x}) = \sum_{x_{h+1} \in R_{X_{h+1}}} \sum_{x_{h+2} \in R_{X_{h+2}}} \cdots \sum_{x_k \in R_{X_k}} p_{\mathbf{X}}(x_1, \dots, x_h, x_{h+1}, \dots, x_k).$$

Demostración. Aplicando la definición de $p_{\mathbf{X}}$

$$\begin{aligned}
p_{\mathbf{X}^*}((x_1, x_2, \dots, x_h)) &= P_{\mathbf{X}^*}(\{(x_1, x_2, \dots, x_h)\}) \\
&= P_{\mathbf{X}}(\{\{x_1\} \times \{x_2\} \times \cdots \times \{x_h\} \times \mathbb{R} \times \cdots \times \mathbb{R}\}).
\end{aligned}$$

Entonces de acuerdo al resultado anterior

$$\begin{aligned}
p_{\mathbf{X}^*}((x_1, x_2, \dots, x_h)) &= P_{\mathbf{X}}(\{x_1\} \times \{x_2\} \times \dots \times \{x_h\} \times \mathbb{R} \times \dots \times \mathbb{R}) \\
&= \sum_{x_k \in \mathbb{R} \cap R_{X_k}} \dots \sum_{x_{h+1} \in \mathbb{R} \cap R_{X_{h+1}}} p_{\mathbf{X}}(x_1, \dots, x_h, x_{h+1}, \dots, x_k) \\
&= \sum_{x_k \in R_{X_k}} \dots \sum_{x_{h+1} \in R_{X_{h+1}}} p_{\mathbf{X}}(x_1, \dots, x_h, x_{h+1}, \dots, x_k). \quad \square
\end{aligned}$$

Ahora vamos a dar una condición necesaria y suficiente de independencia para el caso de variables aleatorias con distribución discreta, en términos de la función de densidad conjunta y sus marginales.

Para esto recordemos que una condición necesaria y suficiente para que el sistema de variables aleatorias X_1, X_2, \dots, X_h sea independiente es que dados borelianos arbitrarios B_1, B_2, \dots, B_h

$$P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_h) = P_{X_1}(B_1) P_{X_2}(B_2) \dots P_{X_h}(B_h). \quad (5.1)$$

Teorema 5.4 *Sea $\mathbf{X} = (X_1, X_2, \dots, X_h)$ un vector aleatorio con distribución discreta. Una condición necesaria y suficiente para que el conjunto de variables aleatorias X_1, X_2, \dots, X_h con distribución discreta sea independiente es que para todo $\mathbf{x} = (x_1, \dots, x_h) \in \mathbb{R}^h$*

$$p_{\mathbf{X}}(\mathbf{x}) = p_{X_1}(x_1) p_{X_2}(x_2) \dots p_{X_h}(x_h). \quad (5.2)$$

Demostración.

Es fácil ver que (5.2) es necesaria. Tomando en particular los borelianos $B_j = \{x_j\}$, $j = 1, 2, \dots, h$ y aplicando (5.1) se obtiene

$$\begin{aligned}
p_{\mathbf{X}}(\mathbf{x}) &= P_{\mathbf{X}}(\{(x_1, x_2, \dots, x_h)\}) = P_{\mathbf{X}}(\{x_1\} \times \{x_2\} \times \dots \times \{x_h\}) \\
&= P_{X_1}(\{x_1\}) P_{X_2}(\{x_2\}) \dots P_{X_h}(\{x_h\}) \\
&= p_{X_1}(x_1) p_{X_2}(x_2) \dots p_{X_h}(x_h).
\end{aligned}$$

Ahora veamos la suficiencia. Tenemos que probar que si ocurre (5.2) entonces las variables X_1, \dots, X_h son independientes. Como (5.1) implica la suficiencia, bastará probar que (5.2) implica (5.1).

Como la demostración para $k = 2$ es similar a la demostración general pero la notación es más simple, lo probaremos en este caso. Consideremos un

vector de dos componentes $\mathbf{X} = (X_1, X_2)$ y sean B_1, B_2 borelianos, entonces

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2) &= \sum_{x_1 \in B_1 \cap R_{X_1}} \sum_{x_2 \in B_2 \cap R_{X_2}} p_{\mathbf{X}}(x_1, x_2) \\ &= \sum_{x_1 \in B_1 \cap R_{X_1}} \sum_{x_2 \in B_2 \cap R_{X_2}} p_{X_1}(x_1) \cdot p_{X_1}(x_2) \\ &= \left(\sum_{x_1 \in B_1 \cap R_{X_1}} p_{X_1}(x_1) \right) \left(\sum_{x_2 \in B_2 \cap R_{X_2}} p_{X_1}(x_2) \right). \quad \square \end{aligned}$$

Observación. En la última igualdad hemos usado la fórmula

$$\sum_{(a,b) \in A \times B} ab = \sum_{a \in A} \sum_{b \in B} ab = \left(\sum_{a \in A} a \right) \cdot \left(\sum_{b \in B} b \right)$$

5.2. Ejemplos de vectores aleatorios con distribución discreta.

5.2.1. Distribución Multinomial.

Supongamos que un experimento que tiene k posibles resultados se repite n veces en forma independiente. Sean $A_i, i = 1, 2, \dots, k$, los posibles resultados del experimento y p_i la probabilidad que el resultado sea A_i . Luego

$$\sum_{i=1}^k p_i = 1.$$

Existen una gran cantidad de ejemplos de este tipo de experimentos. Por ejemplo si “se tira un dado” hay seis posibles resultados con la misma probabilidad. Luego $p_i = 1/6, i = 1, \dots, 6$. Otro experimento puede ser “se registra el voto de n ciudadanos elegidos al azar en una elección donde hay k candidatos”. En este caso en principio los valores de los p_i pueden ser arbitrarios.

Denotamos con X_i a la variable aleatoria “cantidad de veces que ocurre el resultado A_i a lo largo de los n experimentos” $i = 1, 2, \dots, k$ y formemos el vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_k)$. Se dice que el vector aleatorio $X = (X_1, \dots, X_k)$ tiene distribución multinomial con k resultados distintos con probabilidades p_1, \dots, p_k y n repeticiones y será simbolizada por $\mathcal{M}_k(p_1, \dots, p_k, n)$.

Como espacio muestral consideremos

$$\Omega = \{(i_1, i_2, \dots, i_n) : i_j \in \mathbb{N}, 1 \leq i_j \leq k\},$$

donde i_j indica el resultado que ocurrió en la j -ésima repetición del experimento.

Por ejemplo si $n = 4$ y $k = 3$ la 4-upla $(1, 3, 2, 3)$ indica que el resultado A_1 ocurrió la primera vez y nunca más, el resultado A_3 la segunda y cuarta vez y el resultado A_2 la tercera.

Con este espacio muestral, las variables aleatorias $X_j : \Omega \rightarrow \mathbb{N}$ están definidas por

$$X_i((i_1, i_2, \dots, i_n)) = \#\{j : i_j = i\}.$$

y se tiene que

$$\sum_{i=1}^k X_i((i_1, i_2, \dots, i_n)) = n.$$

El espacio Ω no es equiprobable. Vamos a encontrar ahora la probabilidad de cada elemento (i_1, \dots, i_n) de Ω . Consideremos los eventos

$$B_j = \{\text{en el experimento } j \text{ el resultado fue } i_j\}, \quad j = 1, \dots, n$$

Vamos ahora encontrar la probabilidad P definida sobre Ω . Luego el resultado (i_1, i_2, \dots, i_n) es equivalente a la intersección de B_j , $1 \leq j \leq n$. Como suponemos independencia de los experimentos y el evento B_j tiene probabilidad p_j , resulta

$$\begin{aligned} P(\{(i_1, i_2, \dots, i_n)\}) &= p_{i_1} p_{i_2} \dots p_{i_n} \\ &= p_1^{X_1((i_1, i_2, \dots, i_n))} p_2^{X_2((i_1, i_2, \dots, i_n))} \dots p_k^{X_k((i_1, i_2, \dots, i_n))}. \end{aligned} \quad (5.3)$$

El rango de \mathbf{X} es

$$R_{\mathbf{X}} = \left\{ (x_1, \dots, x_k) : 0 \leq x_i \leq n, \sum_{i=1}^k x_i = n \right\}$$

Fijado $\mathbf{x} = (x_1, \dots, x_k) \in R_{\mathbf{X}}$, calcularemos la probabilidad del evento

$$\begin{aligned} A &= \mathbf{X}^{-1}(\mathbf{x}) \\ &= \{(i_1, i_2, \dots, i_n) \in \Omega : \mathbf{X}((i_1, i_2, \dots, i_n)) = (x_1, x_2, \dots, x_k)\}. \end{aligned}$$

El evento A ocurre cuando para cada i , $0 \leq x_i \leq k$, el resultado A_i ocurre x_i veces en las n repeticiones del experimento. En particular si $(i_1, i_2, \dots, i_n) \in A$, de acuerdo a (5.3) se tendrá

$$P(\{(i_1, i_2, \dots, i_n)\}) = p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}.$$

Luego todo los elementos de A tienen la misma probabilidad y por lo tanto la probabilidad de A estará dada por la probabilidad de un elemento

por su cardinal . Un argumento simple de combinatoria muestra que

$$\begin{aligned} \#A &= \binom{n}{x_1} \binom{n-x_1}{x_2} \binom{n-x_1-x_2}{x_3} \cdots \binom{x_k}{x_k} \\ &= \frac{n!}{(x_1)!(n-x_1)!} \frac{(n-x_1)!}{(x_2)!(n-x_1-x_2)!} \frac{(n-x_1-x_2)!}{(x_3)!(n-x_1-x_2-x_3)!} \cdots 1 \\ &= \frac{n!}{(x_1)!(x_2)!(x_3)! \cdots (x_k)!} \end{aligned}$$

Esto resulta del hecho de que para elegir un elemento de A hay que elegir los x_1 lugares donde ocurrió A_1 entre los n , hay que elegir los x_2 lugares en los que ocurren A_2 entre los $n - x_1$ restantes, etc.

Luego tendremos

$$p_{\mathbf{X}}(x_1, x_2, \dots, x_k) = P_{\mathbf{X}}(A) = \frac{n!}{(x_1)!(x_2)!(x_3)! \cdots (x_k)!} \cdot p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}.$$

5.2.2. Distribución Hipergeométrica Multivariada.

Consideremos N objetos que pueden clasificarse en k clases distintas A_1, A_2, \dots, A_k .

Supongamos conocida la cantidad de objetos de cada clase, digamos D_1 de la clase A_1 , D_2 de la clase A_2 , \dots , D_k de la clase A_k , y por lo tanto $\sum_{i=1}^k D_i = N$. Supongamos que se realizan extracciones de n objetos y sea X_i la “cantidad de objetos de la clase i que se obtuvieron en las n extracciones”. Consideremos el vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_k)$.

Existen dos posibilidades

- (a) Las extracciones se hacen *con reposición*. En este caso, el experimento tiene distribución multinomial con parámetros p_1, p_2, \dots, p_k y n , donde $p_i = D_i/N$.
- (b) Las extracciones se hacen *sin reposición*. En este caso la distribución se denomina hipergeométrica multivariada y será denotada por $\text{HGM}_k(D_1, \dots, D_k, n)$.

El rango del vector \mathbf{X} estará dado por

$$R_{\mathbf{X}} = \{(x_1, x_2, \dots, x_k) : 0 \leq x_i \leq D_i, x_1 + x_2 + \cdots + x_k = n\}.$$

Como cada n -upla tiene una probabilidad distinta, no será conveniente tomar como espacio muestral el conjunto de estas k -uplas. Para construir un espacio de probabilidad equiprobable procedemos de la siguiente manera. Comenzamos enumerando todos los objetos de la siguiente manera. Los de clase 1 por

$$M_1 = \{1, 2, \dots, D_1\}.$$

Los de la clase 2 por

$$M_2 = \{D_1 + 1, D_1 + 2, \dots, D_1 + D_2\}.$$

Los de la clase 3 por

$$M_3 = \{D_1 + D_2 + 1, D_1 + D_2 + 2, \dots, D_1 + D_2 + D_3\}.$$

y finalmente los de la clase k por

$$M_k = \left\{ \sum_{i=1}^{k-1} D_i + 1, \sum_{i=1}^{k-1} D_i + 2, \dots, \sum_{i=1}^k D_i \right\}.$$

Definamos entonces el espacio muestral por

$$\Omega = \{A : A \subset \{1, \dots, N\}, \#A = n\},$$

Si el conjunto A se interpreta como el conjunto de los números de las bolillas obtenidas, resultará que todos los elementos de Ω son equiprobables. Por ejemplo si $N = 20$ y $n = 3$ la probabilidad de extraer los elementos $\{1, 2, 17\}$ o $\{2, 6, 8\}$ es la misma.

El número de elementos de Ω es la cantidad de subconjuntos de n elementos que se pueden formar con los N dados. Luego

$$\#(\Omega) = \binom{N}{n}$$

Dado $A \in \Omega$, se define $X_i(A) = \#(A \cap M_i)$, $1 \leq i \leq k$, y $\mathbf{X}(A) = (X_1(A), \dots, X_k(A))$. Consideremos ahora el evento

$$C = \{A : \mathbf{X}(A) = (x_1, x_2, \dots, x_k)\}.$$

El evento C representa todas las extracciones en las que resulta que hay exactamente x_1 elementos de la clase A_1 , x_2 de la clase A_2 , ..., x_k de la clase A . Un argumento combinatorio simple muestra que el cardinal de C es

$$\#(C) = \binom{D_1}{x_1} \binom{D_2}{x_2} \dots \binom{D_k}{x_k},$$

de manera que

$$p_{\mathbf{X}}(x_1, x_2, \dots, x_k) = P(C) = \frac{\binom{D_1}{x_1} \binom{D_2}{x_2} \dots \binom{D_k}{x_k}}{\binom{N}{n}}.$$

5.3. Vectores Aleatorios de tipo absolutamente continuo.

Definición 5.3 Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y $\mathbf{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio. Se dice que el vector es absolutamente continuo si existe una función integrable sobre \mathbb{R}^k , $f_{\mathbf{X}} : \mathbb{R}^k \rightarrow \mathbb{R}_{\geq 0}$ llamada función de densidad de la probabilidad $P_{\mathbf{X}}$ tal que

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2, \dots, x_k) &= \int_{-\infty}^{x_k} \int_{-\infty}^{x_{k-1}} \cdots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_1 dt_2 \dots dt_k \\ &= \int \cdots \int_{(-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_k]} f_{\mathbf{X}}(\mathbf{t}) d\mathbf{t}, \end{aligned}$$

donde $\mathbf{t} = (t_1, t_2, \dots, t_k)$ y $d\mathbf{t} = dt_1 dt_2 \dots dt_k$.

Tomando límite cuando $x_1 \rightarrow \infty, \dots, x_k \rightarrow \infty$, se tendrá

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(\mathbf{t}) d\mathbf{t} = P_{\mathbf{X}}(\mathbb{R}^k) = 1.$$

El siguiente teorema da la probabilidad que un vector aleatorio tome valores en un rectángulo k -dimensional.

Teorema 5.5 Supongamos que $\mathbf{X} = (X_1, \dots, X_k)$ sea un vector aleatorio absolutamente continuo con densidad $f_{\mathbf{X}}$. Sean $a_1 < b_1$, $a_2 < b_2$, $a_3 < b_3, \dots, a_k < b_k$. Luego se tiene

$$\begin{aligned} &P_{\mathbf{X}}((a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_k, b_k]) \\ &= \int_{a_k}^{b_k} \int_{a_{k-1}}^{b_{k-1}} \cdots \int_{a_1}^{b_1} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_1 dt_2 \dots dt_k. \\ &= \int \cdots \int_{(a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_k, b_k]} f_{\mathbf{X}}(\mathbf{t}) d\mathbf{t}, \end{aligned}$$

Demostración. Tenemos que mostrar que

$$\begin{aligned} &\Delta_k(a_k, b_k) \cdots \Delta_1(a_1, b_1) F_{\mathbf{X}}(x_1, x_2, \dots, x_k) \\ &= \int_{a_k}^{b_k} \int_{a_{k-1}}^{b_{k-1}} \cdots \int_{a_1}^{b_1} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_1 dt_2 \dots dt_k. \end{aligned}$$

Para esto bastará probar que para todo $1 \leq h \leq k$ se tiene

$$\begin{aligned} &\Delta_h(a_h, b_h) \cdots \Delta_1(a_1, b_1) F_{\mathbf{X}}(x_1, x_2, \dots, x_h, x_{h+1}, \dots, x_k) \\ &= \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_{h+1}} \int_{a_h}^{b_h} \cdots \int_{a_1}^{b_1} f_{\mathbf{X}}(t_1, t_2, \dots, t_h, t_{h+1}, \dots, t_k) dt_1 dt_2 \dots dt_h, \end{aligned}$$

y esto se prueba por inducción en h . \square

Observación. Usando la integral de Lebesgue, se puede probar, mediante teoría de la medida e integración que para todo boreliano $B \in \mathcal{B}^k$

$$P_{\mathbf{X}}(B) = \int \cdots \int_B f_{\mathbf{X}}(\mathbf{t}) dt. \quad (5.4)$$

Si se usa la integral de Riemman, la integral del segundo miembro de (5.4) puede no existir. Únicamente existe si el borde de B tiene medida de Riemman 0. En cambio la correspondiente integral de Lebesgue siempre existe. Desde el punto de vista práctico en este curso solo se va a trabajar con conjuntos B para los cuales la integral de Riemman existe.

La función de densidad de probabilidad tiene una interpretación análoga a la que hemos visto para el caso univariado. La siguiente propiedad dice que en un punto de continuidad, el límite de la probabilidad de un entorno de un punto sobre su volumen, cuando el entorno se aproxima al punto es el valor de la densidad en el punto. Más precisamente

Teorema 5.6 *Sea $f_{\mathbf{X}}$ la función densidad asociada al vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_k)$ continua en el punto $\mathbf{x}_0 = (x_{10}, x_{20}, \dots, x_{k0})$. Entonces*

$$\lim_{h \rightarrow 0} \frac{P_{\mathbf{X}}([x_{10} - h, x_{10} + h] \times \cdots \times [x_{k0} - h, x_{k0} + h])}{(2h)^k} = f_{\mathbf{X}}(\mathbf{x}_0).$$

Demostración. Es análoga al caso univariado y se deja como ejercicio. \square

Observación. Los entornos cúbicos se pueden reemplazar por otro tipo de entornos, por ejemplo entornos esféricos. En el denominador habrá que poner el volumen correspondiente.

Bajo el supuesto de que la densidad sea continua, se puede escribir la densidad como la derivada parcial cruzada de orden k de la función de distribución.

Teorema 5.7 *Supongamos que $f_{\mathbf{X}}$ sea continua en \mathbf{x}_0 . Entonces*

$$f_{\mathbf{X}}(\mathbf{x}_0) = \left. \frac{\partial^k F_{\mathbf{X}}(x_1, x_2, \dots, x_k)}{\partial x_k \partial x_{k-1} \cdots \partial x_1} \right|_{\mathbf{x}=\mathbf{x}_0}.$$

Demostración. Por Fubini se tiene

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2, \dots, x_k) &= \int_{-\infty}^{x_k} \int_{-\infty}^{x_{k-1}} \cdots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_1 dt_2 \cdots dt_k \\ &= \int_{-\infty}^{x_1} \left[\int_{-\infty}^{x_k} \int_{-\infty}^{x_{k-1}} \cdots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_2 \cdots dt_k \right] dt_1 \end{aligned}$$

y aplicando el teorema fundamental del cálculo resulta

$$\begin{aligned}\frac{\partial F_{\mathbf{X}}(x_1, x_2, \dots, x_k)}{\partial x_1} &= \int_{-\infty}^{x_k} \int_{-\infty}^{x_{k-1}} \cdots \int_{-\infty}^{x_2} f_{\mathbf{X}}(x_1, t_2, \dots, t_k) dt_2 \dots dt_k \\ &= \int_{-\infty}^{x_2} \left[\int_{-\infty}^{x_k} \int_{-\infty}^{x_{k-1}} \cdots \int_{-\infty}^{x_3} f_{\mathbf{X}}(x_1, t_2, \dots, t_k) dt_3 \dots dt_k \right] dt_2\end{aligned}$$

y aplicando nuevamente el teorema fundamental del cálculo obtenemos

$$\frac{\partial F_{\mathbf{X}}(x_1, x_2, \dots, x_k)}{\partial x_2 \partial x_1} = \int_{-\infty}^{x_k} \int_{-\infty}^{x_{k-1}} \cdots \int_{-\infty}^{x_2} f_{\mathbf{X}}(x_1, x_2, t_3, \dots, t_k) dt_3 \dots dt_k.$$

Repitiendo lo mismo k veces se demuestra el teorema. \square

Definición 5.4 Dado un boreliano $B \in \mathcal{B}^k$ se define su volumen de la siguiente manera

$$\text{Vol}(B) = \int \cdots \int_B dx_1 dx_2 \dots dx_k = \int \cdots \int_B d\mathbf{x}.$$

Observación. Un caso típico de conjuntos con volumen 0 resulta ser un punto en \mathbb{R} , una recta en \mathbb{R}^2 , un plano en \mathbb{R}^3 y en general un hiperplano en \mathbb{R}^k . Las uniones a lo sumo numerables de conjuntos de volumen cero tienen volumen cero. En general cualquier subconjunto de \mathbb{R}^k de dimensión j con $j < k$ tendrá volumen 0. Por ejemplo las curvas en \mathbb{R}^2 o las superficies en \mathbb{R}^3 .

Veremos que si el vector aleatorio es absolutamente continuo la función de probabilidad asociada asigna probabilidad 0 a conjuntos cuyo volumen es 0.

Teorema 5.8 Sea \mathbf{X} un vector aleatorio de dimensión k . Si $B \in \mathcal{B}^k$ tal que $\text{Vol}(B) = 0$ entonces $P_{\mathbf{X}}(B) = 0$.

Demostración. Sea

$$C_n = \{\mathbf{x} \in \mathbb{R}^k : f_{\mathbf{X}}(\mathbf{x}) > n\}.$$

Es claro que si $\mathbf{x} \in C_{n+1}$ entonces $f_{\mathbf{X}}(\mathbf{x}) > n+1 > n$ de manera que $\mathbf{x} \in C_n$, es decir la sucesión de conjuntos $\{C_n\}_{n \geq 1}$ es decreciente y además, puesto que la función $f_{\mathbf{X}}$ es finita en todo punto, se tiene $\bigcap_{n=1}^{\infty} C_n = \emptyset$. Luego también se tendrá

$$\lim_{n \rightarrow \infty} P_{\mathbf{X}}(C_n) = 0.$$

Podemos descomponer a $B = (B \cap C_n) \cup (B \cap C_n^c)$. Como esta unión es disjunta, se tiene

$$P_{\mathbf{X}}(B) = P_{\mathbf{X}}(B \cap C_n) + P_{\mathbf{X}}(B \cap C_n^c).$$

Ahora calculamos $P_{\mathbf{X}}(B \cap C_n^c)$. Para ello observemos que para todo $n \in \mathbb{N}$

$$\begin{aligned} P(B \cap C_n^c) &= \int \cdots \int_{B \cap C_n^c} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} \\ &\leq n \int \cdots \int_{B \cap C_n^c} d\mathbf{x} \\ &= n \text{Vol}(B \cap C_n^c) \\ &\leq n \text{Vol}(B) \\ &= 0. \end{aligned}$$

Entonces para todo $n \in \mathbb{N}$ resulta

$$P_{\mathbf{X}}(B) = P_{\mathbf{X}}(B \cap C_n) \leq P_{\mathbf{X}}(C_n),$$

de manera que pasando al límite se concluye que $P_{\mathbf{X}}(B) = 0$. \square

Observación. Existe una diferencia importante entre los vectores discretos y los absolutamente continuos. Recordemos que un vector es discreto si y sólo si sus componentes son variables discretas. Esto no ocurre en el caso de los vectores aleatorios absolutamente continuos. Para demostrarlo daremos un contraejemplo.

Consideremos una variable aleatoria X_1 , con distribución absolutamente continua y sea $X_2 = X_1$ de manera que el vector $\mathbf{X} = (X_1, X_2)$ tiene como componentes variables aleatorias con distribuciones absolutamente continuas. Ahora veamos que el vector \mathbf{X} no puede tener distribución absolutamente continua.

Para ello observemos que

$$B = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = x_2\}$$

es una recta en \mathbb{R}^2 de manera que tiene volumen cero. Pero sin embargo

$$P_{\mathbf{X}}(B) = P(\{\omega \in \Omega : X_1(\omega) = X_2(\omega)\}) = P(\Omega) = 1.$$

Teorema 5.9 Sea $\mathbf{X} = (X_1, X_2, \dots, X_h, X_{h+1}, \dots, X_k)$ un vector aleatorio de dimensión k . Consideremos un subconjunto de coordenadas y formemos el vector aleatorio asociado $\mathbf{X}^* = (X_1, X_2, \dots, X_h)$. Entonces \mathbf{X}^* también es absolutamente continuo y

$$\begin{aligned} f_{\mathbf{X}^*}(x_1, x_2, \dots, x_h) & \tag{5.5} \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x_1, x_2, \dots, x_h, t_{h+1}, \dots, t_k) \, dt_{h+1} dt_{h+2} \cdots dt_k. \end{aligned}$$

Demostración. Tenemos que

$$\begin{aligned}
F_{\mathbf{X}^*}(x_1, x_2, \dots, x_h) &= P_{\mathbf{X}^*}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_h]) \\
&= P_{\mathbf{X}}\left((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_h] \times \underbrace{\mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}}_{k-h \text{ factores}}\right) \\
&= \int \dots \int_{(-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_h] \times \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_1 dt_2 \dots dt_k \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \int_{-\infty}^{x_h} \dots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_1 \dots dt_h dt_{h+1} dt_{h+2} \dots dt_k
\end{aligned}$$

Por lo tanto, usando Fubini, se tendrá

$$\begin{aligned}
F_{\mathbf{X}^*}(x_1, x_2, \dots, x_h) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \int_{-\infty}^{x_h} \dots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_1 \dots dt_h dt_{h+1} dt_{h+2} \dots dt_k \\
&= \int_{-\infty}^{x_h} \dots \int_{-\infty}^{x_1} \left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_{h+1} dt_{h+2} \dots dt_k \right] dt_1 \dots dt_h
\end{aligned}$$

Luego tenemos que

$$F_{\mathbf{X}^*}(x_1, x_2, \dots, x_h) = \int_{-\infty}^{x_h} \dots \int_{-\infty}^{x_1} f_{\mathbf{X}^*}(t_1, t_2, \dots, t_h) dt_1 \dots dt_h,$$

donde $f_{\mathbf{X}^*}$ está dada por (5.5). Esto prueba el Teorema. \square

Observación. Por comodidad hemos escogido las primeras h componentes pero lo mismo puede hacerse para una colección arbitraria de ellas. En el caso de una distribución bivariada $\mathbf{X} = (X_1, X_2)$, $\mathbf{X}^* = X_1$

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x_1, x_2) dx_2.$$

El siguiente Teorema da una condición necesaria y suficiente para que un conjunto de variables absolutamente continuas sean independientes.

Teorema 5.10 Sean X_1, \dots, X_k variables aleatorias absolutamente continuas con densidades f_{X_1}, \dots, f_{X_k} . Luego estas variables son independientes si y sólo si el vector $\mathbf{X} = (X_1, \dots, X_k)$ tiene como densidad conjunta a la función

$$f(x_1, \dots, x_k) = \prod_{i=1}^k f_{X_i}(x_i).$$

Demostración. Como sabemos, por el Teorema 4.13, que X_1, \dots, X_k son independientes si y sólo si

$$F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^k F_{X_i}(x_i), \quad (5.6)$$

por el Teorema 4.5 (Teorema de Extensión para vectores aleatorios) bastará probar que la función de distribución F correspondiente a f está dada por (5.6). Vamos a mostrar que esto es cierto. En efecto, tenemos

$$\begin{aligned} F(x_1, \dots, x_k) &= \int_{-\infty}^{x_k} \dots \int_{-\infty}^{x_1} \prod_{i=1}^k f_{X_i}(x_i) dx_1 \dots dx_k \\ &= \prod_{i=1}^k \int_{-\infty}^{x_i} f_{X_i}(x_i) dx_i \\ &= \prod_{i=1}^k F_{X_i}(x_i), \end{aligned}$$

y luego el Teorema queda probado. \square

El siguiente Teorema que se deja como ejercicio prueba una propiedad similar para vectores.

Teorema 5.11 Sean $\mathbf{X}_1, \dots, \mathbf{X}_k$ vectores aleatorios absolutamente continuos con densidades $f_{\mathbf{X}_1}, \dots, f_{\mathbf{X}_k}$. Luego estos vectores son independientes si y sólo si el vector $\mathbf{X}^* = (\mathbf{X}_1, \dots, \mathbf{X}_k)$ tiene como densidad a la función

$$f(\mathbf{x}_1, \dots, \mathbf{x}_k) = \prod_{i=1}^k f_{\mathbf{X}_i}(\mathbf{x}_i).$$

Capítulo 6

Transformaciones de variables y vectores aleatorios.

En esta sección estudiaremos cómo se obtienen las distribuciones de variables o vectores aleatorios obtenidos a partir de otros a través de cierto tipo de transformaciones.

6.1. Transformaciones monótonas de variables aleatorias.

Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y X una variable aleatoria.

Consideremos una función $g : \mathbb{R} \rightarrow \mathbb{R}$ continua y estrictamente monótona, es decir, estrictamente creciente o bien estrictamente decreciente. Sabemos que $Y = g(X)$ es otra variable aleatoria. Queremos estudiar la relación que existe entre F_X y F_Y .

Caso de g estrictamente creciente.

La imagen de $g(\mathbb{R})$ es un intervalo abierto (a, b) de longitud finita o bien infinita, es decir también puede ser $-\infty$ y $b = \infty$. El siguiente teorema da la relación entre F_X y F_Y .

Teorema 6.1 *Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función estrictamente creciente y sea $(a, b) = g(\mathbb{R})$. Entonces si X es una variable aleatoria con función de distribución F_X , la función de distribución de $Y = g(X)$ será*

$$F_Y(y) = \begin{cases} 0 & \text{si } y \leq a \\ F_X(g^{-1}(y)) & \text{si } y \in (a, b) \\ 1 & \text{si } y \geq b. \end{cases} \quad (6.1)$$

Demostración. Sea $a < y < b$. Como g es estrictamente creciente se tendrá

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

Si $y \leq a$ se tendrá que $\{\omega : g(X(\omega)) \leq y\} = \emptyset$ y luego

$$F_Y(y) = P(\{\omega : g(X(\omega)) \leq y\}) = 0.$$

Del mismo modo, si $y \geq b$ se tendrá $\{\omega : g(X(\omega)) \leq y\} = \Omega$, y luego

$$F_Y(y) = P(\{\omega : g(X(\omega)) \leq y\}) = 1. \quad \square$$

Caso de g estrictamente decreciente.

Nuevamente la imagen de g es un abierto (a, b) de longitud finita o infinita. En este caso tenemos el siguiente teorema.

Teorema 6.2 *Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función estrictamente decreciente $(a, b) = g(\mathbb{R})$. Entonces se tiene*

(a) *Si X es una variable aleatoria con función de distribución F_X , la función de distribución de $Y = g(X)$ será*

$$F_Y(y) = \begin{cases} 0 & \text{si } y \leq a \\ 1 - P(X < g^{-1}(y)) & \text{si } y \in (a, b) \\ 1 & \text{si } y \geq b. \end{cases} \quad (6.2)$$

(b) *Si además F_X es continua se tendrá*

$$F_Y(y) = \begin{cases} 0 & \text{si } y \leq a \\ 1 - F_X(g^{-1}(y)) & \text{si } y \in (a, b) \\ 1 & \text{si } y \geq b. \end{cases} \quad (6.3)$$

Demostración.

(a) Como g es estrictamente decreciente se tiene para $a < y < b$ que

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) \\ &= P(X \geq g^{-1}(y)) = 1 - P(X < g^{-1}(y)). \end{aligned}$$

Los casos $y \leq a$ y $y \geq b$ se demuestran como en el Teorema 6.1.

(b) En este caso se tiene

$$P(X < g^{-1}(y)) = P(X \leq g^{-1}(y)) = 1 - F_X(g^{-1}(y)). \quad \square$$

Ahora caracterizaremos la función de densidad asociada a Y . Supongamos que X tiene distribución absolutamente continua con densidad f_X y además que g es derivable.

Teorema 6.3 *Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función estrictamente creciente o decreciente y derivable con $g'(y) \neq 0$. Sea $(a, b) = g(\mathbb{R})$, entonces si X es una variable aleatoria absolutamente continua con función de densidad f_X , la función de densidad de $Y = g(X)$ será*

$$f_Y(y) = \begin{cases} 0 & \text{si } y \leq a \\ \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|} & \text{si } y \in (a, b) \\ 0 & \text{si } y \geq b. \end{cases} \quad (6.4)$$

Demostración. En el caso de que g es estrictamente creciente, (6.4) se obtiene derivando (6.1) y observando que $g' > 0$. En el caso que g sea estrictamente decreciente, derivando (6.3) y observando que $g' < 0$. \square

Un caso especial de interés ocurre cuando g es una transformación afín, es decir cuando $g(x) = cx + d$ con $c \neq 0$. En este caso $Y = g(X) = cX + d$ y $g'(x) = c$. Como $a = -\infty$ y $b = +\infty$, teniendo en cuenta que $g^{-1}(y) = \frac{y-d}{c}$ obtenemos

$$f_Y(y) = \frac{1}{|c|} f_X\left(\frac{y-d}{c}\right). \quad (6.5)$$

6.1.1. Distribución Normal

Hemos visto la distribución de una variable normal standarizada $X \sim N(0, 1)$ cuya función densidad es

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2).$$

Ahora vamos a definir para todo $\mu \in \mathbb{R}$ y para todo $\sigma \in \mathbb{R}_{>0}$ la distribución normal con media μ y varianza σ^2 que indicaremos con $N(\mu, \sigma^2)$. Esta distribución es la que corresponde a $Y = \sigma X + \mu$, donde X es $N(0, 1)$.

De acuerdo a (6.5) tendremos

$$\begin{aligned} f_Y(y) &= \frac{1}{\sigma} f_X\left(\frac{y-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y-\mu}{\sigma}\right)^2\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right). \end{aligned}$$

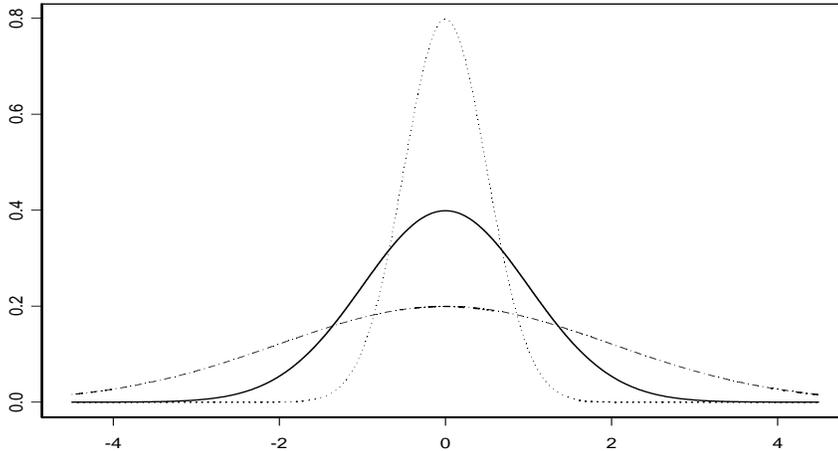


Figura 6.1: Densidad de la normal estándar (en línea sólida), de la $N(0, 4)$ (en línea de puntos) y de la $N(0, \frac{1}{4})$ (en línea de puntos y rayas).

El significado de los parámetros μ y σ se estudiará en la sección 7.7.1. Adelantemos que μ representa un desplazamiento horizontal de la densidad e indica el centro de simetría de la misma. La densidad alcanza su máximo en μ y a medida que nos alejamos de μ , la densidad va decreciendo. El parámetro σ , indica la dispersión de la variable respecto del centro. Un factor σ grande “achata” la curva hacia el eje de abscisas, y en este caso la dispersión es grande. Cuando σ es chico, la probabilidad está más concentrada cerca de μ .

En la Figura 6.1 se muestran densidades normales con diferentes valores de σ ilustrando el significado de este parámetro.

Ejercicio. Se deja como ejercicio mostrar que si Y tiene distribución $N(\mu, \sigma^2)$, entonces $Z = (Y - \mu)/\sigma$ tiene distribución $N(0, 1)$. Esta transformación se llama *estandarización de la variable Y* y permite calcular las probabilidades de cualquier distribución $N(\mu, \sigma^2)$ usando la distribución $N(0, 1)$. Por ejemplo, sea Y con distribución $N(3, 4)$ y supongamos que queremos encontrar $P(3 < Y < 5)$. Luego $Z = (Y - 3)/2$ es $N(0, 1)$ y tendremos

$$\begin{aligned} P(3 < Y < 5) &= P\left(\frac{3-3}{2} < \frac{Y-3}{2} < \frac{5-3}{2}\right) \\ &= P(0 < Z < 1) \\ &= \Phi(1) - \Phi(0) \end{aligned}$$

donde Φ es la función de distribución de una $N(0, 1)$. Usando una tabla de

la $N(0, 1)$ encontramos que $\Phi(0) = 0,50$ y $\Phi(1) = 0,8413$ Luego

$$P(3 < Y < 5) = 0,8413 - 0,50 = 0,3413.$$

6.2. Transformaciones inyectivas de vectores aleatorios.

Recordemos algunos resultados de cálculo integral en varias variables.

Sea $U \subset \mathbb{R}^k$ un abierto y $g : U \rightarrow \mathbb{R}^k$ una función inyectiva de manera que $g : U \rightarrow V = g(U)$ resulta biyectiva. Podemos representar $g = (g_1, \dots, g_k)$, donde $g_i : U \rightarrow \mathbb{R}$. Luego existe $g^{-1} : V \rightarrow U$. Supongamos que g es diferenciable en cada punto $\mathbf{x} \in U$. El jacobiano de g se define por

$$J_g(\mathbf{x}) = \det \begin{pmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \frac{\partial g_1(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial g_1(\mathbf{x})}{\partial x_k} \\ \frac{\partial g_2(\mathbf{x})}{\partial x_1} & \frac{\partial g_2(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial g_2(\mathbf{x})}{\partial x_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_k(\mathbf{x})}{\partial x_1} & \frac{\partial g_k(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial g_k(\mathbf{x})}{\partial x_k} \end{pmatrix} \neq 0.$$

Entonces si $\mathbf{y} \in V$ y $J_g(g^{-1}(\mathbf{y})) \neq 0$, resulta que g^{-1} es diferenciable en \mathbf{y} y se tiene

$$J_{g^{-1}}(\mathbf{y}) = \frac{1}{J_g(g^{-1}(\mathbf{y}))}.$$

El siguiente teorema permite realizar un cambio de variables para integrales múltiples.

Teorema 6.4 *Sea $A \subset U \subset \mathbb{R}^k$ un conjunto tal que el borde tiene medida de Riemann 0, $f : U \rightarrow \mathbb{R}$ una función continua, $g : \mathbb{R}^k \rightarrow \mathbb{R}^k$ una función inyectiva y diferenciable tal que $J_g(\mathbf{x}) \neq 0$ para todo $\mathbf{x} \in A$. Entonces*

$$\int \dots \int_A f(\mathbf{x}) d\mathbf{x} = \int \dots \int_{g(A)} f(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| d\mathbf{y}.$$

donde $d\mathbf{x} = dx_1 dx_2 \dots dx_k$ y $d\mathbf{y} = dy_1 dy_2 \dots dy_k$.

Sea ahora $\mathbf{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio con distribución absolutamente continua y sea $f_{\mathbf{X}}$ su densidad. El siguiente teorema permitirá encontrar la distribución del vector $\mathbf{Y} = g(\mathbf{X})$.

Teorema 6.5 *Sea $\mathbf{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio absolutamente continuo con densidad $f_{\mathbf{X}}$ tal que $P_{\mathbf{X}}(U) = 1$, donde U es un abierto en \mathbb{R}^k . Sea $g : U \rightarrow \mathbb{R}^k$ una función inyectiva diferenciable tal que para todo $\mathbf{x} \in U$*

se tiene $J_g(\mathbf{x}) \neq 0$. Luego el vector $\mathbf{Y} = g(\mathbf{X})$ también es absolutamente continuo y su densidad está dada por

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| I_V(\mathbf{y}),$$

donde $V = g(U)$, e I_V es la función indicadora del conjunto V .

Demostración. Para esto bastará demostrar que para todo $B \in \mathcal{B}^k$

$$P_{\mathbf{Y}}(B) = \int \cdots \int_B f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| I_V(\mathbf{y}) d\mathbf{y}. \quad (6.6)$$

Por definición de función de densidad de \mathbf{X} se tiene que

$$\begin{aligned} P_{\mathbf{Y}}(B) &= P(\mathbf{Y} \in B \cap V) \\ &= P(g(\mathbf{X}) \in B \cap V) \\ &= P(\mathbf{X} \in g^{-1}(B \cap V)) \\ &= \int \cdots \int_{g^{-1}(B \cap V)} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Usando la fórmula de cambio de variables en integrales múltiples resulta

$$\begin{aligned} P_{\mathbf{Y}}(B) &= \int \cdots \int_{g^{-1}(B \cap V)} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \int \cdots \int_{g(g^{-1}(B \cap V))} f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| d\mathbf{y}. \end{aligned}$$

Sea $g : U \rightarrow W$ y $H \subset W$. Es fácil ver que una condición necesaria y suficiente para que $g(g^{-1}(H)) = H$ es que $H \subset g(U)$. Como $B \cap V \subset V = g(U)$ resulta $g(g^{-1}(B \cap V)) = B \cap V$ y por lo tanto

$$\begin{aligned} P_{\mathbf{Y}}(B) &= \int \cdots \int_{g(g^{-1}(B \cap V))} f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| d\mathbf{y} \\ &= \int \cdots \int_{B \cap V} f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| d\mathbf{y} \\ &= \int \cdots \int_B f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| I_V(\mathbf{y}) d\mathbf{y}. \end{aligned}$$

Esto muestra que vale (6.6). \square

El resultado anterior vale cuando g es diferenciable y biunívoca de un abierto de \mathbb{R}^k en \mathbb{R}^k . Veamos ahora que ocurre cuando g es una función

diferenciable de un abierto de \mathbb{R}^k en \mathbb{R}^j con $j \neq k$. Si $j > k$ nada podemos hacer puesto que en tal caso el conjunto $g(U)$ es un conjunto de dimensión k y por lo tanto tiene volumen 0. Luego como $P_{\mathbf{Y}}(g(U)) = 1$, \mathbf{Y} no puede ser un vector absolutamente continuo.

Consideremos ahora $j < k$ y sea U un abierto en \mathbb{R}^k . Supongamos que $g = (g_1, \dots, g_j) : \mathbb{R}^k \rightarrow \mathbb{R}^j$, donde cada $g_i : \mathbb{R}^k \rightarrow \mathbb{R}$, $1 \leq i \leq j$, es una función diferenciable. Trataremos de derivar la densidad $f_{\mathbf{Y}}$ de $\mathbf{Y} = g(\mathbf{X})$. Esto es posible si se pueden encontrar funciones diferenciables $g_i : \mathbb{R}^k \rightarrow \mathbb{R}$, $i = j+1, \dots, k$ tales que si llamamos $\tilde{g} = (g_1, \dots, g_j, g_{j+1}, \dots, g_k)$ la función $\tilde{g} : \mathbb{R}^k \rightarrow \mathbb{R}^k$ resulte inyectiva y $J_{\tilde{g}}(\mathbf{y}) \neq 0$ para todo $\mathbf{y} \in U$. En, efecto en este caso por el teorema anterior podremos encontrar la densidad de $\tilde{\mathbf{Y}} = \tilde{g}(\mathbf{X})$ que denominaremos $f_{\tilde{\mathbf{Y}}}$. Luego la densidad de \mathbf{Y} será

$$f_{\mathbf{Y}}(y_1, \dots, y_j) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\tilde{\mathbf{Y}}}(y_1, \dots, y_j, y_{j+1}, \dots, y_k) dy_{j+1} \dots dy_k.$$

Veamos un ejemplo del uso de este procedimiento. Sea $\mathbf{X} = (X_1, X_2)$ y consideremos $Y = X_1 + X_2$. Si definimos $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ por $g(x_1, x_2) = x_1 + x_2$, vemos que $Y = g(\mathbf{X})$. En este caso $1 = j < k = 2$. Ahora consideremos $\tilde{g} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, definida por $\tilde{g}(x_1, x_2) = (x_1 + x_2, x_2)$ e $\mathbf{Y} = (Y_1, Y_2)$ con $Y_1 = g(\mathbf{X})$ e $Y_2 = X_2$. Luego estamos en las condiciones del teorema puesto que $\tilde{g} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ es biyectiva, diferenciable y su Jacobiano es

$$J_{\tilde{g}}(x_1, x_2) = \det \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = 1.$$

Luego tenemos $\tilde{g}^{-1}(y_1, y_2) = (y_1 - y_2, y_2)$.

En este caso $U = V = \mathbb{R}^2$, y entonces acuerdo al Teorema 6.5, se tendrá

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(\tilde{g}^{-1}(\mathbf{y})) |J_{\tilde{g}^{-1}}(\mathbf{y})| \\ &= f_{\mathbf{X}}(y_1 - y_2, y_2) \end{aligned}$$

y

$$f_Y(y) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(y - y_2, y_2) dy_2.$$

En el caso que X_1 y X_2 son independientes con densidades f_{X_1} y f_{X_2} , se tendrá

$$f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2),$$

y entonces f_Y está dado por

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y - y_2)f_{X_2}(y_2) dy_2. \quad (6.7)$$

La función f_Y dada por (6.7) se denomina convolución de $f_{X_1}(x_1)$ y $f_{X_2}(x_2)$.

6.3. Algunas aplicaciones a la distribución normal.

Sea $\mathbf{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio tal que sus componentes son variables aleatorias independientes con idéntica distribución $N(0, 1)$. Sea $A \in \mathbb{R}^{k \times k}$ una matriz ortogonal, es decir tal que $A^{-1} = A'$ donde A' denota la traspuesta de la matriz A . Definimos la función $g : \mathbb{R}^k \rightarrow \mathbb{R}^k$ dada por $g(\mathbf{x}) = \mathbf{x}A$ y consideramos el vector aleatorio $\mathbf{Y} = \mathbf{X}A$. El siguiente teorema muestra que la distribución de \mathbf{Y} es la misma que la del vector \mathbf{X} .

Teorema 6.6 *La distribución de vector \mathbf{Y} es la misma que la del vector \mathbf{X} .*

Demostración. La función de densidad del vector \mathbf{X} es

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{\sqrt{(2\pi)^k}} \prod_{i=1}^k \exp\left(-\frac{1}{2}x_i^2\right) \\ &= \frac{1}{\sqrt{(2\pi)^k}} \exp\left(-\sum_{i=1}^k x_i^2\right) \\ &= \frac{1}{\sqrt{(2\pi)^k}} \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right). \end{aligned}$$

Sea $g : \mathbb{R}^k \rightarrow \mathbb{R}^k$ definida por $g(\mathbf{x}) = \mathbf{x}A$, luego $g^{-1}(\mathbf{y}) = \mathbf{y}A^{-1} = \mathbf{y}A'$. Calculando el Jacobiano de g vemos que $J_g(\mathbf{x}) = \det A = \pm 1$, de manera que por el Teorema 6.5 y el hecho de que por ser A' ortogonal $\|g^{-1}(\mathbf{y})\| = \|\mathbf{y}A'\| = \|\mathbf{y}\|$, la densidad de \mathbf{Y} está dada por

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| I_{\mathbb{R}^k}(\mathbf{y}) \\ &= f_{\mathbf{X}}(g^{-1}(\mathbf{y})) \\ &= \exp\left(-\frac{1}{2}\|g^{-1}(\mathbf{y})\|^2\right) \\ &= \frac{1}{\sqrt{(2\pi)^k}} \exp\left(-\frac{1}{2}\|\mathbf{y}\|^2\right). \end{aligned}$$

Esto prueba el teorema. \square

El siguiente teorema prueba que combinaciones lineales de variables aleatorias normales independientes son normales.

Teorema 6.7 *(i) Sean X_1, X_2, \dots, X_k variables aleatorias independientes con distribución $N(0, 1)$. Sean b_1, \dots, b_k números reales, tales que $\sum_{i=1}^k b_i^2 = 1$, es decir el vector $\mathbf{b} = (b_1, \dots, b_k)' \in \mathbb{R}^k$ tiene norma unitaria. Luego la variable $Z = b_1X_1 + \dots + b_kX_k$ también distribución $N(0, 1)$.*

(ii) Sean Y_1, Y_2, \dots, Y_k variables aleatorias independientes tales que Y_i tiene distribución $N(\mu_i, \sigma_i^2)$, luego dados números reales $\alpha_1, \dots, \alpha_k$ y β , la distribución de $Z = \sum_{i=1}^k \alpha_i Y_i + \beta$ es

$$N\left(\sum_{i=1}^k \alpha_i \mu_i + \beta, \sum_{i=1}^k \alpha_i^2 \sigma_i^2\right).$$

Demostración.

(i) Sea $\mathbf{a}_1 = (b_1, b_2, \dots, b_k)'$, donde $'$ indica traspuesto. Entonces $\|\mathbf{a}_1\| = 1$. Podemos extender $\{\mathbf{a}_1\}$ a una base ortonormal de \mathbb{R}^k . Es decir existen vectores columnas $\mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_k$ ortogonales y de norma 1 tales que $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\}$ es una base de \mathbb{R}^k . Luego la matriz B cuyas columnas son los vectores \mathbf{a}_j , $j = 1, 2, \dots, k$ es una matriz ortogonal. Definamos el vector aleatorio $\mathbf{Y} = \mathbf{X}B$, y sea Y_i la componente i -ésima de \mathbf{Y} . Por lo visto anteriormente las variables aleatorias Y_i , ($i = 1, 2, \dots, k$) también son independientes con distribución $N(0, 1)$. En particular $Y_1 = \sum_{i=1}^k b_i X_i = Z$ tiene distribución $N(0, 1)$. Luego (i) queda probado.

(ii) Podemos escribir

$$Z = \sum_{i=1}^k \alpha_i \frac{Y_i - \mu_i}{\sigma_i} \sigma_i + \beta + \sum_{i=1}^k \alpha_i \mu_i = \sum_{i=1}^k \alpha_i \sigma_i X_i + \delta,$$

donde $X_i = (Y_i - \mu_i)/\sigma_i$ y

$$\delta = \beta + \sum_{i=1}^k \alpha_i \mu_i. \quad (6.8)$$

Sabemos que para $i = 1, 2, \dots, k$ las variables X_i son independientes con distribución $N(0, 1)$. Luego podemos escribir a Z de la siguiente manera

$$Z = A \sum_{i=1}^k \frac{\alpha_i \sigma_i}{A} X_i + \delta,$$

donde A está dada por

$$A = \left(\sum_{i=1}^k \alpha_i^2 \sigma_i^2 \right)^{\frac{1}{2}}. \quad (6.9)$$

Sea $b_i = \frac{\alpha_i \sigma_i}{A}$, luego

$$\sum_{i=1}^k b_i^2 = \sum_{i=1}^k \left(\frac{\alpha_i \sigma_i}{A} \right)^2 = \frac{1}{A^2} \sum_{i=1}^k (\alpha_i \sigma_i)^2 = 1.$$

Definamos $W = \sum_{i=1}^k b_i X_i$. Luego de acuerdo a la parte (i) de este teorema se tendrá que

$$W = \sum_{i=1}^k b_i X_i$$

tiene distribución $N(0, 1)$. Por lo tanto como

$$Z = A \sum_{i=1}^k \frac{\alpha_i \sigma_i}{A} X_i + \delta = AW + \delta$$

en virtud de la definición de distribución normal se tendrá que Z tiene distribución $N(\delta, A^2)$. Luego el teorema se deduce de (6.8) y (6.9). \square

6.4. Transformaciones no inyectivas

Vamos a tratar el caso donde g no es inyectiva. En ese caso tenemos el siguiente teorema.

Teorema 6.8 *Sea $\mathbf{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio absolutamente continuo con densidad $f_{\mathbf{X}}$. Sean U_1, U_2, \dots, U_h abiertos disjuntos en \mathbb{R}^k tales que $P_{\mathbf{X}}(\bigcup_{i=1}^h U_i) = 1$. Sea $g : \bigcup_{i=1}^h U_i \rightarrow \mathbb{R}^k$ una función tal que es inyectiva y diferenciable en U_i con $J_g(\mathbf{x}) \neq 0$ para todo $\mathbf{x} \in U_i$. Luego el vector $\mathbf{Y} = g(\mathbf{X})$ también es absolutamente continuo y su densidad está dada por*

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{i=1}^h f_{\mathbf{X}}(g_i^{-1}(\mathbf{y})) |J_{g_i^{-1}}(\mathbf{y})| I_{V_i}(\mathbf{y}),$$

donde $V_i = g(U_i)$, $g_i = g|_{U_i}$, $g_i^{-1} : V_i \rightarrow U_i$ es la inversa de g_i .

Demostración. Bastará probar que para todo $B \in \mathcal{B}^k$ se tiene

$$P_{\mathbf{Y}}(B) = \int_B \cdots \int \sum_{i=1}^h f_{\mathbf{X}}(g_i^{-1}(\mathbf{y})) |J_{g_i^{-1}}(\mathbf{y})| I_{V_i}(\mathbf{y}) d\mathbf{y}. \quad (6.10)$$

Usando que los U_i son disjuntos, que

$$P\left(\bigcup_{i=1}^k U_i\right) = 1$$

y que

$$\{\mathbf{Y} \in B\} \cap \{\mathbf{X} \in U_i\} = \{\mathbf{Y} \in B \cap V_i\} \cap \{\mathbf{X} \in U_i\} = \{\mathbf{X} \in g_i^{-1}(B \cap V_i)\}$$

obtenemos

$$\begin{aligned}
P_{\mathbf{Y}}(B) &= P(\mathbf{Y} \in B) \\
&= P\left(\bigcup_{i=1}^h \{\mathbf{Y} \in B\} \cap \{\mathbf{X} \in U_i\}\right) \\
&= \sum_{i=1}^h P(\{\mathbf{Y} \in B\} \cap \{\mathbf{X} \in U_i\}) \\
&= \sum_{i=1}^h P(\mathbf{X} \in g_i^{-1}(B \cap V_i)) \\
&= \sum_{i=1}^h P_{\mathbf{X}}(g_i^{-1}(B \cap V_i)) \\
&= \sum_{i=1}^h \int \cdots \int_{g_i^{-1}(B \cap V_i)} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}
\end{aligned}$$

Como las funciones g_i son biunívocas en cada U_i , usando la fórmula de cambio de variables en integrales múltiples se tiene

$$\begin{aligned}
P_{\mathbf{Y}}(B) &= \sum_{i=1}^h \int \cdots \int_{g_i^{-1}(B \cap V_i)} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \sum_{i=1}^h \int \cdots \int_{B \cap V_i} f_{\mathbf{X}}(g_i^{-1}(\mathbf{y})) |J_{g_i^{-1}}(\mathbf{y})| d\mathbf{y} \\
&= \sum_{i=1}^h \int \cdots \int_B f_{\mathbf{X}}(g_i^{-1}(\mathbf{y})) |J_{g_i^{-1}}(\mathbf{y})| I_{V_i}(\mathbf{y}) d\mathbf{y} \\
&= \int \cdots \int_B \sum_{i=1}^h f_{\mathbf{X}}(g_i^{-1}(\mathbf{y})) |J_{g_i^{-1}}(\mathbf{y})| I_{V_i}(\mathbf{y}) d\mathbf{y},
\end{aligned}$$

y por lo tanto se cumple (6.10). \square

6.4.1. Distribución Chi-cuadrado con un grado de libertad.

Sea $X \sim N(0, 1)$ y consideremos $g : \mathbb{R} \rightarrow \mathbb{R}$ $g(x) = x^2$. Definimos $Y = g(X) = X^2$. Sean $U_1 = \{x : x < 0\}$ y $U_2 = \{x : x > 0\}$. Luego $g_1^{-1}(y) = -\sqrt{y}$ y $g_2^{-1}(y) = \sqrt{y}$.

En este caso $V_1 = V_2 = \mathbb{R}_{>0}$ y

$$\begin{aligned}
J_{g_1^{-1}}(y) &= -\frac{1}{2}y^{-\frac{1}{2}}, \\
J_{g_2^{-1}}(y) &= \frac{1}{2}y^{-\frac{1}{2}}.
\end{aligned}$$

Luego teniendo en cuenta que

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

y que $V_1 = V_2 = \mathbb{R}_{>0}$, por el teorema anterior se tiene

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) \frac{1}{2} y^{-\frac{1}{2}} I_{V_1}(y) + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) \frac{1}{2} y^{-\frac{1}{2}} I_{V_2}(y) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) y^{-\frac{1}{2}} I_{\{y: y>0\}}(y). \end{aligned}$$

A la distribución de la variable Y la denominaremos distribución Chi-cuadrado con un grado de libertad, y lo notaremos por χ_1^2 .

6.5. Algunas distribuciones complementarias.

6.5.1. Distribución Gamma.

En primer lugar introducimos la función Gamma (que denotaremos con Γ), que resulta ser una extensión a los reales positivos de la función factorial definida sobre los números naturales. La función $\Gamma : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 0}$ se define por

$$\Gamma(\alpha) = \int_0^{+\infty} \exp(-x) x^{\alpha-1} dx.$$

Para probar la existencia de este integral la descomponemos como

$$\begin{aligned} \Gamma(\alpha) &= \int_0^1 \exp(-x) x^{\alpha-1} dx + \int_1^{+\infty} \exp(-x) x^{\alpha-1} dx \\ &= I_1 + I_2. \end{aligned}$$

Es fácil ver que I_1 es finita, teniendo en cuenta que $\exp(-x) \leq 1$ sobre $(0, 1)$

$$I_1 = \int_0^1 \exp(-x) x^{\alpha-1} dx \leq \int_0^1 x^{\alpha-1} dx = \frac{x^\alpha}{\alpha} \Big|_0^1 = \frac{1}{\alpha}.$$

Estudiaremos ahora la convergencia de I_2 . Observemos que el desarrollo de Taylor de $\exp(x/2)$ está dado por

$$\exp\left(\frac{x}{2}\right) = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{x}{2}\right)^k.$$

Luego como todos los términos son positivos, tenemos

$$\exp\left(\frac{x}{2}\right) \geq \frac{1}{k!} \left(\frac{x}{2}\right)^k$$

para todo $k \in \mathbb{N}$.

Entonces

$$x^k \leq C_k \exp\left(\frac{x}{2}\right),$$

donde $C_k = k!2^k$. Tomamos ahora $k_0 > \alpha - 1$, luego se obtiene

$$\begin{aligned} I_2 &= \int_1^{+\infty} \exp(-x) x^{\alpha-1} dx \\ &\leq \int_1^{+\infty} \exp(-x) x^{k_0} dx \\ &\leq \int_1^{+\infty} \exp(-x) C_{k_0} \exp\left(\frac{x}{2}\right) dx \\ &\leq C_{k_0} \int_1^{+\infty} \exp\left(\frac{-x}{2}\right) dx < \infty. \end{aligned}$$

Propiedad 6.1 Si $\alpha > 0$ entonces $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$.

Demostración. Para probarlo integraremos por partes tomando $u = x^\alpha$; $dv = \exp(-x) dx$. Luego se tiene $v = -\exp(-x)$ y $du = \alpha x^{\alpha-1}$, de donde resulta

$$\begin{aligned} \Gamma(\alpha + 1) &= \int_0^{+\infty} \exp(-x) x^\alpha dx \\ &= \int_0^{+\infty} u dv \\ &= -x^\alpha \exp(-x) \Big|_0^\infty - \alpha \int_0^{+\infty} (-\exp(-x)) x^{\alpha-1} dx \\ &= -x^\alpha \exp(-x) \Big|_0^\infty + \alpha \int_0^{+\infty} \exp(-x) x^{\alpha-1} dx. \end{aligned}$$

Como $\lim_{x \rightarrow \infty} x^\alpha \exp(-x) = 0$, resulta que $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$. \square

Propiedad 6.2 Γ es una extensión del factorial. Más precisamente para todo $n \in \mathbb{N}$ se tiene $\Gamma(n) = (n-1)!$

Demostración. La prueba se hace por inducción. Si $n = 1$ entonces $\Gamma(1) = 1 = 0!$. Supongamos ahora que la propiedad que vale para n y veamos que entonces vale para $n + 1$. Usando la Propiedad 6.1 y la hipótesis inductiva tenemos

$$\Gamma(n + 1) = n\Gamma(n) = n((n-1)!) = n!,$$

con lo cual la propiedad queda demostrada. \square

Definición 6.1 Dado $\alpha > 0$, se define la distribución Gamma con parámetros α y 1 (será denotada por $\Gamma(\alpha, 1)$) como la distribución absolutamente continua cuya función densidad es

$$f(x) = \frac{1}{\Gamma(\alpha)} \exp(-x) x^{\alpha-1} I_{[0, \infty)}(x).$$

De acuerdo con la definición de la función Gamma es claro que f es una densidad ya que

$$\int_{-\infty}^{+\infty} f(x) dx = 1.$$

Definición 6.2 Dado $\alpha > 0$ y $\lambda > 0$ definiremos la distribución Gamma con parámetros α y λ (que denotaremos por $\Gamma(\alpha, \lambda)$), a la distribución de $Y = X/\lambda$ donde X tiene distribución $\Gamma(\alpha, 1)$. Como $g(x) = x/\lambda$, De acuerdo a (6.5) y teniendo en cuenta que $\lambda > 0$ tendremos

$$\begin{aligned} f_Y(y) &= \lambda f_X(\lambda y) = \\ &= \frac{\lambda}{\Gamma(\alpha)} \exp(-\lambda y) (\lambda y)^{\alpha-1} I_{[0, \infty)}(\lambda y) = \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \exp(-\lambda y) y^{\alpha-1} I_{[0, \infty)}(y). \end{aligned}$$

Obsérvese que como $\Gamma(1) = 0! = 1$, la distribución $\Gamma(1, \lambda)$ tiene como densidad

$$f(y) = \lambda \exp(-\lambda y) I_{[0, \infty)}(y)$$

que es la distribución exponencial con parámetro λ . En la Figura 6.2 muestran varias densidades gamma

Recordemos que si $X \sim N(0, 1)$ entonces $Y = X^2$ tiene, de acuerdo a lo probado en la subsección anterior, una distribución chi-cuadrado con un grado de libertad. Más precisamente probamos que

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} \exp\left(-\frac{y}{2}\right) I_{[0, \infty)}(y). \quad (6.11)$$

Ahora bien si consideramos $Z \sim \Gamma(1/2, 1/2)$ entonces su densidad es

$$\begin{aligned} f_Z(z) &= \frac{\left(\frac{1}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)} \exp\left(-\frac{z}{2}\right) y^{-\frac{1}{2}} I_{[0, \infty)}(z) \\ &= \frac{1}{\sqrt{2}\Gamma\left(\frac{1}{2}\right)} \exp\left(-\frac{z}{2}\right) y^{-\frac{1}{2}} I_{[0, \infty)}(z). \end{aligned} \quad (6.12)$$

Las densidades (6.11) y (6.12) difieren sólo en una constante, luego deben ser iguales. Esto se muestra integrando las densidades sobre \mathbb{R} , ya que ambas

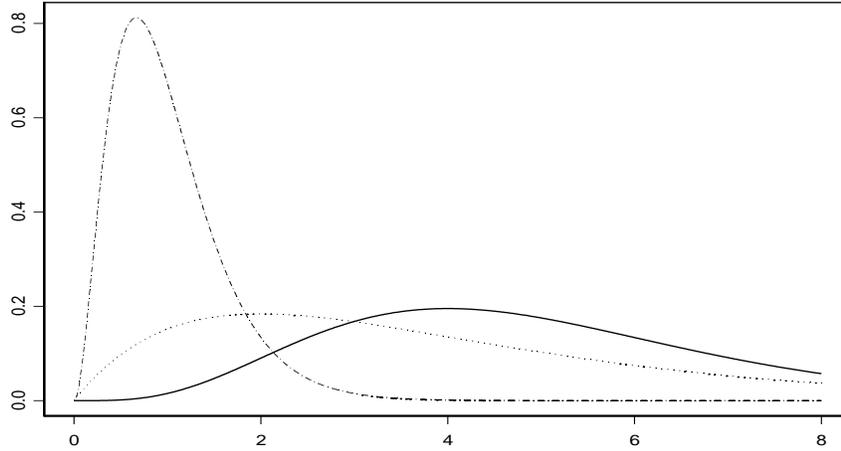


Figura 6.2: Densidad de la $\Gamma\left(2, \frac{1}{2}\right)$ (en línea de puntos y rayas), de la $\Gamma(5, 1)$ (en línea llena) y de la $\Gamma(3, 3)$ (en línea de puntos).

integrales deben ser iguales a 1. Por lo tanto la distribución χ^2 con un grado de libertad coincide con la distribución $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. Además igualando las constantes de ambas densidades se tiene la identidad

$$\frac{1}{\sqrt{2\pi}} = \frac{1}{\sqrt{2}\Gamma\left(\frac{1}{2}\right)},$$

o equivalentemente $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Necesitaremos el siguiente teorema

Teorema 6.9 Sea $\mathbf{W} = (W_1, W_2)$ un vector aleatorio y supongamos que

$$f_{\mathbf{W}}(\mathbf{w}) = g_1(w_1)g_2(w_2),$$

donde g_1 es una función de densidad. Entonces

- (i) $f_{W_2} = g_2$, y por lo tanto g_2 es una función de densidad.
- (ii) $f_{W_1} = g_1$.
- (iii) Las variables W_1 y W_2 son independientes.

Demostración. Como

$$\int_{-\infty}^{+\infty} g_1(w_1) dw_1 = 1,$$

se tiene que

$$\begin{aligned} f_{W_2}(w_2) &= \int_{-\infty}^{+\infty} g_1(w_1) g_2(w_2) dw_1 = \\ &= g_2(w_2) \int_{-\infty}^{+\infty} g_1(w_1) dw_1 = g_2(w_2). \end{aligned}$$

Esto prueba (i). Para ver (ii) se usa el mismo argumento. Como (i) y (ii) implican que

$$f_{\mathbf{W}}(w_1, w_2) = f_{\mathbf{W}_1}(w_1) f_{\mathbf{W}_2}(w_2),$$

resulta que por el Teorema 5.10 W_1 y W_2 son independientes. \square

Teorema 6.10 Sean Y_1, Y_2 variables aleatorias independientes con distribuciones $\Gamma(\alpha_1, \lambda)$ y $\Gamma(\alpha_2, \lambda)$ respectivamente. Definamos $W_1 = Y_1 / (Y_1 + Y_2)$, $W_2 = Y_2 / (Y_1 + Y_2)$. Entonces se tiene

(i) La distribución de W_1 es $W \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$.

(ii) W_2 tiene densidad

$$\frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} w_2^{\alpha_1 - 1} (1 - w_2)^{\alpha_2 - 1} I_{[0,1]}(w_2).$$

(iii) W_1 y W_2 son independientes.

Demostración. La demostración se basa en el Teorema 6.5. Sea el abierto $U \subset \mathbb{R}^2$ definido por $U = \{(y_1, y_2) : y_1 > 0, y_2 > 0\}$. Luego $P_{\mathbf{Y}}(U) = 1$ con $\mathbf{Y} = (Y_1, Y_2)$. Consideremos la transformación $g : U \rightarrow \mathbb{R}^2$ definida por

$$g(y_1, y_2) = \left(y_1 + y_2, \frac{y_1}{y_2 + y_1} \right).$$

Es fácil ver que $V = g(U) = (0, \infty) \times (0, 1)$ y

$$\begin{aligned} g^{-1}(w_1, w_2) &= (w_1 w_2, w_1 - w_1 w_2) \\ &= (w_1 w_2, w_1(1 - w_2)). \end{aligned}$$

Luego

$$\begin{aligned} J_{g^{-1}}(w_1, w_2) &= \det \begin{pmatrix} w_2 & 1 - w_2 \\ w_1 & -w_1 \end{pmatrix} \\ &= -w_1 w_2 - w_1(1 - w_2) \\ &= -w_1, \end{aligned}$$

y por lo tanto $|J_{g^{-1}}(w_1, w_2)| = w_1$.

Consideramos ahora la densidad del vector $\mathbf{Y} = (Y_1, Y_2)$. Como se supuso independencia entre Y_1 e Y_2 , esta densidad es el producto de las densidades marginales y luego

$$f_{\mathbf{Y}}(y_1, y_2) = \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} \exp(-\lambda(y_1 + y_2)) y_1^{\alpha_1 - 1} y_2^{\alpha_2 - 1} I_{(0, \infty)}(y_1) I_{(0, \infty)}(y_2).$$

Luego de acuerdo al Teorema 6.5 y por el hecho de que

$$I_V(w_1, w_2) = I_{(0, \infty) \times (0, 1)}(w_1, w_2) = I_{(0, \infty)}(w_1) I_{(0, 1)}(w_2)$$

se tiene

$$\begin{aligned} f_{\mathbf{W}}(w_1, w_2) &= \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} \exp(-\lambda w_1) (w_1 w_2)^{\alpha_1 - 1} (w_1 (1 - w_2))^{\alpha_2 - 1} w_1 I_V(w_1, w_2) \\ &= \left[\frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} w_1^{\alpha_1 + \alpha_2 - 1} \exp(-\lambda w_1) I_{(0, \infty)}(w_1) \right] \\ &\quad \cdot \left[\frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} w_2^{\alpha_1 - 1} (1 - w_2)^{\alpha_2 - 1} I_{(0, 1)}(w_2) \right] \\ &= g_1(w_1) g_2(w_2) \end{aligned}$$

donde

$$g_1(w_1) = \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} w_1^{\alpha_1 + \alpha_2 - 1} \exp(-\lambda w_1) I_{(0, \infty)}(w_1)$$

y

$$g_2(w_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} w_2^{\alpha_1 - 1} (1 - w_2)^{\alpha_2 - 1} I_{(0, 1)}(w_2).$$

El primer factor g_1 corresponde a una densidad $\Gamma(\alpha_1 + \alpha_2, \lambda)$. Por el Teorema 6.9 resulta que W_1 tiene distribución $\Gamma(\alpha_1 + \alpha_2, \lambda)$ y W_2 tiene como función de densidad a

$$g_2(w_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} w_2^{\alpha_1 - 1} (1 - w_2)^{\alpha_2 - 1} I_{(0, 1)}(w_2).$$

Este teorema también implica que W_1 y W_2 son independientes. \square

6.5.2. Distribución beta.

Definición 6.3 *Se define la distribución beta con parámetros α_1 y α_2 , que denotaremos por $\beta(\alpha_1, \alpha_2)$, como la distribución absolutamente continua cuya función de densidad es:*

$$f(w) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} w^{\alpha_1 - 1} (1 - w)^{\alpha_2 - 1} I_{(0, 1)}(w).$$

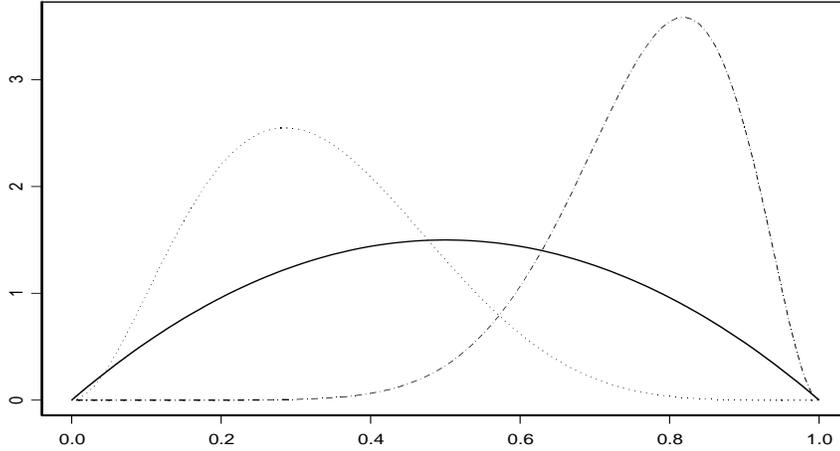


Figura 6.3: Densidad de la $\beta(10, 3)$ (en línea de puntos y rayas), de la $\beta(2, 2)$ (en línea llena) y de la $\beta(3, 6)$ (en línea de puntos).

Observación. Esta función es una densidad por el Teorema 6.10. Por lo tanto podemos deducir que

$$\int_0^1 \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} w^{\alpha_1-1} (1-w)^{\alpha_2-1} dw = 1,$$

y entonces se tiene

$$\int_0^1 w^{\alpha_1-1} (1-w)^{\alpha_2-1} dw = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}.$$

En la Figura 6.3 se muestran varias densidades Beta, para distintos valores de los parámetros α_1 y α_2 .

Teorema 6.11 Sean Y_1, Y_2, \dots, Y_n variables aleatorias independientes tales que Y_i tiene distribución $\Gamma(\alpha_i, \lambda)$. Entonces $\sum_{i=1}^n Y_i$ tiene distribución $\Gamma(\sum_{i=1}^n \alpha_i, \lambda)$.

Demostración. Se deduce de la proposición anterior usando inducción. \square

A continuación definimos las distribuciones chi-cuadrado con n grados de libertad y la t de Student. Ambas distribuciones son de gran importancia en Estadística. Volveremos más adelante sobre ellas.

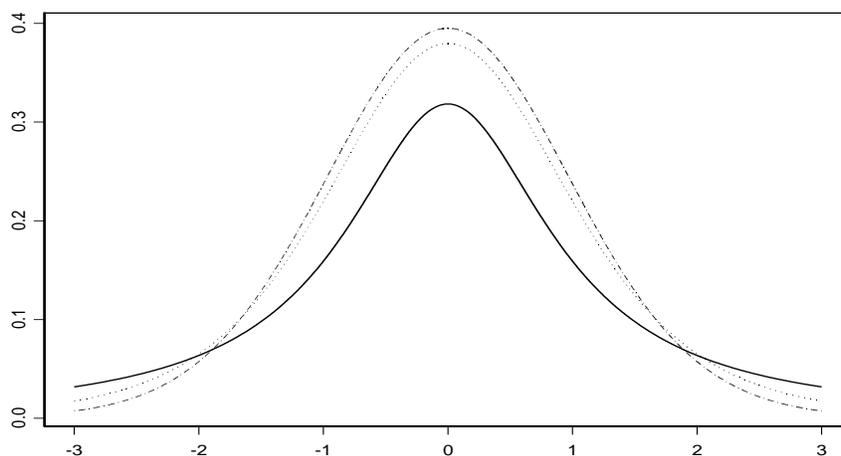


Figura 6.4: Densidad de la t_1 (en línea llena), de la t_5 (en línea de puntos) y de la t_{25} (en línea de rayas).

6.5.3. Distribución Chi-cuadrado.

Supongamos que se tienen n variables independientes X_i , $i = 1, 2, \dots, n$ con distribución $N(0, 1)$. Sabemos que cada $Y_i = X_i^2$ tiene distribución χ^2 con 1 grado de libertad, la cual que coincide con la distribución $\Gamma(1/2, 1/2)$.

Se define la distribución chi-cuadrado con n grados de libertad, que simbolizaremos por χ_n^2 , como la distribución de la variable aleatoria $Y = \sum_{i=1}^n X_i^2$.

De acuerdo al Teorema 6.11, como cada X_i^2 tiene distribución χ_1^2 y estas variables son independientes, se obtiene que Y tiene distribución $\Gamma(n/2, 1/2)$. Por lo tanto la distribución χ_n^2 coincide con la distribución $\Gamma(n/2, 1/2)$.

6.5.4. Distribución t de Student

Supongamos que U tiene distribución $N(0, 1)$ y V distribución χ_n^2 con U y V independientes. Luego se define la distribución de t de Student con n grados de libertad, que simbolizaremos con t_n , como la distribución de

$$T = \frac{U}{\sqrt{V/n}}.$$

En la Figura 6.4 se muestran varias densidades de Student para diferentes grados de libertad

Se deja como ejercicio de la práctica mostrar que la densidad de T es

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

El gráfico de esta densidad es simétrico respecto al origen (función par) y con forma de campana. Se puede probar que cuando n tiende a ∞ , f_T converge a la densidad de la normal.

Capítulo 7

Esperanza Matemática.

7.1. Integral de Riemann-Stieltjes.

7.1.1. Definición de la integral.

Sea $f : [a, b] \rightarrow \mathbb{R}$ y consideremos una partición del intervalo $[a, b]$ que llamaremos $\pi = \{x_0, x_1, \dots, x_n\}$ tal que $a = x_0 < x_1 < \dots < x_n = b$. Sea $\xi = \{\xi_i\}_{1 \leq i \leq n}$ una colección de puntos tal que $\xi_i \in (x_{i-1}, x_i]$ para $i = 1, 2, \dots, n$, que se denominará selección en π .

Definimos la suma de Riemann

$$S_a^b(\pi, \xi, f) = \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}).$$

Se llama *norma de la partición*

$$\|\pi\| = \max_{1 \leq i \leq n} \{x_i - x_{i-1}\}.$$

Definición 7.1 *Se dice que f es integrable Riemann sobre $[a, b]$ con valor $I = \int_a^b f = \int_a^b f(x) dx$ sii para todo $\varepsilon > 0$ existe $\delta > 0$ tal que si $\|\pi\| < \delta$ entonces*

$$|S_a^b(\pi, \xi, f) - I| < \varepsilon.$$

Análogamente se define la *integral de Riemann-Stieltjes*. Dadas g, F funciones definidas sobre $[a, b]$ se define la suma de Riemann-Stieltjes asociada a la partición $\pi = \{x_i\}_{0 \leq i \leq n}$ y la selección $\xi = \{\xi_i\}_{1 \leq i \leq n}$ de π por

$$S_a^b(\pi, \xi, g, F) = \sum_{i=1}^n f(\xi_i)(F(x_i) - F(x_{i-1})).$$

Definición 7.2 Se dice que existe la integral de Riemann-Stieltjes sobre $[a, b]$ con valor $I = \int_a^b g dF = \int_a^b g(x) dF(x)$ sii para todo $\varepsilon > 0$ existe $\delta > 0$ tal que si π es una partición de $[a, b]$ con $\|\pi\| < \delta$ y ξ es cualquier selección en π entonces

$$|S_a^b(\pi, \xi, g, F) - I| < \varepsilon.$$

Observaciones.

1. Si $F(x) = x$, entonces la integral de Riemann-Stieltjes es la integral de Riemann.
2. Una condición suficiente, aunque no necesaria, para que exista la integral de Riemann-Stieltjes, es que g sea continua en $[a, b]$ y F monótona en $[a, b]$. Si tomamos como F una función de distribución el último requisito se cumplirá.
3. Otra condición suficiente (tampoco necesaria) para que exista la integral de Riemann-Stieltjes es que (i) g sea continua en $(a, b]$, (ii) existe $\lim_{x \downarrow a} g(x)$, (iii) F sea monótona en $[a, b]$ y (iv) F es continua en a . En tal caso, vale que

$$\int_a^b g dF = \lim_{c \downarrow a} \int_c^b g dF.$$

A continuación damos algunas propiedades de la integral de Riemann Stieltjes.

Propiedad 7.1 (Linealidad de la Integral de R-S respecto de g) Si $\int_a^b g_1 dF$ y $\int_a^b g_2 dF$ existen y $\alpha_1, \alpha_2 \in \mathbb{R}$ entonces $\int_a^b (\alpha_1 g_1 + \alpha_2 g_2) dF$ existe y además

$$\int_a^b (\alpha_1 g_1 + \alpha_2 g_2) dF = \alpha_1 \int_a^b g_1 dF + \alpha_2 \int_a^b g_2 dF.$$

Propiedad 7.2 (Linealidad de la Integral R-S respecto de F) Si $\int_a^b g dF_1$ y $\int_a^b g dF_2$ existen y $\alpha_1, \alpha_2 \in \mathbb{R}$ entonces $\int_a^b g d(\alpha_1 F_1 + \alpha_2 F_2)$ existe y además

$$\int_a^b g d(\alpha_1 F_1 + \alpha_2 F_2) = \alpha_1 \int_a^b g dF_1 + \alpha_2 \int_a^b g dF_2.$$

Propiedad 7.3 (Aditividad respecto del dominio de integración) Sean $a < b < c$ y supongamos que $\int_a^b g dF$, $\int_b^c g dF$ y $\int_a^c g dF$ existen. Entonces

$$\int_a^c g dF = \int_a^b g dF + \int_b^c g dF.$$

Propiedad 7.4 Si F es no decreciente y $g_1 \leq g_2$ sobre $[a, b]$ entonces

$$\int_a^b g_1 dF \leq \int_a^b g_2 dF.$$

En particular teniendo en cuenta que $-|g| \leq g \leq |g|$ se obtiene la siguiente

Propiedad 7.5 Si las dos integrales existen, entonces

$$\left| \int_a^b g dF \right| \leq \int_a^b |g| dF$$

Estamos interesados en extender el dominio de integración a toda la recta o a semirectas. Esto lleva a la siguiente definición.

Definición 7.3 Supongamos que $\int_a^b g dF$ existe para todo $a, b \in \mathbb{R}$. Decimos que la integral impropia $\int_{-\infty}^{+\infty} g dF$ existe y es igual al número real I sii

$$\lim_{a \rightarrow -\infty; b \rightarrow +\infty} \int_a^b g dF = I. \quad (7.1)$$

De manera análoga se define $\int_a^{+\infty} g dF$ y $\int_{-\infty}^b g dF$. Tendremos el siguiente teorema.

Teorema 7.1 Sea $g \geq 0$ y F no decreciente. Entonces pueden ocurrir dos cosas

(i)

$$M = \sup_{a, b \in \mathbb{R}} \int_a^b g dF < \infty$$

En este caso el límite (7.1) existe y es finito.

(ii)

$$M = \sup_{a,b \in \mathbb{R}} \int_a^b g dF = \infty$$

En este caso el límite (7.1) existe y es ∞ . Luego podemos definir $\int_{-\infty}^{+\infty} g dF = \infty$.

Sea ahora g de signo arbitrario y F no decreciente. El siguiente teorema es válido.

Teorema 7.2 Una condición necesaria y suficiente para que $\int_{-\infty}^{+\infty} g dF$ exista es que

$$\widetilde{M} = \sup_{a,b \in \mathbb{R}} \int_a^b |g| dF < \infty.$$

7.2. Definición de Esperanza Matemática.

7.2.1. Algunas consideraciones heurísticas.

Sea X una variable aleatoria discreta. Para fijar ideas supongamos que toma un número finito de valores, x_1, x_2, \dots, x_k , con probabilidades $p_X(x_1), p_X(x_2), \dots, p_X(x_k)$.

Supongamos que se repite un experimento asociado a la variable aleatoria X , n veces en forma independiente y que el resultado x_i se obtiene n_i veces, $1 \leq i \leq k$. Entonces el promedio de todos los valores es

$$\begin{aligned} \bar{x}_n &= \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n} \\ &= \frac{n_1}{n} x_1 + \frac{n_2}{n} x_2 + \dots + \frac{n_k}{n} x_k. \end{aligned}$$

Luego pasando al límite y dado que la frecuencia observada $\frac{n_j}{n}$ se aproxima a $p_X(x_j)$ obtenemos

$$\begin{aligned} \lim_{n \rightarrow +\infty} \bar{x}_n &= \lim_{n \rightarrow +\infty} \left(\frac{n_1}{n} x_1 + \frac{n_2}{n} x_2 + \dots + \frac{n_k}{n} x_k \right) \\ &= x_1 \lim_{n \rightarrow +\infty} \frac{n_1}{n} + x_2 \lim_{n \rightarrow +\infty} \frac{n_2}{n} + \dots + x_k \lim_{n \rightarrow +\infty} \frac{n_k}{n} \\ &= \sum_{j=1}^k x_j p_X(x_j). \end{aligned}$$

Esto motiva la definición de la esperanza matemática de una variable discreta.

7.2.2. Esperanza de una variable aleatoria discreta.

Definición 7.4 Sea X una variable aleatoria con rango R_X y distribución de probabilidad p_X . Supongamos que

$$\sum_{x \in R_X} |x|p_X(x) < \infty.$$

En tal caso definimos la esperanza matemática de la variable X de la siguiente manera

$$E(X) = \sum_{x \in R_X} xp_X(x).$$

Observaciones.

1. Se sabe que la convergencia absoluta de la serie garantiza la convergencia de la serie.
2. Supongamos $\sum_{x \in R_X} |x|p_X(x) = \infty$. Denotemos con

$$R_X^+ = \{x \in R_X : x > 0\}$$
$$R_X^- = \{x \in R_X : x < 0\}.$$

Entonces pueden ocurrir tres casos distintos.

- a) $\sum_{x \in R_X^+} xp_X(x) = +\infty$ y $\sum_{x \in R_X^-} xp_X(x) = -\infty$.
- b) $\sum_{x \in R_X^+} xp_X(x) = +\infty$ y $\sum_{x \in R_X^-} xp_X(x) > -\infty$.
- c) $\sum_{x \in R_X^+} xp_X(x) < +\infty$ y $\sum_{x \in R_X^-} xp_X(x) = -\infty$.

En el caso (a) no se puede definir la esperanza de X . En el caso (b) se puede definir $E(X) = +\infty$ en el (c) $E(X) = -\infty$. Es decir para que la esperanza esté definida se requiere que $\sum_{x \in R_X^+} xp_X(x)$ o bien $\sum_{x \in R_X^-} xp_X(x)$ sea finita.

7.2.3. Definición general de esperanza matemática.

Ahora queremos definir la esperanza matemática, de manera más general. Supongamos primero que X es una variable aleatoria concentrada en $[a, b]$. Es decir, supongamos que

$$P(a < X < b) = 1.$$

La idea que se utiliza para la definición de la esperanza de esta variable es la siguiente. Se define una sucesión de variables aleatorias discretas X_n que la aproximan y luego como $E(X_n)$ está definida para cada X_n la esperanza de X se define por un paso al límite.

Consideremos para cada n , una partición del intervalo $[a, b]$ formada por n intervalos de longitud $(b - a)/n$. Para esto consideramos la partición $\pi^n = \{x_0^n, x_1^n, \dots, x_n^n\}$ tal que $a = x_0^n < x_1^n < \dots < x_n^n = b$ y $x_i^n - x_{i-1}^n = \frac{b - a}{n}$.

Elegimos para cada i , $1 \leq i \leq n$, $\xi_i^n \in (x_{i-1}^n, x_i^n]$ y definimos la variable aleatoria

$$X_n(\omega) = \xi_i^n \text{ si } X(\omega) \in (x_{i-1}^n, x_i^n].$$

Esta variable toma únicamente un número finito de valores: ξ_i^n , $1 \leq i \leq n$. Además

$$p_{X_n}(\xi_i^n) = F_X(x_i^n) - F_X(x_{i-1}^n).$$

Luego la esperanza de la variable X_n viene dada por

$$\begin{aligned} E(X_n) &= \sum_{i=1}^n \xi_i^n p_{X_n}(\xi_i^n) \\ &= \sum_{i=1}^n \xi_i^n (F_X(x_i^n) - F_X(x_{i-1}^n)) = S_a^b(\pi^n, \xi^n, id, F), \end{aligned}$$

con $id(x) = x$ y se obtiene

$$\lim_{n \rightarrow +\infty} E(X_n) = \lim_{n \rightarrow +\infty} S_a^b(\pi^n, \xi^n, id, F_X) = \int_a^b x dF_X.$$

Por lo tanto definimos la esperanza matemática de X por

$$E(X) = \int_a^b x dF_X.$$

Siendo la función $id(x) = x$ continua y F monótona no decreciente, resulta que $\int_a^b x dF$ existe siempre y por lo tanto también $E(X)$ existe siempre.

Supongamos ahora que X es una variable aleatoria no acotada. El problema que ahora surge es que podría no existir $\int_{-\infty}^{+\infty} x dF$. Sin embargo sabemos que $M = \int_{-\infty}^{+\infty} |x| dF$ siempre está bien definida, eventualmente con el valor $+\infty$.

Si $M < +\infty$ definimos la esperanza de la variable X similarmente al caso anterior por

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

Si $M = +\infty$ hay tres casos y el análisis es análogo al que realizamos anteriormente para variables discretas. Los tres casos son:

(a) $\int_0^{+\infty} x dF = +\infty$ y $\int_{-\infty}^0 x dF = -\infty$.

(b) $\int_0^{+\infty} x dF = +\infty$ y $\int_{-\infty}^0 x dF > -\infty$.

$$(c) \int_0^\infty x dF < +\infty \text{ y } \int_{-\infty}^0 x dF = -\infty.$$

En el caso (a) la esperanza matemática de X no está definida. En el caso (b) se define $E(X) = +\infty$ y en el (c) $E(X) = -\infty$. Nuevamente la esperanza puede estar no definida y para su definición se requiere que al menos una de las dos integrales $\int_0^\infty x dF$ ó $\int_{-\infty}^0 x dF$ converja.

Con esta definición general de esperanza matemática, para el caso de una variable discreta se tienen dos definiciones diferentes. Probaremos ahora que la definición general de esperanza es una extensión de la primera definición dada para el caso discreto, es decir que para variables aleatorias discretas ambas definiciones coinciden.

Teorema 7.3 *Sea F_X la función de distribución de una variable discreta y $g : \mathbb{R} \rightarrow \mathbb{R}$ continua. Luego*

$$\int_a^b g(x) dF_X(x) = \sum_{x \in R_X \cap [a, b]} g(x) p_X(x). \quad (7.2)$$

Observación. Este resultado vale siempre, pero para facilitar la demostración vamos a probarlo para el caso en que $R_X \cap [a, b]$ es finito para todo a y b . Esto se cumple cuando las variables toman valores enteros como sucede, por ejemplo, con las distribuciones Poisson, binomial, etc.

Demostración. Por la hipótesis supuesta $R_X \cap [a, b]$ es un conjunto finito, digamos

$$R_X \cap [a, b] = \{z_1, z_2, \dots, z_k\}.$$

Llamemos δ a

$$\delta = \min_{2 \leq i \leq k} \{z_i - z_{i-1}\}. \quad (7.3)$$

Consideremos una partición $\pi^n = \{x_i^n\}_{0 \leq i \leq n}$ del intervalo $[a, b]$, en n intervalos iguales. Luego tenemos $a = x_0^n < x_1^n < \dots < x_n^n = b$ y $x_i^n - x_{i-1}^n = (b - a)/n$. Teniendo en cuenta que $\|\pi^n\| = (b - a)/n$ es claro que

$$\lim_{n \rightarrow +\infty} \|\pi^n\| = 0.$$

Sea n_0 tal que $(b - a)/n_0 < \delta$. Tomemos $n > n_0$, luego $\|\pi^n\| < \delta$, luego por (7.3) en cada intervalo de π^n hay a lo sumo un elemento de $R_X \cap [a, b]$. Va a ser fundamental para esta demostración la elección de la selección $\xi^n = \{\xi_i^n\}_{1 \leq i \leq n}$ de π^n . Procedemos de la siguiente manera.

(i) Si

$$(R_X \cap [a, b]) \cap (x_{i-1}^n, x_i^n] \neq \emptyset$$

se elige como ξ_i^n el único punto de esta intersección.

(ii) Si

$$(R_X \cap [a, b]) \cap (x_{i-1}^n, x_i^n) = \emptyset$$

ξ_i^n es cualquier punto de $(x_{i-1}, x_i]$.

Sea

$$A = \{i : (R_X \cap [a, b]) \cap (x_{i-1}^n, x_i^n) \neq \emptyset\}$$

y por lo tanto

$$A^c = \{i : (R_X \cap [a, b]) \cap (x_{i-1}^n, x_i^n) = \emptyset\}$$

Entonces podemos realizar la siguiente descomposición de $S_a^b(\pi^n, \xi^n, g, F)$

$$\begin{aligned} S_a^b(\pi^n, \xi^n, g, F) &= \sum_{i=1}^n g(\xi_i^n) [F_X(x_i^n) - F_X(x_{i-1}^n)] \\ &= \sum_{i \in A} g(\xi_i^n) [F_X(x_i^n) - F_X(x_{i-1}^n)] \\ &\quad + \sum_{i \in A^c} g(\xi_i^n) [F_X(x_i^n) - F_X(x_{i-1}^n)]. \end{aligned}$$

Observemos que $F_X(x_i^n) - F_X(x_{i-1}^n) = 0$ si $i \in A^c$ ya que el intervalo $(x_{i-1}, x_i]$ no contiene elementos de R_X . Luego

$$\sum_{i \in A^c} g(\xi_i^n) [F_X(x_i^n) - F_X(x_{i-1}^n)] = 0,$$

y se obtiene

$$S_a^b(\pi^n, \xi^n, g, F_X) = \sum_{i \in A} g(\xi_i^n) [F_X(x_i^n) - F_X(x_{i-1}^n)]. \quad (7.4)$$

Además, como para $i \in A$, el valor ξ_i^n es el único punto de R_X en el intervalo $(x_{i-1}^n, x_i^n]$, resulta

$$p_X(\xi_i^n) = P_X((x_{i-1}^n, x_i^n]) = F_X(x_i^n) - F_X(x_{i-1}^n).$$

Luego de (7.4) obtenemos

$$S_a^b(\pi^n, \xi^n, g, F_X) = \sum_{i \in A} g(\xi_i^n) p_X(\xi_i^n).$$

Pero $(\xi_i^n)_{i \in A}$ coincide con $\{z_j\}_{1 \leq j \leq k} = R_X \cap [a, b]$, y entonces para todo $n \geq n_0$

$$S_a^b(\pi^n, \xi^n, g, F_X) = \sum_{j=1}^k g(z_j) p_X(z_j) = \sum_{x \in R_X \cap [a, b]} g(x) p_X(x). \quad (7.5)$$

Como el miembro derecho de (7.5) no depende de n , obtenemos

$$\int_a^b x dF = \lim_{n \rightarrow \infty} S_a^b(\pi^n, \xi^n, g, F_X) = \sum_{x \in R_X \cap [a, b]} xp_X(x).$$

Esto prueba (7.2) y por lo tanto el teorema queda demostrado. \square

Teorema 7.4 *Supongamos que X es una variable aleatoria discreta y que $E(X)$ existe y es finita. Entonces*

$$\sum_{x \in R_X} xp_X(x) = \int_{-\infty}^{+\infty} x dF_X$$

Demostración. Teniendo en cuenta que

$$\sum_{x \in R_X} xp_X(x) = \lim_{a \rightarrow -\infty; b \rightarrow +\infty} \sum_{x \in R_X \cap [a, b]} xp_X(x),$$

y que

$$\int_{-\infty}^{+\infty} x dF_X = \lim_{a \rightarrow -\infty; b \rightarrow +\infty} \int_a^b x dF_X,$$

bastará probar que para todo $a < b$

$$\sum_{x \in R_X \cap [a, b]} xp_X(x) = \int_a^b x dF_X.$$

Pero esto resulta del teorema 7.3 poniendo $g(x) = x$. \square

7.2.4. Esperanza matemática para una variable absolutamente continua.

El siguiente Teorema prueba que en el caso de que X sea una variable aleatoria absolutamente continua la $E(X)$ se puede calcular a través de una integral de Riemann.

Teorema 7.5 *Supongamos que $\int_{-\infty}^{\infty} |x|f_X(x) dx < \infty$. Luego*

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

Demostración. El teorema vale en general. Sin embargo, para facilitar la demostración, lo probaremos sólo para el caso en que f_X es continua.

Bastará ver que para todo intervalo $[a, b]$, $a < b$ vale que

$$\int_a^b x f_X(x) dx = \int_a^b x dF_X, \quad (7.6)$$

ya que en tal caso el resultado se obtiene pasando al límite.

Consideremos para cada n una partición de puntos equidistantes del intervalo $[a, b]$

$$\pi^n = \{x_0^n, x_1^n, \dots, x_n^n\}$$

tales que $a = x_0^n < x_1^n < \dots < x_n^n = b$ satisfaciendo $x_i^n - x_{i-1}^n = \frac{b-a}{n}$.

Sabemos que $F_X'(x) = f_X(x)$. Por el Teorema del Valor Medio, para todo i , $1 \leq i \leq n$, existe $\xi_i^n \in (x_i^n, x_{i-1}^n]$ tal que

$$F_X(x_i^n) - F_X(x_{i-1}^n) = f_X(\xi_i^n)(x_i^n - x_{i-1}^n). \quad (7.7)$$

Elegiremos la selección $\xi = (\xi_i^n)_{1 \leq i \leq n}$ para formar las sumas de Riemann-Stieltjes. Luego

$$S_a^b(\pi^n, \xi^n, id, F_X) = S_a^b(\pi^n, \xi^n, x, F_X) = \sum_{i=1}^n \xi_i^n (F_X(x_i^n) - F_X(x_{i-1}^n)), \quad (7.8)$$

y se tendrá que

$$\lim_{n \rightarrow \infty} S_a^b(\pi^n, \xi^n, x, F_X) = \int_a^b x dF_X. \quad (7.9)$$

Usando (7.7) y (7.8) obtenemos que $S_a^b(\pi^n, \xi^n, x, F_X)$ es también una suma de Riemann correspondiente a la función $x f_X(x)$. En efecto

$$\begin{aligned} S_a^b(\pi^n, \xi^n, x, F_X) &= \sum_{i=1}^n \xi_i^n f_X(\xi_i^n)(x_i^n - x_{i-1}^n) \\ &= S_a^b(\pi^n, \xi^n, x f_X(x), x). \end{aligned}$$

Luego

$$\lim_{n \rightarrow \infty} S_a^b(\pi^n, \xi^n, x, F_X) = \int_a^b x f_X(x) dx. \quad (7.10)$$

De (7.9) y (7.10) se obtiene (7.6). \square

7.2.5. Algunas propiedades de la esperanza matemática

Propiedad 7.6 Sea X una variable aleatoria tal que $P_X(\{a\}) = 1$. Entonces

$$E(X) = a.$$

Demostración. Esto es inmediato teniendo en cuenta X es una variable discreta con $R_X = \{a\}$ y $p_X(a) = 1$. Luego

$$E(X) = \sum_{x \in R_X} xp_X(x) = a. \square$$

Propiedad 7.7 Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y $A \in \mathcal{A}$. Entonces $E(I_A) = P(A)$.

Demostración. Como

$$I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A. \end{cases}$$

En este caso $R_X = \{0, 1\}$, $p_X(1) = P(A)$, y $p_X(0) = 1 - P(A)$. Entonces

$$E(I_A) = 0(1 - P(A)) + 1P(A) = P(A). \square$$

El siguiente teorema permite la integración por partes de una integral de Riemann-Stieltjes.

Teorema 7.6 (Integración por partes) Sean g y F funciones definidas sobre $[a, b]$ tales que $\int_a^b g dF$ existe. Supongamos que g sea continua en a y que F es acotada en $[a, b]$. Entonces $\int_a^b F dg$ existe y

$$\int_a^b g dF = g(x)F(x) \Big|_a^b - \int_a^b F dg.$$

Demostración. Tenemos que mostrar que

$$\int_a^b F dg = g(x)F(x) \Big|_a^b - \int_a^b g dF. \quad (7.11)$$

Para eso habrá que probar que dado $\varepsilon > 0$ existe $\delta > 0$ tal que para toda $\pi = \{x_i\}_{0 \leq i \leq n}$ partición de $(a, b]$ con $\|\pi\| \leq \delta$ y toda $\xi = \{\xi_i\}_{0 \leq i \leq n}$ selección de puntos en π , se tendrá que

$$\left| S_a^b(\pi, \xi, F, g) - g(x)F(x) \Big|_a^b + \int_a^b g dF \right| < \varepsilon. \quad (7.12)$$

Como $\int_a^b g dF$ existe, dado $\frac{\varepsilon}{2}$ podemos encontrar un δ_1 tal que si $\|\pi\| \leq \delta_1$ para toda selección ξ en π tendremos que

$$\left| S_a^b(g, f, \pi, \xi) - \int_a^b g dF \right| \leq \frac{\varepsilon}{2}. \quad (7.13)$$

Como F es acotada en $[a, b]$ existe un número real $M > 0$ tal que

$$|F(x)| \leq M$$

para todo $x \in [a, b]$. Por la continuidad de g en a , sabemos que existe $\delta_2 > 0$ tal que si $|x - a| \leq \delta_2$ entonces

$$|g(x) - g(a)| < \frac{\varepsilon}{4M}.$$

Pongamos $\delta = \min(\frac{\delta_1}{2}, \delta_2)$. Sea $\pi = \{x_i\}_{0 \leq i \leq n}$ una partición de $(a, b]$, tal que $|\pi| \leq \delta$ y sea $\xi = \{\xi_i\}_{0 \leq i \leq n}$ una selección en la partición.

Vamos a mostrar que (7.12) vale. Sabemos que $x_{n-1} < \xi_n \leq b$. Supondremos que $\xi_n < b$. El caso $\xi_n = b$ se demuestra análogamente. Tenemos que

$$a = x_0 < \xi_1 \leq x_1 < \cdots < \xi_{i-1} \leq x_{i-1} < \xi_i \leq x_i < \cdots < x_{n-1} < \xi_n < x_n = b.$$

Podemos construir una nueva partición $\pi^* = \{x_i^*\}_{0 \leq i \leq n+1}$ con

$$\begin{aligned} x_0^* &= a, \\ x_i^* &= \xi_i, \quad 1 \leq i \leq n, \\ x_{n+1}^* &= b, \end{aligned}$$

y definimos la selección $\xi^* = (\xi_i^*)_{1 \leq i \leq n+1}$ en π^* por

$$\begin{aligned} \xi_1^* &= \xi_1, \\ \xi_i^* &= x_{i-1}, \quad 2 \leq i \leq n+1. \end{aligned}$$

Como

$$\begin{aligned} |x_i^* - x_{i-1}^*| &= |\xi_i - \xi_{i-1}| \leq |\xi_i - x_{i-1}| + |x_{i-1} - \xi_{i-1}| \\ &\leq |x_{i-1} - x_i| + |x_{i-1} - x_{i+1}| \\ &< \delta + \delta = 2\delta \leq \delta_1, \quad \text{para } 2 \leq i \leq n \\ |x_1^* - x_0^*| &= |\xi_1 - a| = |\xi_1 - x_0| \leq |x_1 - x_0| < \delta \leq \delta_1 \\ |x_{n+1}^* - x_n^*| &= |b - \xi_n| = |x_n - \xi_n| \leq |x_n - x_{n-1}| < \delta \leq \delta_1 \end{aligned}$$

tenemos que $|\pi^*| \leq \delta_1$ y entonces por (7.13) resulta

$$\left| S_a^b(\pi^*, \xi^*, g, F) - \int_a^b g dF \right| < \frac{\varepsilon}{2}. \quad (7.14)$$

Por otro lado tenemos

$$\begin{aligned}
S_a^b(\pi^*, \xi^*, g, F) &= \sum_{i=1}^{n+1} g(\xi_i^*) [F(x_i^*) - F(x_{i-1}^*)] \\
&= g(\xi_1^*)F(x_1^*) + \sum_{i=2}^n g(\xi_i^*)F(x_i^*) + g(\xi_{n+1}^*)F(x_{n+1}^*) \\
&\quad - g(\xi_1^*)F(x_0^*) - \sum_{i=2}^{n+1} g(\xi_i^*)F(x_{i-1}^*) \\
&= g(\xi_1)F(\xi_1) + \sum_{i=2}^n g(x_{i-1})F(\xi_i) + g(b)F(b) \\
&\quad - g(\xi_1)F(a) \\
&= g(\xi_1)F(\xi_1) - g(\xi_1)F(a) + \sum_{i=2}^n g(x_{i-1})F(\xi_i) \\
&\quad + g(b)F(b) - \sum_{i=1}^n g(x_i)F(\xi_i) \\
&= g(\xi_1) [F(\xi_1) - F(a)] - \sum_{i=1}^n [g(x_{i-1}) - g(x_i)] F(\xi_i) \\
&\quad + g(b)F(b) - g(x_0)F(\xi_1) \\
&= -\sum_{i=1}^n F(\xi_i) [g(x_{i-1}) - g(x_i)] + g(b)F(b) - g(a)F(a) \\
&\quad + g(\xi_1) [F(\xi_1) - F(a)] + g(a)F(a) - g(a)F(\xi_1) \\
&= -S_a^b(\pi, \xi, F, g) + g(x)F(x)|_a^b \\
&\quad + g(\xi_1) [F(\xi_1) - F(a)] + g(a) [F(a) - F(\xi_1)] \\
&= -S_a^b(\pi, \xi, F, g) + g(x)F(x)|_a^b + [g(\xi_1) - g(a)] [F(\xi_1) - F(a)] \\
&= -S_a^b(F, g, \pi, \xi) + g(x)F(x)|_a^b + r, \tag{7.15}
\end{aligned}$$

donde $r = [g(\xi_1) - g(a)] [F(\xi_1) - F(a)]$. Luego, como $\|\pi^*\| < \delta$ y $|x_0^* - x_1^*| = |a - \xi_1| < \delta \leq \delta_2$ se tendrá

$$|g(a) - g(\xi_1)| \leq \varepsilon/4M.$$

Además $|F(x)| \leq M$, y entonces obtenemos

$$\begin{aligned}
|r| &= |F(\xi_1) - F(a)| |g(\xi_1) - g(a)| \\
&\leq 2M \frac{\varepsilon}{4M} = \frac{\varepsilon}{2}.
\end{aligned}$$

Luego de (7.15) resulta.

$$\left| S_a^b(\pi^*, \xi^*, g, F) - g(x)F(x)|_a^b + S_a^b(\pi, \xi, F, g) \right| \leq \frac{\varepsilon}{2}. \tag{7.16}$$

De (7.14) y (7.16) resulta (7.12) y el teorema queda demostrado. \square

Propiedad 7.8 *Dada una función F monótona se tiene*

$$\int_a^b dF = F(b) - F(a).$$

Demostración. Aplicando integración por partes con $g = 1$ y dado que $dg = 0$, obtenemos

$$\int_a^b dF = 1F(x)|_a^b - \int_a^b F dg = F_X(x)|_a^b = F(b) - F(a). \square$$

Teorema 7.7 *Supongamos que $\int_{-\infty}^{+\infty} |x|dF_X < \infty$. Entonces vale*

(i) $\lim_{x \rightarrow +\infty} x(1 - F_X(x)) = 0$.

(ii) $\lim_{x \rightarrow -\infty} xF_X(x) = 0$.

Demostración.

(i) A partir del hecho de que $\int_{-\infty}^{\infty} |x|dF_X$ es finita se deduce que “las colas” tienden a cero, es decir

$$\lim_{b \rightarrow +\infty} \int_b^{+\infty} x dF_X = 0, \quad (7.17)$$

y

$$\lim_{a \rightarrow -\infty} \int_{-\infty}^a x dF_X = 0. \quad (7.18)$$

Usando la Propiedad 7.8 obtenemos

$$\int_b^{+\infty} dF_X = \lim_{d \rightarrow \infty} \int_b^d dF_X = \lim_{d \rightarrow \infty} F_X(d) - F_X(b) = 1 - F_X(b),$$

y entonces si $b \geq 0$

$$\int_b^{+\infty} x dF_X \geq b \int_b^{+\infty} dF_X = b(1 - F_X(b)) \geq 0.$$

Luego

$$0 = \lim_{b \rightarrow \infty} \int_b^{+\infty} x dF_X \geq \lim_{b \rightarrow \infty} b(1 - F_X(b)) \geq 0.$$

Luego se deduce (i).

(ii) Se prueba de manera análoga y se deja como ejercicio. \square

Ahora estamos en condiciones de dar una expresión de la esperanza como sumas de integrales de Riemann.

Teorema 7.8 *Supongamos que $\int_{-\infty}^{\infty} |x|dF_X < \infty$. Entonces*

$$E(X) = \int_0^{+\infty} (1 - F_X(x)) dx - \int_{-\infty}^0 F_X(x) dx. \quad (7.19)$$

Demostración. Sabemos que

$$E(X) = \int_0^{+\infty} x dF_X + \int_{-\infty}^0 x dF_X.$$

Estudiaremos cada integral por separado. Integrando por partes tenemos que

$$\begin{aligned} \int_0^b x dF_X &= xF_X(x)|_0^b - \int_0^b F_X(x) dx \\ &= bF_X(b) - \int_0^b F_X(x) dx \\ &= bF_X(b) + b - b - \int_0^b F_X(x) dx \\ &= -b(1 - F_X(b)) + b - \int_0^b F_X(x) dx \\ &= -b(1 - F_X(b)) + \int_0^b dx - \int_0^b F_X(x) dx \\ &= -b(1 - F_X(b)) + \int_0^b (1 - F_X(x)) dx. \end{aligned}$$

Luego pasando al límite y teniendo en cuenta el Teorema 7.7 se obtiene

$$\int_0^{+\infty} x dF_X = \int_0^{+\infty} (1 - F_X(x)) dx.$$

Análogamente se prueba

$$\int_{-\infty}^0 x dF_X = - \int_{-\infty}^0 F_X(x) dx.$$

De estas dos últimas igualdades se obtiene el teorema. \square

Propiedad 7.9 *Sean X e Y dos variables aleatorias tal que $P(X \leq Y) = 1$, y tal que sus esperanzas $E(X)$, $E(Y)$ existen. Entonces*

$$(i) F_X(t) \geq F_Y(t), \quad \forall t, y$$

(ii) $E(X) \leq E(Y)$.

Demostración.

(i) Consideremos el evento $U = \{\omega : X(\omega) \leq Y(\omega)\}$. Claramente $P(U) = 1$ y $P(U^c) = 0$. Podemos escribir

$$\{Y \leq t\} = (\{Y \leq t\} \cap U) \cup (\{Y \leq t\} \cap U^c). \quad (7.20)$$

y luego como $P(\{Y \leq t\} \cap U^c) \leq P(U^c) = 0$, resulta

$$P(\{Y \leq t\}) = P(\{Y \leq t\} \cap U) + P(\{Y \leq t\} \cap U^c) \quad (7.21)$$

$$= P(\{Y \leq t\} \cap U). \quad (7.22)$$

Si $\omega \in \{Y \leq t\} \cap U$ entonces $X(\omega) \leq Y(\omega) \leq t$ de manera que

$$\{Y \leq t\} \cap U \subset \{X \leq t\}.$$

Tomando probabilidades y teniendo en cuenta (7.21) se obtiene que

$$P(\{Y \leq t\}) = P(\{Y \leq t\} \cap U) \leq P(\{X \leq t\}),$$

o bien

$$F_Y(t) \leq F_X(t) \quad (7.23)$$

y por lo tanto (i) se cumple.

(ii) También se tiene

$$1 - F_X(t) \leq 1 - F_Y(t), \quad (7.24)$$

y usando el Teorema 7.8 resulta

$$E(X) = \int_0^{+\infty} (1 - F_X(t)) dt - \int_{-\infty}^0 F_X(t) dt,$$

$$E(Y) = \int_0^{+\infty} (1 - F_Y(t)) dt - \int_{-\infty}^0 F_Y(t) dt.$$

Luego la Propiedad 7.9 se deduce de (7.23) y (7.24). \square

Supongamos que $P(X = 0) = 1$. Por la Propiedad 7.6 es claro que $E(X) = 0$.

Ahora bien, del hecho de que $E(X) = 0$ no se deduce que $P(X = 0) = 1$. ¿Qué condición podemos agregar para que se cumpla? La propiedad 7.10 responde a esta pregunta.

Propiedad 7.10 $E(X) = 0$ y $P(X \geq 0) = 1$ implica que $P(X = 0) = 1$.

Demostración. Supongamos que esta propiedad no fuera cierta, luego tendríamos una variable aleatoria X tal que $E(X) = 0$, $P(X \geq 0) = 1$ y $P(X = 0) < 1$. Luego teniendo en cuenta que $P(X \geq 0) = 1$ obtenemos que $P(X > 0) = P(X \geq 0) - P(X = 0) = 1 - P(X = 0) = a > 0$.

Ahora consideremos los eventos $A_n = \{X > \frac{1}{n}\}$. La sucesión $\{A_n\}$ es monótona creciente ya que $A_n \subset A_{n+1}$ y además

$$\{X > 0\} = \bigcup_{n \in \mathbb{N}} A_n,$$

de manera que

$$\lim_{n \rightarrow \infty} P(A_n) = P(\{X > 0\}) = a > 0.$$

Por lo tanto existe un número natural n_0 tal que $P(A_{n_0}) > a/2$ y entonces

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x dF_X \\ &= \int_0^{+\infty} x dF_X \\ &= \int_0^{\frac{1}{n_0}} x dF_X + \int_{\frac{1}{n_0}}^{+\infty} x dF_X \\ &\geq \int_{\frac{1}{n_0}}^{+\infty} x dF_X \\ &\geq \frac{1}{n_0} \int_{\frac{1}{n_0}}^{+\infty} dF_X \\ &= \frac{1}{n_0} \left(1 - F_X \left(\frac{1}{n_0} \right) \right) \\ &= \frac{1}{n_0} P \left(X > \frac{1}{n_0} \right) = \frac{1}{n_0} \frac{a}{2} > 0. \end{aligned}$$

lo cual es un absurdo ya que contradice la hipótesis. \square

Observación. La igualdad $\int_{-\infty}^{+\infty} x dF_X = \int_0^{+\infty} x dF_X$ se justifica teniendo en cuenta que $P(X \geq 0) = 1$.

Sea X una variable aleatoria discreta, R_X su rango y p_X su densidad. Sabemos que

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

El siguiente teorema permite hallar la esperanza de una variable aleatoria Y que es función medible de otra variable aleatoria X sin necesidad de de hallar antes la función de probabilidad puntual de la variable Y .

Teorema 7.9 Consideremos \mathbf{X} un vector aleatorio discreto de dimensión k y sea $g : \mathbb{R}^k \rightarrow \mathbb{R}$ una función medible. Definamos $Y = g(\mathbf{X})$. Entonces

$$E(Y) = \sum_{\mathbf{x} \in R_{\mathbf{X}}} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}).$$

Demostración. Sea $y \in g(R_{\mathbf{X}}) = R_Y$ y definamos

$$A_y = \{\mathbf{x} \in R_{\mathbf{X}} : g(\mathbf{x}) = y\} = g^{-1}(\{y\}).$$

Es fácil ver que la familia de subconjuntos $\{A_y\}_{y \in R_Y}$ es una partición de $R_{\mathbf{X}}$, es decir $R_{\mathbf{X}} = \bigcup_{y \in R_Y} A_y$ y si $y \neq y'$ entonces $A_y \cap A_{y'} = \emptyset$.

Teniendo en cuenta que

$$p_Y(y) = P_{\mathbf{X}}(A_y) = \sum_{\mathbf{x} \in A_y} p_{\mathbf{X}}(\mathbf{x}),$$

y que para todo $\mathbf{x} \in A_y$ se tiene $g(\mathbf{x}) = y$, obtenemos

$$\begin{aligned} E(Y) &= \sum_{y \in R_Y} y p_Y(y) \\ &= \sum_{y \in R_Y} y \sum_{\mathbf{x} \in A_y} p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{y \in R_Y} \sum_{\mathbf{x} \in A_y} y p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{y \in R_Y} \sum_{\mathbf{x} \in A_y} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{\mathbf{x} \in R_{\mathbf{X}}} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}), \end{aligned}$$

y por lo tanto queda demostrado el Teorema. \square

Ahora pasamos al caso absolutamente continuo. Sea X una variable aleatoria absolutamente continua y f_X su función de densidad. Sabemos que

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

El siguiente teorema es el análogo al teorema anterior cuando \mathbf{X} es un vector absolutamente continuo.

Teorema 7.10 Sea \mathbf{X} un vector aleatorio absolutamente continuo de dimensión k , con densidad $f_{\mathbf{X}}$. Sea $g : \mathbb{R}^k \rightarrow \mathbb{R}$ una función medible que toma un conjunto a lo sumo numerable de valores y definamos $Y = g(\mathbf{X})$. Luego

$$E(Y) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_k. \quad (7.25)$$

Demostración. Como en el teorema anterior consideramos la partición

$$A_y = \{\mathbf{x} \in R_{\mathbf{X}} : g(\mathbf{x}) = y\} = g^{-1}(\{y\}).$$

En este caso $\mathbb{R}^k = \bigcup_{y \in R_Y} A_y$ y si $y \neq y'$ entonces $A_y \cap A_{y'} = \emptyset$. Además $p_Y(y) = P_{\mathbf{X}}(g^{-1}(\{y\})) = P_{\mathbf{X}}(A_y)$. Entonces usando que para $\mathbf{x} \in A_y$ se tiene $g(\mathbf{x}) = y$, que además

$$\sum_{y \in R_Y} I_{A_y}(\mathbf{x}) = 1$$

y que

$$P_{\mathbf{X}}(A_y) = \int \cdots \int_{A_y} f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_k \quad (7.26)$$

obtenemos

$$\begin{aligned} E(Y) &= \sum_{y \in R_Y} y p_Y(y) \\ &= \sum_{y \in R_Y} y P_{\mathbf{X}}(A_y) \\ &= \sum_{y \in R_Y} y \int \cdots \int_{A_y} f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_k \\ &= \sum_{y \in R_Y} \int \cdots \int_{A_y} y f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_k \\ &= \sum_{y \in R_Y} \int \cdots \int_{A_y} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_k \\ &= \sum_{y \in R_Y} \int \cdots \int_{\mathbb{R}^k} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) I_{A_y}(\mathbf{x}) dx_1 \dots dx_k \\ &= \int \cdots \int_{\mathbb{R}^k} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \left(\sum_{y \in R_Y} I_{A_y}(\mathbf{x}) \right) dx_1 \dots dx_k = \\ &= \int \cdots \int_{\mathbb{R}^k} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_k. \quad \square \end{aligned}$$

Observación. En la demostración usamos (7.26). Como se comenta en la observación que sigue al Teorema 5.5, para demostrar esta propiedad para todo boreliano se requiere teoría de la medida y se debe usar la integral de Lebesgue.

Propiedad 7.11 *Sea X una variable aleatoria con esperanza finita. Entonces $E(X + c) = E(X) + c$.*

Demostración. Sea $Y = X + c$. Supongamos primero que $c > 0$. Sabemos que $F_Y(x) = F_X(x - c)$. Utilizando el Teorema 7.8 tenemos

$$\begin{aligned} E(Y) &= \int_0^{\infty} (1 - F_Y(y)) dy - \int_{-\infty}^0 F_Y(y) dy \\ &= \int_0^{\infty} (1 - F_X(y - c)) dy - \int_{-\infty}^0 F_X(y - c) dy. \end{aligned}$$

Haciendo el cambio de variable $x = y - c$ dentro de las integrales, resulta

$$\begin{aligned} E(Y) &= \int_{-c}^{\infty} (1 - F_X(x)) dx - \int_{-\infty}^{-c} F_X(x) dx \\ &= \int_{-c}^0 (1 - F_X(x)) dx + \int_0^{\infty} (1 - F_X(x)) dx - \int_{-\infty}^0 F_X(x) dx + \int_{-c}^0 F_X(x) dx \\ &= E(X) + \int_{-c}^0 (1 - F_X(x)) dx + \int_{-c}^0 F_X(x) dx \\ &= E(X) + \int_{-c}^0 dx - \int_{-c}^0 F_X(x) dx + \int_{-c}^0 F_X(x) dx \\ &= E(X) + \int_{-c}^0 dx \\ &= E(X) + x \Big|_{-c}^0 \\ &= E(X) + c. \end{aligned}$$

El caso de $c < 0$ se demuestra de la misma manera. \square

Recordemos el concepto de convergencia uniforme.

Definición 7.5 Sea $(f_n)_{n \geq 1}$ una sucesión de funciones definidas sobre A un conjunto cualquiera. Se dice que la sucesión de funciones $(f_n)_{n \geq 1}$ converge uniformemente a la función f sobre A para cada $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces para todo $x \in A$

$$|f_n(x) - f(x)| < \varepsilon.$$

Observación. La diferencia con la convergencia puntual es que el n_0 en este caso sirve para todo x , es decir sólo depende de ε .

La convergencia uniforme implica la puntual pero no al revés. En particular nos interesa la convergencia uniforme de variables aleatorias. Hacemos notar que el límite puntual de funciones medibles, y en consecuencia el límite uniforme, también resulta ser una función medible.

Teorema 7.11 *Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias definidas en (Ω, \mathcal{A}, P) que convergen uniformemente a una variable aleatoria X sobre Ω . Supongamos que $E(X)$ existe. Entonces*

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

Observación. La existencia de $E(X)$ implica la existencia de $E(X_n)$ para todo n a partir de un valor n_0 . Se deja como ejercicio.

Demostración. Sea (Ω, \mathcal{A}, P) el espacio de probabilidades donde están definidas las variables aleatorias X_n , $n \geq 1$ y X . Teniendo en cuenta la convergencia uniforme dado $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces

$$\sup_{\omega \in \Omega} |X_n(\omega) - X(\omega)| < \varepsilon.$$

Esto significa que si $n \geq n_0$ entonces

$$|X_n(\omega) - X(\omega)| < \varepsilon, \quad \forall \omega \in \Omega,$$

o bien

$$X(\omega) - \varepsilon < X_n(\omega) < X(\omega) + \varepsilon, \quad \forall \omega \in \Omega.$$

Por las propiedades 7.9 y 7.11 se obtiene que si $n \geq n_0$ entonces

$$E(X) - \varepsilon \leq E(X_n) \leq E(X) + \varepsilon.$$

Por lo tanto $\lim E(X_n) = E(X)$. \square

El siguiente teorema muestra que cualquier función medible puede aproximarse por otra que toma un conjunto a lo sumo numerable de valores.

Teorema 7.12 (i) *Sea $g: \mathbb{R}^k \rightarrow \mathbb{R}$ una función tal que $g(\mathbb{R}^k)$ es un conjunto finito o numerable. Luego una condición necesaria y suficiente para que g sea medible es que para todo $y \in g(\mathbb{R}^k) = R_g$, se tenga que $g^{-1}(y)$ pertenezca a \mathcal{B}^k .*

(ii) *Dada una función $g: \mathbb{R}^k \rightarrow \mathbb{R}$ medible, existe una sucesión $g_n: \mathbb{R}^k \rightarrow \mathbb{R}$ de funciones medibles tales que R_{g_n} es numerable, y $|g_n(\omega) - g(\omega)| \leq \frac{1}{n}$ para todo $\omega \in \Omega$. Luego g_n converge a g uniformemente.*

(iii) Sea \mathbf{X} un vector aleatorio de dimensión k y sea $Y = g(\mathbf{X})$ donde $g : \mathbb{R}^k \rightarrow \mathbb{R}$ es una función medible. Entonces si $g_n : \mathbb{R}^k \rightarrow \mathbb{R}$ es una sucesión de funciones medibles que converge uniformemente a g , resulta que $Y_n = g_n(\mathbf{X})$ converge uniformemente a Y .

(iv) Dada una variable aleatoria X existe una sucesión de variables aleatorias discretas X_n , $n \geq 1$ que converge uniformemente a X .

Demostración.

(i) Sea $y \in R_g$. Como $\{y\} \in \mathcal{B}$, $y \in R_g$, para que g sea medible es necesario que $g^{-1}(y) \in \mathcal{B}^k$. Supongamos ahora que esta condición se cumpla. Entonces

$$\begin{aligned} g^{-1}((-\infty, x]) &= g^{-1}((-\infty, x] \cap R_g) \\ &= \bigcup_{y \in (-\infty, x] \cap R_g} g^{-1}(y). \end{aligned}$$

como $(-\infty, x] \cap R_g$ es numerable y $g^{-1}(y) \in \mathcal{B}^k$, resulta $g^{-1}((-\infty, x]) \in \mathcal{B}^k$ y por lo tanto g es medible.

(ii) Dado n , todo $y \in \mathbb{R}$ pertenece a un intervalo de la forma $(i/n, (i+1)/n)$ para algún i entero. Luego definimos g_n por

$$g_n(x) = \frac{(i+1)}{n} \text{ si } g(x) \in (i/n, (i+1)/n].$$

Luego $|g_n(x) - g(x)| \leq 1/n$ y R_{g_n} es numerable. Por otro lado

$$g_n^{-1}\left(\frac{i+1}{n}\right) = g^{-1}\left(\left[\frac{i}{n}, \frac{i+1}{n}\right]\right)$$

pertenece a \mathcal{B}^k ya que g es medible. Por lo tanto por (i) g_n es medible.

(iii) Se deja como ejercicio.

(iv) Por (ii) podemos encontrar una sucesión de funciones medibles $g_n : \mathbb{R} \rightarrow \mathbb{R}$ tales que g_n converja uniformemente a la función identidad $g(x) = x$ y tal que además tomen un conjunto a lo sumo numerable de valores. Luego las variables $X_n = g_n(X)$ son discretas y por (iii) $X_n = g_n(X)$ converge uniformemente a $g(X) = X$. \square

El siguiente teorema generaliza el Teorema 7.10 para una función g medible cualquiera. La estrategia de la demostración es la siguiente y será usada a menudo: se aproxima uniformemente a la función g por una sucesión de funciones g_n que toman un número a lo sumo numerable de valores y que satisfacen la propiedad pedida. Luego usando que el Teorema 7.12 vale para las funciones g_n y pasando al límite se demuestra que la propiedad vale para g .

Teorema 7.13 Sea $\mathbf{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio absolutamente continuo con función de densidad $f_{\mathbf{X}}$ y $g : \mathbb{R}^k \rightarrow \mathbb{R}$ una función medible arbitraria. Si definimos la variable aleatoria $Y = g(\mathbf{X})$ entonces

$$E(Y) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

Demostración. Por el Teorema 7.12 (ii) existe una sucesión de funciones medibles g_n tal que R_{g_n} es a lo sumo numerable y que converge uniformemente a g . Definimos las variables aleatorias $Y_n = g_n(\mathbf{X})$. Por el Teorema 7.12 (iii), $(Y_n)_n$ converge uniformemente a Y .

Como ya hemos demostrado en el Teorema 7.10 que esta propiedad vale para funciones que toman un conjunto a lo sumo numerable de valores, se tendrá

$$E(Y_n) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g_n(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

Además por el Teorema 7.11 se tiene que $\lim_{n \rightarrow \infty} E(Y_n) = E(Y)$. Luego bastará probar que

$$\lim_{n \rightarrow +\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g_n(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (7.27)$$

Para probar esto observemos que

$$\begin{aligned} & \left| \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g_n(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} - \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right| \\ &= \left| \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (g_n(\mathbf{x}) - g(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right| \\ &\leq \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} |(g_n(\mathbf{x}) - g(\mathbf{x}))| f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &\leq \frac{1}{n} \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}}_{=1} = \frac{1}{n}, \end{aligned}$$

y por lo tanto se cumple (7.27). \square

Ahora vamos a probar la linealidad de la esperanza.

Teorema 7.14 Sean X_1 y X_2 dos variables aleatorias con esperanza finita. Entonces para todo escalar α y β vale que

$$E(\alpha X_1 + \beta X_2) = \alpha E(X_1) + \beta E(X_2).$$

Demostración.

Primero probaremos el Teorema cuando X_1 y X_2 son discretas. Sean X_1 y X_2 variables aleatorias discretas con esperanza finita y sea $Z = \alpha X_1 + \beta X_2$.

Definamos $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ por

$$g(x_1, x_2) = \alpha x_1 + \beta x_2.$$

Entonces si $\mathbf{X} = (X_1, X_2)$ se tiene que $Z = g(\mathbf{X})$. Definamos $g_i : \mathbb{R}^2 \rightarrow \mathbb{R}$, $i = 1, 2$ por $g_i(x_1, x_2) = x_i$. Luego $g(\mathbf{x}) = \alpha g_1(\mathbf{x}) + \beta g_2(\mathbf{x})$. Usando el Teorema 7.9 podemos escribir

$$\begin{aligned} E(Z) &= \sum_{(x_1, x_2) \in R_{\mathbf{X}}} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{(x_1, x_2) \in R_{\mathbf{X}}} [\alpha g_1(\mathbf{x}) + \beta g_2(\mathbf{x})] p_{\mathbf{X}}(\mathbf{x}) \\ &= \alpha \sum_{(x_1, x_2) \in R_{\mathbf{X}}} g_1(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) + \beta \sum_{(x_1, x_2) \in R_{\mathbf{X}}} g_2(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) \\ &= \alpha E(g_1(\mathbf{X})) + \beta E(g_2(\mathbf{X})) \\ &= \alpha E(X_1) + \beta E(X_2). \end{aligned}$$

Ahora bien, si X_1 y X_2 son variables aleatorias arbitrarias, entonces por Teorema 7.12 (iii) podemos definir dos sucesiones de variables aleatorias discretas $(X_{1n})_{n \geq 1}$ e $(X_{2n})_{n \geq 1}$ tales que convergen uniformemente a X_1 y X_2 respectivamente. Es fácil ver que también se tendrá que $\alpha X_{1n} + \beta X_{2n}$ converge uniformemente a $\alpha X_1 + \beta X_2$.

Hemos demostrado que para el caso de variables aleatorias discretas se cumple la linealidad de la esperanza. Luego tenemos

$$E(\alpha X_{1n} + \beta X_{2n}) = \alpha E(X_{1n}) + \beta E(X_{2n}). \quad (7.28)$$

Aplicando el Teorema 7.11 se obtiene

$$\lim_{n \rightarrow \infty} E(\alpha X_{1n} + \beta X_{2n}) = E(\alpha X_1 + \beta X_2), \quad (7.29)$$

y

$$\lim_{n \rightarrow \infty} E(X_{jn}) = E(X_j), \quad j = 1, 2. \quad (7.30)$$

Luego por (7.28), (7.29) y (7.30) se obtiene

$$\begin{aligned} E(\alpha X_1 + \beta X_2) &= \lim_{n \rightarrow \infty} E(\alpha X_{1n} + \beta X_{2n}) \\ &= \lim_{n \rightarrow \infty} (\alpha E(\alpha X_{1n}) + \beta E(X_{2n})) \\ &= \alpha \lim_{n \rightarrow \infty} E(X_{1n}) + \beta \lim_{n \rightarrow \infty} E(X_{2n}) \\ &= \alpha E(X_1) + \beta E(X_2), \end{aligned}$$

y esto prueba el teorema. \square

7.3. Esperanza del producto de variables aleatorias independientes.

Otro problema interesante es estudiar la esperanza de un producto de variables aleatorias. Si las variables aleatorias X e Y tienen esperanzas finitas y definimos la variable aleatoria $Z = XY$ entonces nos podemos preguntar: ¿cuándo vale que $E(Z) = E(XY) = E(X)E(Y)$? Veremos en el siguiente Teorema que una condición suficiente es la independencia de las variables X e Y .

Teorema 7.15 Sean X e Y variables aleatorias independientes con esperanza finita. Si $Z = XY$ entonces

$$E(Z) = E(XY) = E(X)E(Y).$$

Demostración. En principio lo probaremos para el caso discreto. Luego aproximaremos a X e Y por variables discretas uniformemente y probaremos el teorema para el caso general pasando al límite.

Sean X e Y variables aleatorias discretas independientes con esperanza finita y definamos $g : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$g(x, y) = xy.$$

Entonces como $Z = g(X, Y)$, por el Teorema 7.9 resulta

$$\begin{aligned} E(Z) &= \sum_{(x,y) \in R_{(X,Y)}} g(x, y) p_{(X,Y)}(x, y) \\ &= \sum_{(x,y) \in R_X \times R_Y} xy p_{(X,Y)}(x, y) \\ &= \sum_{(x,y) \in R_X \times R_Y} (xp_X(x)) (yp_Y(y)) \\ &= \left(\sum_{x \in R_X} xp_X(x) \right) \left(\sum_{y \in R_Y} yp_Y(y) \right) \\ &= E(X)E(Y). \end{aligned}$$

Observemos que $R_{(X,Y)} \subset R_X \times R_Y$ pero para $(x, y) \in R_X \times R_Y - R_{(X,Y)}$ se tiene $p_{(X,Y)}(x, y) = 0$, lo que justifica la segunda igualdad. La tercera se justifica por el hecho de que dado que X e Y son independientes se tiene $p_{(X,Y)}(x, y) = p_X(x)p_Y(y)$.

Por el Teorema 7.12 (ii) existe una sucesión de funciones medibles $g_n : \mathbb{R} \rightarrow \mathbb{R}$ que toman un conjunto a lo sumo numerable de valores y que converge uniformemente a la función identidad $g(x) = x$. Consideremos

las sucesiones de variables aleatorias discretas $g_n(X) = X_n$ e $Y_n = g_n(Y)$. Dado que X e Y son independientes, se tiene que X_n e Y_n también lo son.

Luego, como ya hemos probado que el teorema vale para el caso discreto, se tiene

$$E(X_n Y_n) = E(X_n) E(Y_n).$$

Ahora como por el Teorema 7.12 (iii) X_n converge uniformemente a X e Y_n converge uniformemente a Y se tendrá

$$\lim_{n \rightarrow \infty} E(X_n Y_n) = \lim_{n \rightarrow \infty} E(X_n) \lim_{n \rightarrow \infty} E(Y_n) = E(X) E(Y).$$

Luego basta probar que $\lim_{n \rightarrow \infty} E(X_n Y_n) = E(XY)$. Para ver esto observemos que

$$\begin{aligned} |E(X_n Y_n) - E(XY)| &= |E(X_n Y_n - XY)| \\ &\leq E|X_n Y_n - XY| \\ &= E|X_n Y_n - X_n Y + X_n Y - XY| \\ &= E|X_n(Y_n - Y) + Y(X_n - X)| \\ &\leq E(|X_n(Y_n - Y)| + |Y(X_n - X)|) \\ &\leq E(|X_n||Y_n - Y|) + E(|Y||X_n - X|). \end{aligned} \quad (7.31)$$

Por la convergencia uniforme de X_n a X y de Y_n a Y tenemos

$$\lim_{n \rightarrow \infty} \max_{\omega \in \Omega} |X_n(\omega) - X(\omega)| = 0 \quad (7.32)$$

y

$$\lim_{n \rightarrow \infty} \max_{\omega \in \Omega} |Y_n(\omega) - Y(\omega)| = 0. \quad (7.33)$$

Además como $|X_n| \rightarrow |X|$ uniformemente, resulta por el Teorema 7.11

$$\lim_{n \rightarrow \infty} E(|X_n|) = E(|X|). \quad (7.34)$$

De (7.31), (7.32), (7.33) y (7.34) se obtiene que

$$\lim_{n \rightarrow \infty} |E(X_n Y_n) - E(XY)| = 0,$$

y esto prueba el teorema. \square

Damos a continuación un ejemplo que muestra que la recíproca es falsa, es decir es falso que $E(XY) = E(X)E(Y)$ implique que X e Y son independientes.

Ejemplo 7.1 *Consideremos un vector (X, Y) discreto tal que*

$$R_{(X,Y)} = \{(-1, 0), (1, 0), (0, -1), (0, 1)\}$$

y tal que $p(x, y) = 1/4$ para cada $(x, y) \in R_{(X,Y)}$.

Como para todo $(x, y) \in R_{(X,Y)}$, se tiene $xy = 0$, resulta $P(XY \equiv 0) = 1$. Luego $E(XY) = 0$. También se ve que $R_X = \{-1, 0, 1\}$ y $p_X(-1) = 1/4$, $p_X(0) = 1/2$ y $p_X(1) = 1/4$, por lo tanto resulta

$$E(X) = -1(1/4) + 0(1/2) + 1(1/4) = 0.$$

De manera que se cumple que

$$E(XY) = E(X)E(Y) = 0.$$

Pero X e Y no son independientes pues $p_X(1) = \frac{1}{4} = p_Y(1)$ y dado que $(1, 1) \notin R_{(X,Y)}$ se tiene $p_{(X,Y)}(1, 1) = 0$.

Sin embargo si X, Y fueran independientes debiera cumplirse

$$p_{(X,Y)}(1, 1) = p_X(1)p_Y(1) = \frac{1}{4} \frac{1}{4} = \frac{1}{16}.$$

lo cual es una contradicción. Por lo tanto X e Y no son independientes.

7.4. Una fórmula general para la esperanza de una variable transformada

Teorema 7.16 Sea X una variable aleatoria con esperanza finita y $g : \mathbb{R} \rightarrow \mathbb{R}$ tal que $g(X)$ tiene esperanza finita. Supongamos además que existen un número finito de puntos $-\infty = d_0 < d_1 < \dots < d_k = \infty$, tales que en $D_i = (d_i, d_{i+1}]$ la función g es continua y estrictamente creciente o estrictamente decreciente o constante y que $\lim_{x \downarrow d_i} g(x)$ existe. Supongamos además que en d_i , $1 \leq i \leq k-1$ la función g es continua o F_X es continua. Luego se tiene

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

Demostración. Podemos escribir

$$g(X) = \sum_{i=1}^k g(X) I_{D_i}(X).$$

Vamos a ver que para probar el teorema bastará mostrar que

$$E(g(X) I_{D_i}(X)) = \int_{d_i}^{d_{i+1}} g dF_X. \quad (7.35)$$

Es importante observar que de acuerdo a las observaciones 2 y 3 de la página 126 la integral de Riemann-Stieltjes en el lado derecho de (7.35) existe. En efecto, si (7.35) se cumple se tendrá por el Teorema 7.14 y el hecho de que en los puntos $d_i, 1 \leq i \leq k-1$ la función F_X o g es continua, que

$$\begin{aligned} E(g(X)) &= \sum_{i=1}^k E(g(X)I_{D_i}(X)) \\ &= \sum_{i=1}^k \int_{d_i}^{d_{i+1}} g dF_X \\ &= \int_{-\infty}^{\infty} g dF_X. \end{aligned}$$

Veamos que (7.35) para el caso que g es constante en D_i . En este caso sea c el valor de la función en D_i . Luego $g(X)I_{D_i}(X)$ toma valores c con probabilidad $F_X(d_{i+1}) - F_X(d_i)$ y 0 con probabilidad $1 - (F_X(d_{i+1}) - F_X(d_i))$. Luego

$$\begin{aligned} E(g(X)I_{D_i}(X)) &= c(F_X(d_{i+1}) - F_X(d_i)) \\ &= \int_{d_i}^{d_{i+1}} g dF_X, \end{aligned}$$

y por lo tanto (7.35) se cumple.

Veamos ahora que (7.35) vale en los intervalos D_i donde g es estrictamente creciente. Sean $a_i^* = \lim_{x \downarrow d_i} g(x)$ y $b_i^* = \lim_{x \uparrow d_{i+1}} g(x)$ donde $\lim_{x \downarrow a}$ indica límite cuando x tiende a a por la derecha y $\lim_{x \uparrow a}$ indica el límite cuando x tiende a a por la izquierda. Sea $Y_i = g(X)I_{D_i}(X)$. De acuerdo al Teorema 6.1

$$F_{Y_i}(y) = \begin{cases} 0 & \text{si } y \leq a_i^* \\ F_X(g_i^{-1}(y)) & \text{si } a_i^* < y < b_i^* \\ 1 & \text{si } y \geq b_i^*, \end{cases} \quad (7.36)$$

donde g_i es la restricción de g a D_i . Luego

$$E(Y_i) = \int_{a_i^*}^{b_i^*} y dF_{Y_i}.$$

Como $\lim_{a \downarrow a_i^*} g_i^{-1}(a) = d_i$ y $\lim_{b \uparrow b_i^*} g_i^{-1}(b) = d_{i+1}$, para probar (7.35) bastará demostrar que para todo $a_i^* < a < b < b_i^*$ se tiene

$$\int_a^b y dF_Y = \int_{g_i^{-1}(a)}^{g_i^{-1}(b)} g(x) dF_X. \quad (7.37)$$

En efecto si (7.37), vale entonces resulta

$$\begin{aligned}
E(Y_i) &= \int_{a_i^*}^{b_i^*} y dF_{Y_i} \\
&= \lim_{a \downarrow a_i^*, b \uparrow b_i^*} \int_a^b y dF_{Y_i} \\
&= \lim_{a \downarrow a_i^*, b \uparrow b_i^*} \int_{g_i^{-1}(a)}^{g_i^{-1}(b)} g(x) dF_X \\
&= \int_{d_i}^{d_{i+1}} g(x) dF_X.
\end{aligned}$$

y por lo tanto (7.35) vale.

Para mostrar (7.37) consideremos una sucesión de particiones π_n del intervalo $[a, b]$ en n intervalos de igual longitud. Entonces tenemos $\pi_Y^n = \{y_0^n, y_1^n, \dots, y_n^n\}$ con $a = y_0^n < y_1^n < \dots < y_n^n = b$ e $y_{j+1} - y_j = 1/n$, $1 \leq j \leq n$. Tomemos una selección arbitraria de puntos en esta partición $y_j^n < \eta_j^n \leq y_{j+1}^n$, la llamamos $\eta^n = (\eta_j^n)_{1 \leq j \leq n}$. Luego por 7.36 tenemos que

$$\begin{aligned}
S_a^b(\pi_Y^n, \eta^n, y, F_Y) &= \sum_{j=1}^n \eta_j^n (F_Y(y_{j+1}^n) - F_Y(y_j^n)) \\
&= \sum_{j=1}^n \eta_j^n (F_X(g_i^{-1}(y_{j+1}^n)) - F_X(g_i^{-1}(y_j^n))). \quad (7.38)
\end{aligned}$$

Entonces como la función $id(y) = y$ es continua en $[a, b]$ y F_Y es monótona, existe la integral de Riemann-Stieltjes $\int_a^b y dF_Y$ y se tiene que

$$\lim_{n \rightarrow \infty} S_a^b(\pi_Y^n, \eta^n, y, F_Y) = \int_a^b y dF_Y. \quad (7.39)$$

Llamemos ahora

$$x_j^n = g_i^{-1}(y_j^n), \quad 0 \leq j \leq n, \quad \xi_j^n = g_i^{-1}(\eta_j^n), \quad 1 \leq j \leq n.$$

Luego por la monotonía de g_i^{-1} obtenemos $g_i^{-1}(a) = x_0^n < x_1^n < \dots < x_n^n = g_i^{-1}(b)$ y $x_j^n < \xi_j^n \leq x_{j+1}^n$. Por lo tanto $\pi_X^n = \{x_0^n, x_1^n, \dots, x_n^n\}$ es una partición de $[g_i^{-1}(a), g_i^{-1}(b)]$ y $\xi^n = (\xi_j^n)_{1 \leq j \leq n}$ una selección en esta partición. Además

$$\begin{aligned}
\|\pi_X^n\| &= \max_{1 \leq j \leq n} (x_{j+1}^n - x_j^n) \\
&= \max_{1 \leq j \leq n} (g_i^{-1}(y_{j+1}^n) - g_i^{-1}(y_j^n))
\end{aligned}$$

tiende a 0 con n por la continuidad uniforme de g_i^{-1} en $[g_i^{-1}(a), g_i^{-1}(b)]$ y el hecho de que

$$\lim_{n \rightarrow \infty} \max_{1 \leq j \leq n} (y_{j+1}^n - y_j^n) = 0.$$

Luego, como g es continua en $[g_i^{-1}(a), g_i^{-1}(b)]$ y F_X es monótona, existe la integral de Riemann-Stieltjes $\int_{g_i^{-1}(a)}^{g_i^{-1}(b)} g(x) dF_X$ y resulta que

$$\lim_{n \rightarrow \infty} S_{g_i^{-1}(a)}^{g_i^{-1}(b)}(\pi_X^n, \xi^n, g, F_X) = \int_{g_i^{-1}(a)}^{g_i^{-1}(b)} g(x) dF_X. \quad (7.40)$$

Finalmente observemos de (7.38) que

$$\begin{aligned} S_a^b(\pi_Y^n, \eta^n, y, F_Y) &= \sum_{j=1}^n \eta_j^n (F_X(g_i^{-1}(y_{j+1}^n)) - F_X(g_i^{-1}(y_j^n))) \\ &= \sum_{j=1}^n g(g_i^{-1}(\eta_j^n)) (F_X(x_{j+1}) - F_X(x_j)) \\ &= \sum_{j=1}^n g(\xi_j^n) (F_X(x_{j+1}) - F_X(x_j)) \\ &= S_{g_i^{-1}(a)}^{g_i^{-1}(b)}(\pi_X^n, \xi^n, g, F_X). \end{aligned} \quad (7.41)$$

Luego de (7.39) (7.40) y (7.41) obtenemos (7.37), y por lo tanto (7.35) queda demostrada para el caso que g es estrictamente creciente en D_i .

Para el caso que g es estrictamente decreciente, tenemos que $-g$ es estrictamente creciente. Por lo tanto (7.35) vale para $-g$ y entonces

$$E(-g(X)I_{D_i}(X)) = \int_{d_i}^{d_{i+1}} -g dF_X.$$

Pero esto es equivalente a

$$E(g(X)I_{D_i}(X)) = \int_{d_i}^{d_{i+1}} g dF_X,$$

y luego (7.35) también vale. Esto prueba el teorema. \square

7.5. Esperanza de distribuciones simétricas

El concepto de esperanza matemática está ligado con el valor central de la distribución. Ciertas variables llamadas simétricas tienen un centro

natural. Por ejemplo aquellas que tienen densidad simétrica respecto a un punto.

Definición 7.6 *Dada una variable aleatoria X cualquiera, se dice que tiene distribución simétrica respecto de μ si*

$$P_X([\mu - x, \mu]) = P_X((\mu, \mu + x]). \quad (7.42)$$

para todo $x > 0$.

Teorema 7.17 *X tiene distribución simétrica respecto de 0 si y sólo si $F_X = F_{-X}$*

Demostración. X tiene distribución simétrica respecto de 0 si y sólo si

$$P_X([-x, 0)) = P_X((0, x]), \quad \forall x > 0. \quad (7.43)$$

Se tiene

$$P_X((0, x]) = F_X(x) - F_X(0) \quad (7.44)$$

y

$$\begin{aligned} P_X([-x, 0)) &= P(-x \leq X < 0) \\ &= P(x \geq -X > 0) \\ &= P(0 < -X \leq x) \\ &= F_{-X}(x) - F_{-X}(0). \end{aligned} \quad (7.45)$$

Luego, de (7.43), (7.44) y (7.45) resulta que X tiene distribución simétrica respecto de 0 si y sólo si

$$F_X(x) - F_X(0) = F_{-X}(x) - F_{-X}(0), \quad \forall x > 0. \quad (7.46)$$

Tomando límite cuando x tiende a infinito resulta

$$1 - F_X(0) = 1 - F_{-X}(0)$$

y luego

$$F_X(0) = F_{-X}(0). \quad (7.47)$$

De (7.46) y (7.47) resulta que si X tiene distribución simétrica respecto de 0 entonces

$$F_X(x) = F_{-X}(x), \quad \forall x. \quad (7.48)$$

Veamos la recíproca. Supongamos que

$$F_X(x) = F_{-X}(x), \quad \forall x.$$

Luego, para todo $x \in \mathbb{R}$ se tiene

$$P(X \leq x) = F_X(x) = F_{-X}(x) = P(-X \leq x) = P(X \geq -x).$$

En particular

$$P(X \leq 0) = P(X \geq 0).$$

Luego, si $x > 0$

$$\begin{aligned} P(0 < X \leq x) &= P(X \leq x) - P(X \leq 0) \\ &= P(X \geq -x) - P(X \geq 0) \\ &= P(-x \leq X < 0). \end{aligned}$$

Es decir, (7.48) implica que

$$P_X([-x, 0)) = P_X((0, x]), \quad \forall x > 0,$$

de lo que se deduce que X es simétrica. \square

Teorema 7.18 *X tiene distribución simétrica respecto de μ si y sólo si $Y = X - \mu$ tiene distribución simétrica respecto de 0.*

Demostración. Sea $x > 0$. Se tiene

$$\begin{aligned} P_X([\mu - x, \mu)) &= P(\mu - x \leq X < \mu) \\ &= P(-x \leq X - \mu \leq 0) \\ &= P(-x \leq Y \leq 0) \\ &= P_Y([-x, 0)), \end{aligned}$$

y

$$\begin{aligned} P_X((\mu, \mu + x]) &= P(\mu < X \leq \mu + x) \\ &= P(0 < X - \mu \leq x) \\ &= P(0 < Y \leq x) \\ &= P_Y((0, x]). \end{aligned}$$

Luego $P_X([\mu - x, \mu)) = P_X((\mu, \mu + x])$ es equivalente a $P_Y([-x, 0)) = P_Y((0, x])$ y por lo tanto el teorema es cierto. \square

Teorema 7.19 *Si X tiene esperanza finita y tiene distribución simétrica respecto de μ , entonces $E(X) = \mu$.*

Demostración. Primero probaremos el teorema cuando $\mu = 0$. En este caso por el Teorema 7.14

$$E(-X) = -E(X). \quad (7.49)$$

Ademas como $F_X = F_{-X}$, y la esperanza depende solamente de la función de distribución se tendrá

$$E(-X) = E(X). \quad (7.50)$$

De (7.49) y (7.50) resulta $E(X) = E(-X) = 0$.

Supongamos ahora que X tenga distribución simétrica respecto de μ . Entonces $X - \mu$ tiene distribución simétrica respecto de 0. Luego usando la Propiedad 7.11 resulta

$$0 = E(X - \mu) = E(X) - \mu,$$

y el teorema queda demostrado. \square

Teorema 7.20 (i) Si X es absolutamente continua, entonces X tiene distribución simétrica respecto de μ si y sólo si

$$f_X(\mu - x) = f_X(\mu + x). \quad (7.51)$$

(ii) Si X es discreta, entonces X tiene distribución simétrica respecto de μ si y sólo si

$$p_X(\mu - x) = p_X(\mu + x).$$

Demostración.

(i) Si llamamos $Y = X - \mu$, como $f_Y(x) = f_X(x + \mu)$, (7.51) es equivalente a

$$f_Y(-x) = f_Y(x).$$

Por otro lado por las fórmulas de cambio de variable

$$f_{-Y}(x) = f_Y(-x).$$

Luego (7.51) es equivalente a $f_{-Y} = f_Y$ y esto es equivalente a $F_{-Y} = F_Y$. Aplicando el Teorema 7.17 esto es equivalente a que Y sea simétrica respecto de 0 y por Teorema 7.18 a que X sea simétrica respecto de μ .

(ii) Es similar a (i). Se deja como ejercicio. \square

7.6. Mediana de una variable aleatoria.

Dijimos que la esperanza describe un valor central de una variable aleatoria. En particular, si la variable aleatoria X es simétrica y tiene esperanza finita, entonces esta coincide con su centro de simetría. Una desventaja de la esperanza es que es muy inestable, es decir es muy sensible a las pequeñas perturbaciones, pequeños cambios en la distribución de la variable se ven reflejados en importantes cambios en los valores de la esperanza.

Otra desventaja de la esperanza es que puede ocurrir que no exista. Incluso esto puede darse en el caso de una distribución simétrica. Un ejemplo de distribución simétrica que no tiene esperanza es la distribución de Cauchy. Su densidad está dada por

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

Es fácil ver que efectivamente es una densidad. Tenemos que

$$\begin{aligned} \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{1+x^2} &= \frac{2}{\pi} \int_0^{\infty} \frac{1}{1+x^2} \\ &= \frac{2}{\pi} \operatorname{arctg}(x) \Big|_0^{\infty} \\ &= \frac{2}{\pi} \left(\frac{\pi}{2} - 0 \right) \\ &= 1 \end{aligned}$$

El gráfico de esta densidad es parecido al de la densidad normal aunque las colas tienden a 0 más lentamente. Es una función par y por lo tanto simétrica respecto del eje y . Esta distribución no tiene esperanza puesto que un cálculo sencillo prueba que

$$\frac{1}{\pi} \int_0^{+\infty} x \frac{1}{1+x^2} dx = -\frac{1}{\pi} \int_{-\infty}^0 x \frac{1}{1+x^2} dx = +\infty.$$

En efecto haciendo la transformación $y = 1+x^2$ en la primer integral se tiene $dy = 2x dx$ y entonces

$$\begin{aligned} \frac{1}{\pi} \int_0^{+\infty} x \frac{1}{1+x^2} dx &= \frac{1}{2\pi} \int_1^{+\infty} \frac{1}{y} dy \\ &= \frac{1}{2\pi} \log(y) \Big|_1^{\infty} = \infty. \end{aligned}$$

Por lo tanto la simetría no garantiza la existencia de la esperanza. En este sentido no es una buena medida de centralidad, puesto que cualquier medida de centralidad debiera coincidir con el centro de simetría de f_X en el caso de existir éste.

Otra medida de centralidad es la mediana. Si existe un valor que deja la misma probabilidad a su derecha que a la izquierda, ese valor es la mediana. Esto se podrá lograr siempre en el caso de una variable aleatoria continua. Si X es simétrica entonces la mediana coincide con el centro de simetría. Una definición general de mediana es la siguiente.

Definición 7.7 *Se dice que m es una mediana de la variable aleatoria X si se cumple que*

$$(i) P(X \geq m) \geq \frac{1}{2}, y$$

$$(ii) P(X \leq m) \geq \frac{1}{2}.$$

Veremos que siempre existe, y que si no es única, el conjunto de las medianas es conexo, es decir es un intervalo en \mathbb{R} . Para mostrar esto necesitaremos recurrir a la función

$$F_X^{-1}(y) = \inf A_y,$$

donde $A_y = \{x : F_X(x) \geq y\}$. Hemos visto que el ínfimo es en verdad un mínimo, de manera que $F_X(F_X^{-1}(y)) \geq y$ es decir

$$P(X \leq F_X^{-1}(y)) \geq y. \quad (7.52)$$

Probaremos ahora una propiedad adicional.

Teorema 7.21

$$P(X \geq F_X^{-1}(y)) \geq 1 - y. \quad (7.53)$$

Demostración. Sea $x < F_X^{-1}(y)$, entonces, dado que $F_X^{-1}(y)$ es el mínimo de A_y se tiene que $F_X(x) < y$. Luego si ponemos $x = F_X^{-1}(y) - \frac{1}{n} < F_X^{-1}(y)$ obtenemos

$$F_X\left(F_X^{-1}(y) - \frac{1}{n}\right) < y,$$

es decir

$$P\left(X \leq F_X^{-1}(y) - \frac{1}{n}\right) < y.$$

La sucesión de eventos

$$A_n = \{X \leq F_X^{-1}(y) - \frac{1}{n}\}$$

es monótona no decreciente y además

$$\bigcup_{n=1}^{\infty} A_n = \{X < F_X^{-1}(y)\}.$$

Luego pasando al límite se tiene

$$\lim_{n \rightarrow \infty} P\left(X \leq F_X^{-1}(y) - \frac{1}{n}\right) \leq y,$$

y además

$$\lim_{n \rightarrow \infty} P\left(X \leq F_X^{-1}(y) - \frac{1}{n}\right) = P(\{X < F_X^{-1}(y)\}).$$

Por lo tanto

$$P(\{X < F_X^{-1}(y)\}) \leq y,$$

o equivalentemente

$$P(\{X \geq F_X^{-1}(y)\}) \geq 1 - y. \quad \square$$

Teorema 7.22 *Sea X una variable aleatoria y F_X su distribución. Entonces*

(i) $F_X^{-1}\left(\frac{1}{2}\right)$ es una mediana.

(ii) Si m es mediana de X entonces

$$F_X^{-1}\left(\frac{1}{2}\right) \leq m.$$

(iii) Si m_1 y m_2 son medianas de X entonces para todo $m \in (m_1, m_2)$, m es mediana de X .

Demostración.

(i) Se deduce de (7.52) y (7.53) tomando $y = \frac{1}{2}$.

(ii) Si m es otra mediana, entonces como $P(X \leq m) \geq \frac{1}{2}$, resulta que $m \in A_{\frac{1}{2}}$. Como $F_X^{-1}\left(\frac{1}{2}\right) = \inf A_{\frac{1}{2}}$ resulta $F_X^{-1}\left(\frac{1}{2}\right) \leq m$.

(iii) Se deja como ejercicio. \square

También se propone como ejercicio dar ejemplos de distribuciones en las que el intervalo de las medianas sea cerrado a derecha y ejemplos en los que sea abierto a derecha.

En el caso de que se trate de un intervalo podemos definir la mediana central como el punto medio del intervalo. Es decir si el conjunto de medianas es el intervalo $[a, b)$ o el $[a, b]$, la mediana central es $m_c(X) = \frac{a+b}{2}$.

7.7. Varianza de una variable aleatoria.

La esperanza y la mediana de una variable aleatoria son características de su distribución que describen un valor central. Sin embargo, variables aleatorias con distribuciones muy distintas pueden tener la misma esperanza. Por ejemplo pueden diferir en cuan dispersos alrededor de la esperanza están los valores que toma la variable. Variables con la misma esperanza pueden estar más o menos dispersas. Esto nos lleva a definir otras características de una variable aleatoria, que midan la dispersión alrededor de un valor central.

Tampoco existe una única manera de medir dicha “dispersión”. Consideremos una variable aleatoria X . Podríamos considerar la distancia entre los valores que toma X y su esperanza, es decir $|X - E(X)|$ y como esto resulta ser una variable aleatoria, calcular su esperanza $E(|X - E(X)|)$. Sin embargo, dado que la función valor absoluto no es derivable en el origen, será conveniente reemplazarla por la función cuadrática.

Definición 7.8 *Definimos la varianza de la variable aleatoria X por*

$$\text{Var}(X) = E((X - E(X))^2).$$

Se la suele notar por σ_X^2 . La desviación típica o desvío estándar de una variable aleatoria X es definida como la raíz cuadrada de la varianza

$$ds(X) = \sqrt{\text{Var}(X)} = \sigma_X.$$

Observación. Es Inmediato observar que $\text{Var}(X) \geq 0$ pues se trata de la esperanza de una variable aleatoria no negativa. También es claro que siempre existe si admitimos como medida el valor $+\infty$.

La varianza tiene las siguientes propiedades.

Propiedad 7.12 *Si X tiene varianza finita, entonces*

$$\text{Var}(X) = E(X^2) - E^2(X).$$

Luego para el caso discreto resulta

$$\text{Var}(X) = \sum_{x \in R_X} x^2 p_X(x) - \left(\sum_{x \in R_X} x p_X(x) \right)^2,$$

y para el continuo

$$\text{Var}(X) = \int_{-\infty}^{\infty} x^2 f_X(x) dx - \left(\int_{-\infty}^{\infty} x f_X(x) dx \right)^2.$$

Demostración. Teniendo en cuenta las propiedades de la esperanza, se obtiene que:

$$\begin{aligned}
 \text{Var}(X) &= E((X - E(X))^2) \\
 &= E(X^2 - 2E(X)X + E^2(X)) \\
 &= E(X^2) - 2E(X)E(X) + E(E^2(X)) \\
 &= E(X^2) - 2E^2(X) + E^2(X) \\
 &= E(X^2) - E^2(X). \square
 \end{aligned}$$

Propiedad 7.13 $\text{Var}(X) = 0$ es equivalente a $P(X = E(X)) = 1$.

Demostración. Supongamos que $\text{Var}(X) = E((X - E(X))^2) = 0$. Como $(X - E(X))^2$ es no negativa, resulta por la Propiedad 7.10 que

$$P((X - E(X))^2 = 0) = 1.$$

Esto equivale a que

$$P(X - E(X) = 0) = 1,$$

o

$$P(X = E(X)) = 1.$$

Se deja como ejercicio probar que si

$$P(X = E(X)) = 1,$$

entonces $\text{Var}(X) = 0$. Para eso obsérvese que la variable aleatoria $(X - E(X))^2$ es cero con probabilidad uno. \square

Propiedad 7.14 Sea X una variable aleatoria e $Y = \alpha X + \beta$, con α, β escalares. Entonces $\text{Var}(Y) = \alpha^2 \text{Var}(X)$.

Demostración. Como $E(Y) = \alpha E(X) + \beta$ resulta

$$\begin{aligned}
 \text{Var}(Y) &= E((Y - E(Y))^2) \\
 &= E([\alpha X + \beta - (\alpha E(X) + \beta)]^2) \\
 &= E([\alpha(X - E(X))]^2) \\
 &= \alpha^2 E([X - E(X)]^2) \\
 &= \alpha^2 \text{Var}(X). \square
 \end{aligned}$$

Se mostrará que en el caso de suma de variables aleatorias independientes, la varianza es aditiva.

Propiedad 7.15 Sean X e Y variables aleatorias independientes. Luego si $Z = X + Y$ resulta $\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y)$.

Demostración. Tenemos

$$\begin{aligned}\text{Var}(Z) &= E\left([Z - E(Z)]^2\right) \\ &= E\left([X + Y - E(X) - E(Y)]^2\right) \\ &= E\left([(X - E(X)) + (Y - E(Y))]^2\right) \\ &= E\left([X - E(X)]^2\right) + 2E\left([X - E(X)][Y - E(Y)]\right) + E\left([Y - E(Y)]^2\right) \\ &= \text{Var}(X) + 2E\left([X - E(X)][Y - E(Y)]\right) + \text{Var}(Y).\end{aligned}$$

Luego, bastará probar que

$$E\left([X - E(X)][Y - E(Y)]\right) = 0.$$

Usando la independencia de X e Y y teniendo en cuenta que

$$E(X - E(X)) = 0 = E(Y - E(Y)),$$

resulta

$$\begin{aligned}E\left([X - E(X)][Y - E(Y)]\right) &= E(X - E(X))E(Y - E(Y)) \\ &= 0. \quad \square\end{aligned}\tag{7.54}$$

7.7.1. Esperanzas y varianzas de distribuciones normales

Calcularemos ahora $E(Y)$ y $\text{Var}(Y)$ para una variable Y con distribución $N(\mu, \sigma^2)$.

Teorema 7.23 Si $Y \sim N(\mu, \sigma^2)$ entonces $E(Y) = \mu$ y $\text{Var}(Y) = \sigma^2$.

Demostración. Tomemos primero una variable X con distribución $N(0, 1)$. Mostraremos que $E(X) = 0$ y $\text{Var}(X) = 1$. La densidad de X es

$$f(x) = \frac{1}{(2\pi)^{1/2}}e^{-x^2/2}.$$

Como X es simétrica respecto de 0, para mostrar que $E(X) = 0$, bastara mostrar que $E(|X|) < \infty$. Tenemos que

$$\begin{aligned}E(|X|) &= \int_{-\infty}^{\infty} |x|f(x)dx \\ &= 2 \int_0^{\infty} xf(x)dx \\ &= \frac{2}{(2\pi)^{1/2}} \int_0^{\infty} xe^{-x^2/2}dx.\end{aligned}\tag{7.55}$$

Definamos $u = x^2/2$ y entonces $du = xdx$. Luego

$$\begin{aligned}
 E(|X|) &= \frac{2}{(2\pi)^{1/2}} \int_0^\infty xe^{-x^2/2} dx \\
 &= \frac{2}{(2\pi)^{1/2}} \int_0^\infty e^{-u} du \\
 &= \frac{2}{(2\pi)^{1/2}} (-e^{-u}|_0^\infty) \\
 &= \frac{2}{(2\pi)^{1/2}} < \infty.
 \end{aligned}
 \tag{7.56}$$

Vamos ahora a calcular la integral indefinida

$$\int x^2 e^{-x^2/2} dx.$$

Haciendo $u = x$ y $dv = xe^{-x^2/2} dx$ para integrar por partes, se tiene $du = dx$ y por (7.56) $v = -e^{-x^2/2}$. Luego

$$\begin{aligned}
 \int x^2 e^{-x^2/2} dx &= \int u dv \\
 &= uv - \int v du \\
 &= -xe^{-x^2/2} + \int e^{-x^2/2} dx.
 \end{aligned}$$

Luego

$$\int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = -[xe^{-x^2/2}]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-x^2/2} dx,$$

y como $[xe^{-x^2/2}]_{-\infty}^{\infty} = 0$, resulta

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

Entonces se tiene

$$\begin{aligned}
 \text{Var}(X) &= \int_{-\infty}^{\infty} x^2 f(x) dx \\
 &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \\
 &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \\
 &= \int_{-\infty}^{\infty} f(x) dx \\
 &= 1.
 \end{aligned}$$

De acuerdo a su definición, la distribución $N(\mu, \sigma^2)$ es la distribución de $Y = \sigma X + \mu$, con $X \sim N(\mu, \sigma^2)$. Luego $E(Y) = \sigma E(X) + \mu = \mu$ y $\text{Var}(Y) = \sigma^2 \text{Var}(X) = \sigma^2$. \square

Observación. De acuerdo a este resultado, los parámetros de una distribución normal coinciden con la esperanza y la varianza.

7.8. Covarianza

La ecuación (7.54) motiva la definición del concepto de covarianza.

Definición 7.9 Sean X e Y variables aleatorias. Se define la covarianza de X e Y como

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

Las siguientes Propiedades 7.16 y 7.17 son inmediatas

Propiedad 7.16 $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.

Propiedad 7.17 Si X, Y son independientes, $\text{Cov}(X, Y) = 0$

La recíproca es falsa: la covarianza igual a cero no garantiza la independencia de las variables. Se puede dar el mismo contraejemplo que se usó luego del Teorema 7.15 para mostrar que $E(XY) = E(X)E(Y)$ no implica que X e Y sean independientes.

Diremos que dos variables aleatorias X e Y están positivamente correlacionadas si $\text{Cov}(X, Y) > 0$ y negativamente correlacionadas si $\text{Cov}(X, Y) < 0$.

Si $\text{Cov}(X, Y) = E([X - EX][Y - E(Y)]) > 0$, $X - EX$ y $Y - E(Y)$ tienden a tener el mismo signo, es decir tienden a situarse del mismo lado de sus respectivas esperanzas. Lo contrario ocurre si $\text{Cov}(X, Y) < 0$.

Propiedad 7.18 Si X e Y son variables aleatorias y ponemos $X' = \alpha X + \beta$ e $Y' = \gamma Y + \delta$ entonces

$$\text{Cov}(X', Y') = \alpha\gamma \text{Cov}(X, Y).$$

Demostración. Para probarlo obsérvese que

$$\begin{aligned} X' - E(X') &= \alpha X + \beta - (\alpha E(X) + \beta) = \alpha(X - E(X)), \\ Y' - E(Y') &= \gamma Y + \delta - (\gamma E(Y) + \delta) = \gamma(Y - E(Y)). \end{aligned}$$

Luego

$$\begin{aligned} E([X' - E(X')] [Y' - E(Y')]) &= E(\alpha\gamma [X - E(X)] [Y - E(Y)]) \\ &= \alpha\gamma E([X - E(X)] [Y - E(Y)]) \end{aligned}$$

de donde se obtiene el resultado enunciado. \square

Ahora enunciaremos la desigualdad de Cauchy-Schwarz para variables aleatorias.

Teorema 7.24 (Desigualdad de Cauchy-Schwarz) Sean X e Y variables aleatorias. Entonces si las varianzas de ambas variables son finitas se tiene

$$E^2(XY) \leq E(X^2) E(Y^2), \quad (7.57)$$

y la igualdad ocurre si y sólo si existe α tal que $P(Y = \alpha X) = 1$. Además

$$\text{Cov}^2(X, Y) \leq \text{Var}(X)\text{Var}(Y), \quad (7.58)$$

y la igualdad ocurre si y sólo si existen escalares α, β tal que

$$P(Y = \alpha X + \beta) = 1. \quad (7.59)$$

Demostración.

Sea $Z = Y - \alpha X$. Entonces

$$Q(\alpha) = E(Z^2) = \alpha^2 E(X^2) + E(Y^2) - 2\alpha E(XY) \geq 0.$$

es un polinomio de segundo grado en α , no negativo y como tiene a lo sumo una raíz su discriminante es no positivo.

$$\Delta = 4E^2(XY) - 4E(X^2) E(Y^2) = 4(E^2(XY) - E(X^2) E(Y^2)) \leq 0.$$

Luego

$$E^2(XY) - E(X^2) E(Y^2) \leq 0,$$

de donde obtiene el resultado.

La igualdad se cumple si y sólo si $\Delta = 0$. Esto ocurre si y sólo si existe un único α tal que $Q(\alpha) = 0$. Esto es equivalente a que $E((Y - \alpha X)^2) = 0$, y esto a que $P(Y = \alpha X) = 1$.

La desigualdad (7.58) se obtiene aplicando (7.57) a $X^* = X - E(X)$ e $Y^* = Y - E(Y)$. Luego resulta que la correspondiente igualdad se cumple si y sólo si existe α tal que

$$P(Y - E(Y) = \alpha(X - E(X))) = 1.$$

Poniendo $\beta = E(Y) + \alpha E(X)$, esto es equivalente a (7.59). \square

Definición 7.10 Dadas las variables aleatorias X e Y se define el cuadrado del coeficiente de correlación entre ellas, y se denota por $\rho^2(X, Y)$ a

$$\rho^2(X, Y) = \frac{\text{Cov}^2(X, Y)}{\text{Var}(X) \text{Var}(Y)}.$$

También definimos el coeficiente de correlación entre X e Y por

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{[\text{Var}(X)]^{\frac{1}{2}} [\text{Var}(Y)]^{\frac{1}{2}}}.$$

De la desigualdad de Cauchy-Schwarz se deduce la siguiente propiedad.

Propiedad 7.19 Se tiene que

$$0 \leq \rho(X, Y)^2 \leq 1$$

y por lo tanto

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

Ademas $\rho(X, Y)^2 = 1$ es equivalente a que para algún α y β se tenga $P(Y = \alpha X + \beta) = 1$, es decir a que haya una relación lineal perfecta entre las variables X e Y .

7.9. Distribución Normal Bivariada.

En esta sección vamos a definir la distribución normal con medias, varianzas y covarianzas arbitrarias.

Queremos definir la distribución conjunta de un vector aleatorio $\mathbf{Y} = (Y_1, Y_2)$ a partir de fijar la distribución marginal de cada una de sus coordenadas y establecer un valor para la covarianza entre sus coordenadas. Es decir que queremos que la distribución conjunta del vector \mathbf{Y} sea tal que $Y_1 \sim N(\mu_1, \sigma_1^2)$, $Y_2 \sim N(\mu_2, \sigma_2^2)$, y tal que $\text{Cov}(Y_1, Y_2) = \sigma_{12}$, con las constantes $\mu_1, \mu_2, \sigma_1, \sigma_2$ y σ_{12} prefijadas arbitrariamente. Para que esto sea posible se tendrán que cumplir ciertas restricciones sobre estas constantes. Los valores μ_1, μ_2 no tienen que cumplir ningún requisito en particular, pero $\sigma_1^2 > 0$, $\sigma_2^2 > 0$ y σ_{12} debe cumplir la desigualdad de Cauchy-Schwarz que se puede escribir como

$$\sigma_{12}^2 \leq \sigma_1^2 \sigma_2^2.$$

Ahora bien si queremos una distribución bivariada absolutamente continua, no podrá cumplirse $\sigma_{12}^2 = \sigma_1^2 \sigma_2^2$, ya que en este caso (Y_1, Y_2) estaría sobre una recta que es un conjunto de superficie 0. Luego se deberá cumplir

$$\sigma_{12}^2 < \sigma_1^2 \sigma_2^2.$$

Sea la matriz Σ definida por

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}. \quad (7.60)$$

Luego $\det(\Sigma) = \sigma_1^2\sigma_2^2 - \sigma_{12}^2 > 0$.

Definamos la matriz de covarianza del vector \mathbf{Y} por

$$\Sigma_{\mathbf{Y}} = \begin{pmatrix} \text{Var}(Y_1) & \text{Cov}(Y_1, Y_2) \\ \text{Cov}(Y_2, Y_1) & \text{Var}(Y_2) \end{pmatrix}.$$

Luego queremos que

$$\Sigma_{\mathbf{Y}} = \Sigma.$$

Como $\det(\Sigma) = \sigma_1^2\sigma_2^2 - \sigma_{12}^2 > 0$ y $\sigma_1^2 > 0$, Σ resulta simétrica y definida positiva. Luego tiene al menos una “raíz cuadrada”. Es decir existe una matriz (no única)

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad (7.61)$$

tal que

$$\Sigma = AA^t, \quad (7.62)$$

donde A^t designa su traspuesta.

Estamos ahora en condiciones de construir el vector aleatorio buscado. Lo haremos en el siguiente teorema.

Teorema 7.25 *Sea $\Sigma \in \mathbb{R}^{2 \times 2}$ una matriz definida positiva dada por (7.60), $\mu = (\mu_1, \mu_2) \in \mathbb{R}^2$. Sea $A \in \mathbb{R}^{2 \times 2}$ dada por (7.61) que cumple (7.62). Sea $\mathbf{X} = (X_1, X_2)$ un vector aleatorio tal que X_1 y X_2 variables aleatorias independientes con distribución $N(0, 1)$. Se define el vector aleatorio $\mathbf{Y} = (Y_1, Y_2)$ por*

$$\mathbf{Y} = \mathbf{X}A^t + \mu.$$

Entonces resulta que

- (i) Y_1 tiene distribución $N(\mu_1, \sigma_1^2)$ e Y_2 tiene distribución $N(\mu_2, \sigma_2^2)$.
- (ii) $\text{Cov}(Y_1, Y_2) = \sigma_{12}$.
- (iii) La densidad del vector \mathbf{Y} está dada por

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{2\pi \det(\Sigma)^{\frac{1}{2}}} \exp\left(\frac{-1}{2} (\mathbf{y} - \mu) \Sigma^{-1} (\mathbf{y} - \mu)^t\right).$$

(iv) La forma cuadrática $Q(\mathbf{y}) = (\mathbf{y} - \boldsymbol{\mu}) \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})^t$ es igual a

$$\frac{1}{(1 - \rho^2)} \left[\frac{(y_1 - \mu_1)^2}{\sigma_1^2} + \frac{(y_2 - \mu_2)^2}{\sigma_2^2} - 2\rho (y_1 - \mu_1)(y_2 - \mu_2) \right].$$

Demostración.

(i) y (ii) Observemos que el vector \mathbf{Y} satisface

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \mu_1, \quad (7.63)$$

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \mu_2. \quad (7.64)$$

Como $E(X_1) = E(X_2) = 0$, resulta

$$E(Y_1) = \mu_1, \quad E(Y_2) = \mu_2.$$

Además como $\text{Cov}(X_1, X_2) = 0$, $\text{Var}(X_1) = \text{Var}(X_2) = 1$, resulta

$$\begin{aligned} \text{Var}(Y_1) &= a_{11}^2 \text{Var}(X_1) + a_{12}^2 \text{Var}(X_2) \\ &= a_{11}^2 + a_{12}^2. \end{aligned} \quad (7.65)$$

De modo análogo,

$$\text{Var}(Y_2) = a_{21}^2 + a_{22}^2, \quad (7.66)$$

y como $E(X_1X_2) = 0$, resulta

$$\begin{aligned} \text{Cov}(Y_1, Y_2) &= E([a_{11}X_1 + a_{12}X_2][a_{21}X_1 + a_{22}X_2]) \\ &= a_{11}a_{21}E(X_1^2) + a_{12}a_{22}E(X_2^2) + (a_{12}a_{21} + a_{11}a_{22})E(X_1X_2) \\ &= a_{11}a_{21} + a_{12}a_{22}. \end{aligned} \quad (7.67)$$

Luego

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{Y}} &= \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} \\ a_{11}a_{21} + a_{12}a_{22} & a_{21}^2 + a_{22}^2 \end{pmatrix} \\ &= \mathbf{A}\mathbf{A}^t \\ &= \boldsymbol{\Sigma} \\ &= \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}. \end{aligned} \quad (7.68)$$

De acuerdo al Teorema 6.7, como Y_1 e Y_2 son combinaciones lineales de normales independientes serán normales. Por (7.63), (7.65) y (7.68) resulta que la distribución de Y_1 es $N(\mu_1, \sigma_1^2)$. Por (7.64), (7.66) y (7.68) resulta que la distribución de Y_2 es $N(\mu_2, \sigma_2^2)$. Además, de (7.67) y (7.68) resulta que $\text{Cov}(Y_1, Y_2) = \sigma_{12}$. Esto prueba (i) y (ii).

- (iii) Vamos a calcular la distribución conjunta del vector \mathbf{Y} . Comencemos escribiendo la distribución conjunta del vector \mathbf{X} . Como X_1 y X_2 son independientes, la distribución conjunta de \mathbf{X} es el producto de las marginales,

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{2\pi} \exp\left(\frac{-x_1^2}{2}\right) \exp\left(\frac{-x_2^2}{2}\right) \\ &= \frac{1}{2\pi} \exp\left(\frac{-(x_1^2 + x_2^2)}{2}\right) \\ &= \frac{1}{2\pi} \exp\left(\frac{1}{2} \mathbf{x} \mathbf{x}^t\right), \end{aligned}$$

donde $\mathbf{x} \mathbf{x}^t = \|\mathbf{x}\|^2$.

Teniendo en cuenta que $\mathbf{X} = (\mathbf{Y} - \mu) (A^t)^{-1}$ se obtiene que el Jacobiano de esta transformación es $J = 1/\det(A^t)$. Además, como $\Sigma = AA^t$ se obtiene que $(\det(A))^2 = \det(\Sigma)$ o sea $\det(A) = \det(\Sigma)^{\frac{1}{2}}$ y por lo tanto $J = 1/\det(\Sigma)^{\frac{1}{2}}$. Entonces, a partir de la igualdad $(A^t)^{-1} A^{-1} = \Sigma^{-1}$ usando la fórmula para transformaciones de vectores aleatorios dada en el teorema 6.4, resulta

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \frac{1}{2\pi \det(\Sigma)^{\frac{1}{2}}} \exp\left(\frac{-1}{2} (\mathbf{y} - \mu) (A^t)^{-1} A^{-1} (\mathbf{y} - \mu)^t\right) \\ &= \frac{1}{2\pi \det(\Sigma)^{\frac{1}{2}}} \exp\left(\frac{-1}{2} (\mathbf{y} - \mu) \Sigma^{-1} (\mathbf{y} - \mu)^t\right). \end{aligned}$$

- (iv) Para hallar la forma cuadrática, calculemos primero el determinante de Σ

$$\det(\Sigma) = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2 = \sigma_1^2 \sigma_2^2 \left(1 - \frac{\sigma_{12}^2}{\sigma_1^2 \sigma_2^2}\right) = \sigma_1^2 \sigma_2^2 (1 - \rho^2).$$

Luego la inversa de Σ viene dada por

$$\Sigma^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}.$$

Entonces la forma cuadrática se puede escribir como

$$\begin{aligned} (\mathbf{y} - \mu) \Sigma^{-1} (\mathbf{y} - \mu)^t &= (\mathbf{y} - \mu) \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix} (\mathbf{y} - \mu)^t \\ &= \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \left[(y_1 - \mu_1)^2 \sigma_2^2 + (y_2 - \mu_2)^2 \sigma_1^2 - \right. \\ &\quad \left. 2(y_1 - \mu_1)(y_2 - \mu_2) \sigma_{12} \right]. \end{aligned}$$

Luego se tiene

$$\begin{aligned}
& (\mathbf{y} - \mu) \Sigma^{-1} (\mathbf{y} - \mu)^t \\
&= \frac{1}{1 - \rho^2} \left(\frac{(y_1 - \mu_1)^2}{\sigma_1^2} + \frac{(y_2 - \mu_2)^2}{\sigma_2^2} - 2 \frac{\sigma_{12}}{\sigma_1^2 \sigma_2^2} (y_1 - \mu_1) (y_2 - \mu_2) \right) \\
&= \frac{1}{1 - \rho^2} \left(\frac{(y_1 - \mu_1)^2}{\sigma_1^2} + \frac{(y_2 - \mu_2)^2}{\sigma_2^2} - 2 \frac{\rho}{\sigma_1 \sigma_2} (y_1 - \mu_1) (y_2 - \mu_2) \right). \square
\end{aligned}$$

Observación. El teorema anterior se demostró para el caso de dos variables. Sin embargo la densidad normal multivariada de cualquier dimensión que se define para vectores aleatorios $\mathbf{Y} \in \mathbb{R}^k$ tiene una expresión similar a la escrita en el punto (iii).

Observación. El máximo valor de $f_{\mathbf{Y}}$ se logra cuando se hace mínimo el exponente de la exponencial, esto es en $\mathbf{y} = \mu$. Por otro lado las curvas de nivel $f_{\mathbf{Y}}(\mathbf{y}) = c$ (con c constante) son elipses cuyas direcciones principales vienen dadas por los autovectores de Σ^{-1} . Si la $\text{Cov}(Y_1, Y_2) = 0$ entonces, la matriz Σ es diagonal y las direcciones son paralelas a los ejes coordenados, dando lugar a circunferencias como curvas de nivel en este caso.

Definición 7.11 *Se dice que el vector \mathbf{Y} tiene distribución normal bivariada con media μ y matriz de covarianza Σ definida positiva, que se denotará por $N_2(\mu, \Sigma)$ si su función densidad es*

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{2\pi \det(\Sigma)^{\frac{1}{2}}} \exp\left(\frac{-1}{2} (\mathbf{y} - \mu) \Sigma^{-1} (\mathbf{y} - \mu)^t\right).$$

Capítulo 8

Teoría de la Predicción.

8.1. Error cuadrático medio y predictores óptimos.

En esta sección veremos como utilizar ciertas variables conocidas para predecir otras variables que no se pueden observar en un determinado momento. Por ejemplo se quiere predecir la cantidad de lluvia que mañana caerá en determinada región, utilizaremos otras variables que se puedan medir hoy. Quisiéramos encontrar el predictor que se aproxime más a la variable a predecir, entre todos los predictores pertenecientes a un conjunto dado.

Sea \mathcal{P} un conjunto de predictores para la variable aleatoria Y , que forman un espacio vectorial. Cada elemento de \mathcal{P} es una variables aleatoria observable. Supongamos que se quiere predecir a Y a través de $\hat{Y} \in \mathcal{P}$. ¿Cómo se puede medir la bondad de un predictor \hat{Y} cualquiera? Se pueden considerar las siguientes alternativas:

Definición 8.1 *El error cuadrático medio del predictor \hat{Y} para predecir a Y está dado por*

$$\text{ECM}(\hat{Y}, Y) = E\left(\left(Y - \hat{Y}\right)^2\right)$$

y el error absoluto medio

$$\text{EAM}(\hat{Y}, Y) = E\left(|Y - \hat{Y}|\right).$$

Si usamos como criterio de bondad de un predictor el error cuadrático medio, diremos que $\hat{Y}_0 \in \mathcal{P}$ es un *predictor óptimo de Y en \mathcal{P}* , si dado otro $\hat{Y} \in \mathcal{P}$ se tiene

$$\text{ECM}(\hat{Y}_0, Y) \leq \text{ECM}(\hat{Y}, Y).$$

A continuación damos un criterio suficiente para obtener un predictor óptimo usando el criterio del error cuadrático medio.

Teorema 8.1 Una condición suficiente para que $\hat{Y}_0 \in \mathcal{P}$ sea un predictor óptimo usando el criterio del error cuadrático medio es que

$$E\left(\left(Y - \hat{Y}_0\right) \hat{Y}\right) = 0 \quad (8.1)$$

para todo $\hat{Y} \in \mathcal{P}$. Además, si \hat{Y}_0 satisface (8.1), es esencialmente el único predictor óptimo. Es decir si $\hat{Y} \in \mathcal{P}$ satisface $\text{ECM}(\hat{Y}_0, Y) = \text{ECM}(\hat{Y}, Y)$ entonces $P(\hat{Y} = \hat{Y}_0) = 1$.

Observación. La condición (8.1) se puede interpretar como que el error de predicción $(Y - \hat{Y}_0)$ es ortogonal a todo elemento de \mathcal{P} cuando el producto escalar está definido por $\langle Y, X \rangle = E(YX)$ en el espacio de Hilbert de las variables aleatorias.

Demostración. Sea $\hat{Y} \in \mathcal{P}$. Entonces

$$\begin{aligned} \text{ECM}(\hat{Y}, Y) &= E\left(\left(Y - \hat{Y}\right)^2\right) = E\left(\left[\left(Y - \hat{Y}_0\right) + \left(\hat{Y}_0 - \hat{Y}\right)\right]^2\right) \\ &= E\left(\left(Y - \hat{Y}_0\right)^2\right) + E\left(\left(\hat{Y}_0 - \hat{Y}\right)^2\right) \\ &\quad + 2E\left(\left(\hat{Y}_0 - \hat{Y}\right)\left(Y - \hat{Y}_0\right)\right). \end{aligned}$$

Usando la condición de ortogonalidad, como $\hat{Y}_0 - \hat{Y} \in \mathcal{P}$ se tiene

$$E\left(\left(\hat{Y}_0 - \hat{Y}\right)\left(Y - \hat{Y}_0\right)\right) = 0,$$

y luego

$$\begin{aligned} \text{ECM}(\hat{Y}, Y) &= E\left(\left(Y - \hat{Y}_0\right)^2\right) + E\left(\left(\hat{Y}_0 - \hat{Y}\right)^2\right) \\ &\geq E\left(\left(Y - \hat{Y}_0\right)^2\right) \\ &= \text{ECM}(\hat{Y}_0, Y), \end{aligned}$$

y por lo tanto \hat{Y}_0 es óptimo.

Además si \hat{Y} fuera también óptimo se tendría $E\left(\left(\hat{Y}_0 - \hat{Y}\right)^2\right) = 0$ y siendo $\left(\hat{Y}_0 - \hat{Y}\right)^2 \geq 0$ resultaría $P(\hat{Y} = \hat{Y}_0) = 1$, en virtud de la Propiedad 7.10. \square .

El siguiente Teorema simplifica la verificación de la condición (8.1).

Teorema 8.2 Sea \mathcal{P} un espacio vectorial de predictores de la variable aleatoria Y de dimensión finita y sea $\{\widehat{Y}_1, \dots, \widehat{Y}_k\}$ una base de \mathcal{P} . La condición necesaria y suficiente para que se cumpla (8.1) es que

$$E\left(\left(Y - \widehat{Y}_0\right) \widehat{Y}_i\right) = 0, \quad 1 \leq i \leq k. \quad (8.2)$$

Demostración. Claramente es una condición necesaria. Veamos que es suficiente. Sea \widehat{Y} cualquier elemento de \mathcal{P} , entonces existen escalares $\alpha_1, \dots, \alpha_k$ tal que $\widehat{Y} = \sum_{i=1}^k \alpha_i \widehat{Y}_i$. Luego si para $i = 1, 2, \dots, k$ se cumple que

$$E\left(\left(Y - \widehat{Y}_0\right) \widehat{Y}_i\right) = 0,$$

resulta también que

$$\begin{aligned} E\left(\left(Y - \widehat{Y}_0\right) \widehat{Y}\right) &= E\left(\left(Y - \widehat{Y}_0\right) \sum_{i=1}^k \alpha_i \widehat{Y}_i\right) \\ &= \sum_{i=1}^k \alpha_i E\left(\left(Y - \widehat{Y}_0\right) \widehat{Y}_i\right) = 0. \quad \square \end{aligned}$$

8.2. Predictores constantes.

Se pueden considerar distintos conjuntos de predictores. Comenzaremos con los predictores constantes.

Sea (Ω, \mathcal{A}, P) un espacio de probabilidad, Y una variable aleatoria a predecir y consideremos

$$\mathcal{P}_1 = \{\widehat{Y} : \widehat{Y} \text{ es una variable aleatoria constante}\}.$$

El siguiente Teorema determina el predictor óptimo perteneciente a \mathcal{P}_1 .

Teorema 8.3 El predictor $\widehat{Y}_0 = E(Y)$ es el de menor error cuadrático medio en \mathcal{P}_1 . Además $\text{ECM}(\widehat{Y}_0, Y) = \text{Var}(Y)$.

Demostración. Una base de \mathcal{P}_1 es $\{\widehat{Y}_1\}$ donde $\widehat{Y}_1 = 1$. Como

$$E\left(\left(Y - \widehat{Y}_0\right) 1\right) = E(Y - E(Y)) = E(Y) - E(Y) = 0,$$

resulta $\widehat{Y}_0 = E(Y)$ el predictor de menor error cuadrático medio.

Además

$$\begin{aligned} \text{ECM}(\widehat{Y}_0, Y) &= E((Y - \widehat{Y}_0)^2) \\ &= E((Y - E(Y))^2) \\ &= \text{Var}(Y). \quad \square \end{aligned}$$

Designamos el predictor óptimo para Y en \mathcal{P}_1 por $\widehat{Y}_{0,C}$. En la práctica únicamente se usa un predictor constante si no se observan otras variables vinculadas a Y .

8.3. Predictores lineales.

Sea ahora (Ω, \mathcal{A}, P) un espacio de probabilidad, Y una variable aleatoria a predecir y X otra variable aleatoria observada. Consideremos el siguiente conjunto de predictores

$$\mathcal{P}_2 = \{\widehat{Y} : \widehat{Y} = \alpha X + \beta, \quad \alpha, \beta \in \mathbb{R}\}.$$

\mathcal{P}_2 es el conjunto de variables aleatorias que se obtiene por una transformación lineal de la variable X . Claramente $\mathcal{P}_1 \subset \mathcal{P}_2$, y por lo tanto el error cuadrático medio del predictor óptimo en \mathcal{P}_2 será menor o igual que el del predictor óptimo en \mathcal{P}_1 . Por esta razón, si denotamos por $\widehat{Y}_{0,L}$ el predictor óptimo en \mathcal{P}_2 , resulta claro que

$$\text{ECM}(Y, \widehat{Y}_{0,L}) \leq \text{ECM}(Y, \widehat{Y}_{0,C}).$$

El siguiente Teorema caracteriza el predictor óptimo en \mathcal{P}_2 .

Teorema 8.4 (i) *El predictor de menor error cuadrático medio en \mathcal{P}_2 está dado por $\widehat{Y}_{0,L} = \alpha X + \beta$ con*

$$\beta = E(Y) - \alpha E(X) \tag{8.3}$$

y

$$\alpha = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}. \tag{8.4}$$

(ii) *El error cuadrático medio de $\widehat{Y}_{0,L}$ está dado por*

$$\text{ECM}(\widehat{Y}_{0,L}, Y) = \text{Var}(Y) - \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)}. \tag{8.5}$$

Demostración. Una base de \mathcal{P}_2 es $\{\widehat{Y}_1, \widehat{Y}_2\}$ donde $\widehat{Y}_1 = X$ e $\widehat{Y}_2 = 1$. Luego el predictor óptimo $\widehat{Y}_{0,L}$ debe satisfacer

$$E((Y - \alpha X - \beta) X) = 0 \tag{8.6}$$

y

$$E((Y - \alpha X - \beta) 1) = 0. \tag{8.7}$$

De la condición (8.6) se obtiene

$$E(Y) - \alpha E(X) - \beta = 0,$$

de donde resulta (8.3).

Ahora multiplicando (8.7) por $E(X)$ resulta

$$E((Y - \alpha X - \beta) E(X)) = 0,$$

y restándola de (8.6) obtenemos

$$E((Y - \alpha X - \beta)(X - E(X))) = 0.$$

Reemplazando β por (8.3) obtenemos

$$E((Y - \alpha X - E(Y) + \alpha E(X))(X - E(X))) = 0,$$

y por lo tanto

$$E((Y - E(Y)) - \alpha(X - E(X)))(X - E(X)) = 0.$$

Entonces distribuyendo la esperanza se obtiene

$$\begin{aligned} \text{Cov}(X, Y) &= E[(Y - E(Y))(X - E(X))] \\ &= \alpha E[(X - E(X))^2] \\ &= \alpha \text{Var}(X), \end{aligned}$$

y por lo tanto resulta (8.4).

Ahora calcularemos el error cuadrático medio de $\hat{Y}_{0,L}$. Usando (8.3) obtenemos

$$\begin{aligned} \text{ECM}(\hat{Y}_{0,L}, Y) &= E\left([Y - \alpha X - \beta]^2\right) \\ &= E\left([Y - \alpha X - E(Y) + \alpha E(X)]^2\right) = \\ &= E\left([(Y - E(Y)) - \alpha(X - E(X))]^2\right) = \\ &= E\left([Y - E(Y)]^2\right) + \alpha^2 E\left([X - E(X)]^2\right) \\ &\quad - 2\alpha E\left([Y - E(Y)][X - E(X)]\right). \end{aligned}$$

Luego, usando (8.4) se obtiene

$$\begin{aligned} \text{ECM}(\hat{Y}_{0,L}, Y) &= \text{Var}(Y) + \alpha^2 \text{Var}(X) - 2\alpha \text{Cov}(X, Y) \\ &= \text{Var}(Y) + \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)} - 2 \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)} \\ &= \text{Var}(Y) - \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)}. \quad \square \end{aligned}$$

Para evaluar cuánto mejora el error cuadrático medio cuando se usa $\hat{Y}_{0,L}$

en vez de $\widehat{Y}_{0,C}$, calculemos su decrecimiento relativo

$$\begin{aligned} & \frac{\text{ECM}\left(\widehat{Y}_{0,C}, Y\right) - \text{ECM}\left(\widehat{Y}_{0,L}, Y\right)}{\text{ECM}\left(\widehat{Y}_{0,C}, Y\right)} \\ &= \frac{\text{Var}(Y) - \left(\text{Var}(Y) - \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)}\right)}{\text{ECM}\left(\widehat{Y}_{0,C}, Y\right)} \\ &= \frac{\frac{\text{Cov}^2(X, Y)}{\text{Var}(X)}}{\text{Var}(Y)} = \frac{\text{Cov}^2(X, Y)}{\text{Var}(X) \text{Var}(Y)} = \rho^2(X, Y). \end{aligned}$$

Esto permite interpretar el coeficiente $\rho^2(X, Y)$ como el decrecimiento relativo del error cuadrático medio cuando se usa un predictor lineal basado en X en vez de un predictor constante. Por lo tanto $\rho^2(X, Y)$ mide la utilidad de la variable X para predecir a Y por una función lineal. Observemos que a partir de esta igualdad puede obtenerse nuevamente la desigualdad de Cauchy-Schwarz. En efecto, como $0 \leq \text{ECM}\left(\widehat{Y}_{0,C}, Y\right) - \text{ECM}\left(\widehat{Y}_{0,L}, Y\right) \leq \text{ECM}\left(Y, \widehat{Y}_{0,C}\right)$, se obtiene $0 \leq \rho^2(X, Y) \leq 1$.

Veremos ahora el significado de los casos extremos $\rho^2(X, Y) = 1$ y $\rho^2(X, Y) = 0$. $\rho^2(X, Y) = 1$ es equivalente a $\text{ECM}\left(Y, \widehat{Y}_{0,L}\right) = 0$ y esto es equivalente a $E\left(\left(Y - \widehat{Y}_{0,L}\right)^2\right) = 0$, que a su vez es equivalente a

$$P\left(Y = \widehat{Y}_{0,L}\right) = P\left(Y = \alpha X + \beta\right) = 1,$$

en virtud de la Propiedad 7.10.

Es decir $\rho^2(X, Y) = 1$ es equivalente a que hay una relación lineal perfecta entre X e Y con probabilidad 1.

Existen dos posibilidades para $\rho^2(X, Y) = 1$: o bien $\rho(X, Y) = 1$ o $\rho(X, Y) = -1$. El signo de $\rho(X, Y)$ coincide con el de $\text{Cov}(X, Y)$ que es el mismo que el de la pendiente del predictor lineal óptimo. Luego $\rho(X, Y) = 1$ indica que la relación entre la X y la Y es creciente y $\rho(X, Y) = -1$ que la relación es decreciente.

Veremos ahora como se interpreta $\rho^2(X, Y) = 0$. En este caso

$$\text{ECM}\left(\widehat{Y}_{0,L}, Y\right) = \text{ECM}\left(\widehat{Y}_{0,C}, Y\right)$$

y $\text{Cov}(X, Y) = 0$. Por lo tanto $\alpha = 0$, y se puede concluir que la variable X no tiene utilidad para predecir a Y cuando se utilizan predictores lineales.

Se deja como ejercicio probar que la recta $Y = \alpha X + \beta$ pasa por el punto $(E(X), E(Y))$, es decir que cuando la variable X toma el valor $E(X)$ el valor predicho para la variable Y es $E(Y)$.

Capítulo 9

Esperanza y distribución condicional.

9.1. Caso discreto.

Sean dos variables aleatorias discretas X, Y definidas sobre un mismo espacio de probabilidad (Ω, \mathcal{A}, P) . Sea $R_X = \{x : p_X(x) > 0\}$ y $R_Y = \{y : p_Y(y) > 0\}$. Luego, para cada $x \in R_X$ definimos la función de probabilidad de Y condicional a $X = x$ como

$$p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)}.$$

A veces se utiliza la notación $p_{Y|X=x}(y)$ para la función de probabilidad condicional.

Para cada $x \in R_X$ fijo esta función es una función de densidad de probabilidad ya que

$$\sum_{y \in R_Y} p_{Y|X}(y|x) = \sum_{y \in R_Y} \frac{p_{XY}(x, y)}{p_X(x)} = \frac{1}{p_X(x)} \sum_{y \in R_Y} p_{XY}(x, y) = \frac{p_X(x)}{p_X(x)} = 1,$$

y representa la distribución de Y una vez conocido que el valor de $X = x$.

Si se tienen dos vectores discretos $\mathbf{X} = (X_1, \dots, X_k)$, $\mathbf{Y} = (Y_1, \dots, Y_h)$ podemos definir una noción análoga. Sea $R_{\mathbf{X}} = \{\mathbf{x} \in \mathbb{R}^k : p_{\mathbf{X}}(\mathbf{x}) > 0\}$, luego para todo $\mathbf{x} \in R_{\mathbf{X}}$ definimos

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})}, \quad (9.1)$$

y también se tendrá

$$\sum_{\mathbf{y} \in R_Y} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = 1.$$

Esto permite calcular probabilidades que involucran a \mathbf{Y} cuando sabemos que el evento $\{\mathbf{X} = \mathbf{x}\}$ ha ocurrido. En efecto, si $B \in \mathcal{B}^h$ (borelianos de \mathbb{R}^h) definimos

$$P(\mathbf{Y} \in B \mid \mathbf{X} = \mathbf{x}) = \sum_{\mathbf{y} \in R_{\mathbf{Y}} \cap B} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}).$$

Sea ahora Y una variable aleatoria y \mathbf{X} un vector aleatorio de dimensión k , ambos discretos. La esperanza condicional de la variable Y condicional a $\mathbf{X} = \mathbf{x}$ se define como la esperanza de Y utilizando como distribución de esta variable la distribución determinada por (9.1). Es decir, la esperanza condicional se define por

$$E(Y|\mathbf{X} = \mathbf{x}) = \sum_{y \in R_Y} yp_{Y|\mathbf{X}}(y|\mathbf{x}). \quad (9.2)$$

Este valor representa la esperanza de la variable Y cuando se sabe que el vector \mathbf{X} ha tomado el valor \mathbf{x} .

Llamemos $g(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$, luego $g : R_X \rightarrow \mathbb{R}$. Vamos a definir ahora una variable aleatoria que llamaremos esperanza de Y condicional a X , y que notaremos por $E(Y|\mathbf{X})$. Esta variable se define por

$$E(Y|\mathbf{X}) = g(\mathbf{X}).$$

Vamos ahora a mostrar el siguiente teorema, que relaciona las esperanzas de ambas variables aleatorias.

Teorema 9.1 *Si Y tiene esperanza finita, entonces se tiene que*

$$E(E(Y|\mathbf{X})) = E(Y).$$

Demostración. Tenemos que

$$E(E(Y|\mathbf{X})) = E(g(\mathbf{X})) = \sum_{\mathbf{x} \in R_{\mathbf{X}}} g(\mathbf{x})p_{\mathbf{X}}(\mathbf{x}).$$

Utilizando que $g(\mathbf{x})$ viene dado por (9.2), se tiene

$$\begin{aligned}
 E(E(Y|\mathbf{X})) &= \sum_{\mathbf{x} \in R_{\mathbf{X}}} \left(\sum_{y \in R_Y} y p_{Y|\mathbf{X}}(y|\mathbf{x}) \right) p_{\mathbf{X}}(\mathbf{x}) \\
 &= \sum_{\mathbf{x} \in R_{\mathbf{X}}} \left(\sum_{y \in R_Y} y \frac{p_{\mathbf{X}Y}(\mathbf{x}, y)}{p_{\mathbf{X}}(\mathbf{x})} \right) p_{\mathbf{X}}(\mathbf{x}) \\
 &= \sum_{\mathbf{x} \in R_{\mathbf{X}}} \left(\sum_{y \in R_Y} y p_{\mathbf{X}Y}(\mathbf{x}, y) \right) \\
 &= \sum_{y \in R_Y} y \left(\sum_{\mathbf{x} \in R_{\mathbf{X}}} p_{\mathbf{X}Y}(\mathbf{x}, y) \right) \\
 &= \sum_{y \in R_Y} y p_Y(y) \\
 &= E(Y).
 \end{aligned}$$

El cambio en el orden de la suma se encuentra justificado pues la suma converge. Luego el teorema queda demostrado. \square

Ejemplo 9.1 *Supongamos que se hace una primera serie de n tiradas de una moneda y sea X el número de caras obtenido. En base al resultado de la primera serie de tiradas, se inicia una segunda serie de X tiradas. Sea Y el número de caras obtenidas en esta segunda serie. Calcular la $E(Y)$.*

Si $X = x$, la distribución de Y condicional a $X = x$ es binomial $\text{Bi}(0,50, x)$. Luego $g(x) = E(Y|X = x) = 0,50x$. Luego $E(Y|X) = g(X) = 0,50X$, y por lo tanto $E(Y) = E(E(Y|X)) = 0,50E(X)$. Como X es $\text{Bi}(0,50, n)$, entonces $E(X) = 0,5n$. Por lo tanto $E(Y) = 0,25n$.

Teorema 9.2 (i) *Si \mathbf{X} e \mathbf{Y} son dos vectores aleatorios independientes, entonces se tiene*

- a) $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = p_{\mathbf{Y}}(\mathbf{y})$
- b) *Si Y es una variable aleatoria y $E(Y)$ existe y es finita entonces $E(Y|\mathbf{X} = \mathbf{x}) = E(Y)$.*

(ii) *Sean \mathbf{X} e \mathbf{Y} dos vectores aleatorios tales $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = p(\mathbf{y})$ para todo $\mathbf{x} \in R_{\mathbf{X}}$. Entonces $p_{\mathbf{Y}}(\mathbf{y}) = p(\mathbf{y})$, y \mathbf{X} e \mathbf{Y} son independientes.*

Demostración.

- (i) a) se deduce del hecho de que $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = p_{\mathbf{Y}}(\mathbf{y})$ implica que $p_{\mathbf{X}Y}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}}(\mathbf{y})$.

b) es inmediata.

(ii) Para probar (ii) observemos que $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = p(\mathbf{y})$ implica que

$$p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x})p(\mathbf{y}), \quad (9.3)$$

y por lo tanto

$$p_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{x} \in R_{\mathbf{X}}} p_{\mathbf{X}}(\mathbf{x})p(\mathbf{y}) = p(\mathbf{y}) \sum_{\mathbf{x} \in R_{\mathbf{X}}} p_{\mathbf{X}}(\mathbf{x}) = p(\mathbf{y}).$$

Luego reemplazando en (9.3) se obtiene

$$p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}}(\mathbf{y}), \quad (9.4)$$

y esto implica que \mathbf{X} e \mathbf{Y} son independientes. \square

Teorema 9.3 Si $P(Y = c) = 1$, entonces, cualquiera sea el vector \mathbf{X} , se tiene

$$(i) p_{Y|\mathbf{X}}(c|\mathbf{x}) = 1.$$

$$(ii) E(Y|\mathbf{X} = \mathbf{x}) = c.$$

Demostración. Tenemos que

$$\{\mathbf{X} = \mathbf{x}\} = (\{\mathbf{X} = \mathbf{x}\} \cap \{Y = c\}) \cup (\{\mathbf{X} = \mathbf{x}\} \cap \{Y \neq c\}).$$

Como $P(\{\mathbf{X} = \mathbf{x}\} \cap \{Y \neq c\}) = 0$, se tiene

$$\begin{aligned} p_{\mathbf{X}}(\mathbf{x}) &= P(\mathbf{X} = \mathbf{x}) = P(\mathbf{X} = \mathbf{x}, Y = c) \\ &= p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, c). \end{aligned}$$

Por lo tanto

$$p_{Y|\mathbf{X}}(c|\mathbf{x}) = \frac{p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, c)}{p_{\mathbf{X}}(\mathbf{x})} = 1.$$

Como en este caso $R_Y = \{c\}$, se tiene

$$\begin{aligned} E(Y|\mathbf{X} = \mathbf{x}) &= \sum_{y \in R_Y} yp_{Y|\mathbf{X}}(y|\mathbf{x}) \\ &= cp_{Y|\mathbf{X}}(c|\mathbf{x}) \\ &= c1 \\ &= c, \end{aligned}$$

y el teorema queda demostrado. \square

Sean ahora dos vectores aleatorios discretos, $\mathbf{X} = (X_1, \dots, X_k)$, $\mathbf{Y} = (Y_1, \dots, Y_j)$, y sea $Z = h(\mathbf{X}, \mathbf{Y})$, donde $h : \mathbb{R}^{k+j} \rightarrow \mathbb{R}$ es una función medible. El siguiente Teorema muestra cómo se calcula $E(Z|\mathbf{X} = \mathbf{x})$.

Teorema 9.4 Sean \mathbf{X} , \mathbf{Y} dos vectores aleatorios discretos de dimensiones k y j , y sea $h : \mathbb{R}^{k+j} \rightarrow \mathbb{R}$ una función medible. Definamos la variable aleatoria discreta $Z = h(\mathbf{X}, \mathbf{Y})$, y supongamos que tiene esperanza finita. Entonces para todo $\mathbf{x} \in R_{\mathbf{X}}$ se tiene

$$E(Z|\mathbf{X} = \mathbf{x}) = \sum_{\mathbf{y} \in R_{\mathbf{Y}}} h(\mathbf{x}, \mathbf{y}) p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}).$$

Demostración. Comenzaremos calculando la función de probabilidad conjunta de (\mathbf{X}, Z) . Sea $R_Z^{\mathbf{x}} = \{z : z = h(\mathbf{x}, \mathbf{y}) \text{ para } \mathbf{y} \in R_{\mathbf{Y}}\}$, y para todo $z \in R_Z^{\mathbf{x}}$ definamos $A_z^{\mathbf{x}} = \{\mathbf{y} : h(\mathbf{x}, \mathbf{y}) = z\}$. Es fácil ver que:

si $z \neq z'$ entonces $A_z^{\mathbf{x}} \cap A_{z'}^{\mathbf{x}} = \emptyset$, y que

$$\bigcup_{z \in R_Z^{\mathbf{x}}} A_z^{\mathbf{x}} = R_{\mathbf{Y}}. \quad (9.5)$$

Es inmediato que

$$p_{\mathbf{X}Z}(\mathbf{x}, z) = \begin{cases} P(\mathbf{X} = \mathbf{x}, \mathbf{Y} \in A_z^{\mathbf{x}}) = \sum_{\mathbf{y} \in A_z^{\mathbf{x}}} p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) & \text{si } \mathbf{x} \in R_{\mathbf{X}}, z \in R_Z^{\mathbf{x}} \\ 0 & \text{en otro caso,} \end{cases}$$

y luego, para $\mathbf{x} \in R_{\mathbf{X}}$ se tiene

$$p_{Z|\mathbf{X}}(z|\mathbf{x}) = \frac{p_{\mathbf{X}Z}(\mathbf{x}, z)}{p_{\mathbf{X}}(\mathbf{x})} = \begin{cases} \sum_{\mathbf{y} \in A_z^{\mathbf{x}}} \frac{p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})} & \text{si } z \in R_Z^{\mathbf{x}} \\ 0 & \text{en otro caso.} \end{cases}$$

Por lo tanto se tiene

$$p_{Z|\mathbf{X}}(z|\mathbf{x}) = \begin{cases} \sum_{\mathbf{y} \in A_z^{\mathbf{x}}} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) & \text{si } z \in R_Z^{\mathbf{x}} \\ 0 & \text{en otro caso.} \end{cases} \quad (9.6)$$

Luego utilizando (9.6) se tiene

$$\begin{aligned} E(Z|\mathbf{X} = \mathbf{x}) &= \sum_{z \in R_Z^{\mathbf{x}}} z p_{Z|\mathbf{X}}(z|\mathbf{x}) \\ &= \sum_{z \in R_Z^{\mathbf{x}}} z \sum_{\mathbf{y} \in A_z^{\mathbf{x}}} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \\ &= \sum_{z \in R_Z^{\mathbf{x}}} \sum_{\mathbf{y} \in A_z^{\mathbf{x}}} z p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}), \end{aligned}$$

y como para $\mathbf{y} \in A_z^{\mathbf{x}}$, se tiene $h(\mathbf{x}, \mathbf{y}) = z$, utilizando (9.5) obtenemos

$$\begin{aligned} E(Z|\mathbf{X} = \mathbf{x}) &= \sum_{z \in R_Z^{\mathbf{x}}} \sum_{\mathbf{y} \in A_z^{\mathbf{x}}} h(\mathbf{x}, \mathbf{y}) p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \\ &= \sum_{\mathbf{y} \in R_{\mathbf{Y}}} h(\mathbf{x}, \mathbf{y}) p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}), \end{aligned}$$

probando por lo tanto el teorema. \square

El Teorema 9.4 se puede interpretar como que $E(Z|\mathbf{X} = \mathbf{x})$ se calcula como la esperanza de $h(\mathbf{Y}, \mathbf{x})$ (variable aleatoria que depende únicamente del vector aleatorio \mathbf{Y} , ya que \mathbf{x} es tratada como si fuera constante) utilizando $p_{\mathbf{Y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$ como función de probabilidad puntual de \mathbf{Y}

Veamos qué propiedades de la esperanza condicional se deducen del Teorema 9.4.

Propiedad 9.1 Sean \mathbf{X} un vector aleatorio discreto de dimensión k e \mathbf{Y} un vector aleatorio discreto de dimensión j , y sean $r : \mathbb{R}^k \rightarrow \mathbb{R}$ y $s : \mathbb{R}^j \rightarrow \mathbb{R}$ funciones medibles tales que las variables aleatorias $r(\mathbf{X})s(\mathbf{Y})$, $r(\mathbf{X})$ y $s(\mathbf{Y})$ tienen esperanza finita. Entonces se tiene

$$E(r(\mathbf{X})s(\mathbf{Y})|\mathbf{X} = \mathbf{x}) = r(\mathbf{x})E(s(\mathbf{Y})|\mathbf{X} = \mathbf{x}).$$

Demostración. Utilizando el Teorema 9.4 con $h(\mathbf{x}, \mathbf{y}) = r(\mathbf{x})s(\mathbf{y})$ que tiene esperanza finita, se tiene

$$\begin{aligned} E(r(\mathbf{X})s(\mathbf{Y}) | \mathbf{X} = \mathbf{x}) &= \sum_{\mathbf{y} \in R_{\mathbf{Y}}} r(\mathbf{x})s(\mathbf{y})p_{\mathbf{Y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) \\ &= r(\mathbf{x}) \sum_{\mathbf{y} \in R_{\mathbf{Y}}} s(\mathbf{y})p_{\mathbf{Y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) \\ &= r(\mathbf{x})E(s(\mathbf{Y})|\mathbf{X} = \mathbf{x}), \end{aligned}$$

y luego la propiedad queda demostrada. \square

Propiedad 9.2 Sea \mathbf{X} un vector aleatorio discreto de dimensión k , y sea $r : \mathbb{R}^k \rightarrow \mathbb{R}$ una función medible tal que la variable $r(\mathbf{X})$ tiene esperanza finita. Luego

$$E(r(\mathbf{X})|\mathbf{X} = \mathbf{x}) = r(\mathbf{x}).$$

Demostración. La demostración resulta de la Propiedad 9.1 tomando $s(\mathbf{y}) = 1$, ya que entonces $E(r(\mathbf{X})|\mathbf{X} = \mathbf{x}) = r(\mathbf{x})E(1|\mathbf{X} = \mathbf{x})$. Luego por el Teorema 9.4 resulta la Propiedad 9.2. \square

Propiedad 9.3 (Linealidad de la esperanza condicional) Sean Y_1 e Y_2 variables aleatorias discretas con esperanza finita, y sea \mathbf{X} un vector aleatorio discreto, entonces

$$E(c_1Y + c_2Y_2|\mathbf{X} = \mathbf{x}) = c_1E(Y_1|\mathbf{X} = \mathbf{x}) + c_2E(Y_2|\mathbf{X} = \mathbf{x}).$$

Demostración.

Sea $\mathbf{Y} = (Y_1, Y_2)$ y definamos $h(\mathbf{x}, \mathbf{y}) = c_1 y_1 + c_2 y_2$, $h_1(\mathbf{x}, \mathbf{y}) = y_1$ y $h_2(\mathbf{x}, \mathbf{y}) = y_2$. Entonces se tiene $h(\mathbf{x}, \mathbf{y}) = c_1 h_1(\mathbf{x}, \mathbf{y}) + c_2 h_2(\mathbf{x}, \mathbf{y})$. Luego tenemos

$$\begin{aligned}
 E(c_1 Y_1 + c_2 Y_2 | \mathbf{X} = \mathbf{x}) &= E(h(\mathbf{X}, \mathbf{Y}) | \mathbf{X} = \mathbf{x}) \\
 &= \sum_{\mathbf{y} \in R_{\mathbf{Y}}} h(\mathbf{x}, \mathbf{y}) p_{\mathbf{Y} | \mathbf{X}}(\mathbf{y} | \mathbf{x}) \\
 &= \sum_{\mathbf{y} \in R_{\mathbf{Y}}} (c_1 h_1(\mathbf{x}, \mathbf{y}) + c_2 h_2(\mathbf{x}, \mathbf{y})) p_{\mathbf{Y} | \mathbf{X}}(\mathbf{y} | \mathbf{x}) \\
 &= c_1 \sum_{\mathbf{y} \in R_{\mathbf{Y}}} h_1(\mathbf{x}, \mathbf{y}) p_{\mathbf{Y} | \mathbf{X}}(\mathbf{y} | \mathbf{x}) + c_2 \sum_{\mathbf{y} \in R_{\mathbf{Y}}} h_2(\mathbf{x}, \mathbf{y}) p_{\mathbf{Y} | \mathbf{X}}(\mathbf{y} | \mathbf{x}) \\
 &= c_1 E(h_1(\mathbf{X}, \mathbf{Y}) | \mathbf{X} = \mathbf{x}) + c_2 E(h_2(\mathbf{X}, \mathbf{Y}) | \mathbf{X} = \mathbf{x}) \\
 &= c_1 E(Y_1 | \mathbf{X} = \mathbf{x}) + c_2 E(Y_2 | \mathbf{X} = \mathbf{x}),
 \end{aligned}$$

y la Propiedad 9.3 queda demostrada. \square

Propiedad 9.4 (i) Si $P(Y \geq 0) = 1$, $E(Y | \mathbf{X} = \mathbf{x}) \geq 0$.

(ii) $E(Y^2 | \mathbf{X} = \mathbf{x}) \geq E^2(Y | \mathbf{X} = \mathbf{x})$.

(iii) Si $E(Y^2) < \infty$, entonces $E(E^2(Y | \mathbf{X})) < \infty$.

Demostración.

(i) Es inmediato de la definición.

(ii) Para demostrar (ii), observemos que por (i)

$$\begin{aligned}
 0 &\leq E([Y - E(Y | \mathbf{X} = \mathbf{x})]^2 | \mathbf{X} = \mathbf{x}) \\
 &= E([Y^2 - 2YE(Y | \mathbf{X} = \mathbf{x}) + E^2(Y | \mathbf{X} = \mathbf{x})] | \mathbf{X} = \mathbf{x}) \\
 &= E(Y^2 | \mathbf{X} = \mathbf{x}) - 2E(Y | \mathbf{X} = \mathbf{x})E(Y | \mathbf{X} = \mathbf{x}) + E^2(Y | \mathbf{X} = \mathbf{x}) \\
 &= E(Y^2 | \mathbf{X} = \mathbf{x}) - E^2(Y | \mathbf{X} = \mathbf{x}),
 \end{aligned}$$

En la penúltima igualdad utilizamos la Propiedad 9.1 y la Propiedad 9.3. Luego (ii) queda demostrado.

(iii) Ahora demostraremos (iii). Observemos que por (ii)

$$E(Y^2 | \mathbf{X}) \geq E^2(Y | \mathbf{X})$$

y luego, en virtud del Teorema 9.1 tenemos

$$\infty > E(Y^2) = E(E(Y^2 | \mathbf{X})) \geq E(E^2(Y | \mathbf{X})),$$

demostrando (iii).

Propiedad 9.5 Sea Y una variable aleatoria discreta con esperanza finita y \mathbf{X} un vector aleatorio discreto de dimensión k . Luego si $g(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$, entonces para toda $t : \mathbb{R}^k \rightarrow \mathbb{R}$ medible tal que $Yt(\mathbf{X})$ tiene esperanza finita resulta

$$E[(Y - g(\mathbf{X}))t(\mathbf{X})] = 0.$$

Demostración. Sea $Z = h(\mathbf{X}, Y) = (Y - g(\mathbf{X}))t(\mathbf{X})$. Luego bastará demostrar que

$$E(Z) = 0.$$

Utilizando el Teorema 9.1 bastará demostrar que

$$E(Z|\mathbf{X}) = 0. \tag{9.7}$$

De acuerdo a la Propiedad 9.1, tenemos que

$$E(Z|\mathbf{X} = \mathbf{x}) = t(\mathbf{x})E((Y - g(\mathbf{X}))|\mathbf{X} = \mathbf{x}),$$

y por lo tanto

$$E(Z|\mathbf{X}) = t(\mathbf{X})E((Y - g(\mathbf{X}))|\mathbf{X}).$$

Luego para mostrar (9.7) bastará demostrar que

$$E(Y - g(\mathbf{X})|\mathbf{X}) = 0.$$

Pero esto es cierto ya que por Propiedades 9.3 y luego la Propiedad 9.2 se tiene

$$\begin{aligned} E(Y - g(\mathbf{X})|\mathbf{X}) &= E(Y|\mathbf{X}) - E(g(\mathbf{X})|\mathbf{X}) \\ &= E(Y|\mathbf{X}) - g(\mathbf{X}) \\ &= g(\mathbf{X}) - g(\mathbf{X}) \\ &= 0, \end{aligned}$$

y por lo tanto queda demostrada esta propiedad. \square

Propiedad 9.6 Sea Y una variable aleatoria discreta con varianza finita y \mathbf{X} un vector aleatorio discreto de dimensión k . Luego $\hat{Y} = g(\mathbf{X}) = E(Y|\mathbf{X})$ es el único predictor con menor error cuadrático medio en la clase de predictores

$$\mathcal{P} = \left\{ \hat{Y} = t(\mathbf{X}) : t \text{ medible, } \text{Var}(t(\mathbf{X})) < \infty \right\}.$$

Demostración. Se deja como ejercicio ver que \mathcal{P} es un espacio vectorial. Vamos a mostrar primero que $g(\mathbf{X}) \in \mathcal{P}$ o sea que

$$\text{Var}(g^2(\mathbf{X})) < \infty. \tag{9.8}$$

Pero esto resulta de Propiedad 9.4 (iii). Luego el resultado se obtiene del Teorema 8.1 y de la Propiedad 9.5. \square

9.2. Caso general

Vamos ahora dar una definición de $E(Y|\mathbf{X})$ para el caso de una variable Y cualesquiera, y un vector \mathbf{X} cualquiera de dimensión k . Ambos, Y y \mathbf{X} no tienen porque ser discretos ni absolutamente continuos

Definición 9.1 *La variable aleatoria esperanza de Y condicional \mathbf{X} se define por $E(Y|\mathbf{X}) = g(\mathbf{X})$, donde $g : \mathbb{R}^k \rightarrow \mathbb{R}$ es una función medible tal que*

$$E((Y - g(\mathbf{X}))t(\mathbf{X})) = 0 \quad (9.9)$$

para toda $t : \mathbb{R}^k \rightarrow \mathbb{R}$ medible tal que $Yt(\mathbf{X})$ tiene esperanza finita. Definiremos $E(Y|\mathbf{X} = \mathbf{x}) = g(\mathbf{x})$.

La Propiedad 9.5 demostrada anteriormente muestra que en el caso de Y y \mathbf{X} discretos esta definición coincide con la dada anteriormente, y por lo tanto en este caso siempre existe.

El siguiente teorema muestra que siempre existe una única variable aleatoria $g(\mathbf{X}) = E(Y|\mathbf{X})$ satisfaciendo (9.9).

Teorema 9.5 *Sea Y una variable aleatoria con esperanza finita y sea \mathbf{X} un vector aleatorio cualquiera de dimensión k . Luego*

(i) *Siempre existe una función medible $g : \mathbb{R}^k \rightarrow \mathbb{R}$ satisfaciendo (9.9).*

(ii) *Si g_1 y g_2 son dos funciones medibles satisfaciendo (9.9), entonces*

$$P(g_1(\mathbf{X}) = g_2(\mathbf{X})) = 1.$$

Demostración.

(i) No lo demostraremos en general en este curso. Más adelante haremos una demostración para el caso absolutamente continuo.

(ii) Sean g_1 y g_2 son dos funciones medibles satisfaciendo (9.9), entonces

$$E((Y - g_1(\mathbf{X}))t(\mathbf{X})) = 0 \quad (9.10)$$

y

$$E((Y - g_2(\mathbf{X}))t(\mathbf{X})) = 0 \quad (9.11)$$

para toda $t(\mathbf{X})$ tal que $Yt(\mathbf{X})$ tenga esperanza finita. Luego restando (9.11) de (9.10) se obtiene

$$E((g_2(\mathbf{X}) - g_1(\mathbf{X}))t(\mathbf{X})) = 0,$$

y tomando $t(\mathbf{X}) = \text{signo}(g_2(\mathbf{X}) - g_1(\mathbf{X}))$ resulta que

$$E(|t(\mathbf{X})Y|) = E(|t(\mathbf{X})||Y|) = E(|Y|) < \infty$$

y por lo tanto $t(\mathbf{X})Y$ tiene esperanza finita, y

$$\begin{aligned} 0 &= E((g_2(\mathbf{X}) - g_1(\mathbf{X}))t(\mathbf{X})) \\ &= E(\text{signo}(g_2(\mathbf{X}) - g_1(\mathbf{X}))(g_2(\mathbf{X}) - g_1(\mathbf{X}))) \\ &= E(|g_2(\mathbf{X}) - g_1(\mathbf{X})|). \end{aligned}$$

Esto implica que

$$\begin{aligned} P(|g_2(\mathbf{X}) - g_1(\mathbf{X})| = 0) &= P(g_2(\mathbf{X}) = g_1(\mathbf{X})) \\ &= 1. \quad \square \end{aligned}$$

Vamos ahora a demostrar que todas las propiedades de esperanza condicional que valían para el caso discreto también valen para la definición general.

Teorema 9.6 *Si Y tiene esperanza finita, entonces $E(E(Y|\mathbf{X})) = E(Y)$.*

Demostración. Apliquemos (9.9) con $t(\mathbf{X}) = 1$. Luego se tiene

$$\begin{aligned} 0 &= E(Y - g(\mathbf{X})) \\ &= E(Y) - E(g(\mathbf{X})) \\ &= E(Y) - E(E(Y|\mathbf{X})), \end{aligned}$$

y por lo tanto se cumple el Teorema 9.6. \square

Teorema 9.7 *Sean Y una variable aleatoria con esperanza finita y \mathbf{X} un vector aleatorio independientes. Entonces se tiene $E(Y|\mathbf{X}) = E(Y)$.*

Demostración. Veamos que poniendo $g(\mathbf{X}) = E(Y)$ se cumple (9.9). En efecto dado que $(Y - E(Y))$ y $t(\mathbf{X})$ son independientes se tiene

$$E((Y - E(Y))t(\mathbf{X})) = E(Y - E(Y))E(t(\mathbf{X})).$$

Luego como $E(Y - E(Y)) = E(Y) - E(Y) = 0$, el Teorema 9.7 queda demostrado. \square

Teorema 9.8 *Si $P(Y = c) = 1$, entonces, cualquiera sea el vector \mathbf{X} , se tiene $E(Y|\mathbf{X}) = c$.*

Demostración. Poniendo $g(\mathbf{X}) = c$, resulta inmediatamente (9.9). \square

Vamos ahora a probar las propiedades 9.1-9.4 para la definición general de $E(Y|\mathbf{X})$.

Propiedad 9.7 Sean \mathbf{X} un vector aleatorio de dimensión k e \mathbf{Y} un vector aleatorio de dimensión j , y sea $r : \mathbb{R}^k \rightarrow \mathbb{R}$ y $s : \mathbb{R}^j \rightarrow \mathbb{R}$. Entonces se tiene

$$E(r(\mathbf{X})s(\mathbf{Y})|\mathbf{X}) = r(\mathbf{X})E(s(\mathbf{Y})|\mathbf{X}).$$

Demostración. Vamos a probar que si ponemos $g(\mathbf{X}) = r(\mathbf{X})E(s(\mathbf{Y})|\mathbf{X})$, entonces (9.9) se cumple. En efecto

$$\begin{aligned} E((r(\mathbf{X})s(\mathbf{Y}) - g(\mathbf{X}))t(\mathbf{X})) &= E((r(\mathbf{X})s(\mathbf{Y}) - r(\mathbf{X})E(s(\mathbf{Y})|\mathbf{X}))t(\mathbf{X})) \\ &= E((s(\mathbf{Y}) - E(s(\mathbf{Y})|\mathbf{X}))m(\mathbf{X})), \end{aligned}$$

con $m(\mathbf{X}) = r(\mathbf{X})t(\mathbf{X})$. Luego por la definición de $E(s(\mathbf{Y})|\mathbf{X})$ obtenemos $E((s(\mathbf{Y}) - E(s(\mathbf{Y})|\mathbf{X}))m(\mathbf{X})) = 0$. Por lo tanto la propiedad queda demostrada. \square

Propiedad 9.8 Sea X un vector aleatorio de dimensión k y sea $r : \mathbb{R}^k \rightarrow \mathbb{R}$, una función medible. Luego $E(r(\mathbf{X})|\mathbf{X}) = r(\mathbf{X})$.

Demostración. Se obtiene de la Propiedad 9.7 tomando $s(\mathbf{Y}) = 1$. \square

Propiedad 9.9 Si Y_1 e Y_2 son variables aleatorias con esperanza finita, y \mathbf{X} es un vector aleatorio, entonces

$$E(c_1Y_1 + c_2Y_2|\mathbf{X}) = c_1E(Y_1|\mathbf{X}) + c_2E(Y_2|\mathbf{X}).$$

Demostración. Vamos a ver que se cumple (9.9) poniendo

$$g(\mathbf{X}) = c_1E(Y_1|\mathbf{X}) + c_2E(Y_2|\mathbf{X}).$$

En efecto si $Z = c_1Y_1 + c_2Y_2$ usando la linealidad de la esperanza y la definición de esperanza condicional se tiene

$$\begin{aligned} E((Z - g(\mathbf{X}))t(\mathbf{X})) &= E((c_1(Y_1 - E(Y_1|\mathbf{X})) + c_2(Y_2 - E(Y_2|\mathbf{X})))t(\mathbf{X})) \\ &= c_1E((Y_1 - E(Y_1|\mathbf{X}))t(\mathbf{X})) + c_2E((Y_2 - E(Y_2|\mathbf{X}))t(\mathbf{X})) \\ &= c_1 \cdot 0 + c_2 \cdot 0 \\ &= 0, \end{aligned}$$

y la propiedad queda demostrada. \square

La generalización de la Propiedad 9.5 usando la definición general de $E(Y|\mathbf{X})$ es obvia a partir de la definición.

Propiedad 9.10 Sea Y una variable aleatoria con varianza finita y \mathbf{X} un vector aleatorio de dimensión k . Luego $\hat{Y} = g(\mathbf{X}) = E(Y|\mathbf{X})$ es el único predictor con menor error cuadrático medio en la clase de predictores $\mathcal{P} = \{\hat{Y} = t(\mathbf{X}) : t \text{ medible, } \text{Var}(t(\mathbf{X})) < \infty\}$.

Demostración. Es totalmente similar a la Propiedad 9.6. \square

De acuerdo a esta propiedad $E(Y|\mathbf{X})$ es el predictor de Y óptimo basado en cualquier función medible (lineal o no lineal) de \mathbf{X} . Por esta razón lo denotaremos con $\hat{Y}_{O,NL}$.

9.3. Caso continuo

Supongamos ahora que tenemos dos vectores $\mathbf{X} = (X_1, \dots, X_k)$ e $\mathbf{Y} = (Y_1, \dots, Y_j)$ de dimensiones k y j respectivamente con distribución conjunta absolutamente continua y densidad $f_{\mathbf{X},\mathbf{Y}}$, y sea $h : \mathbb{R}^{k+j} \rightarrow \mathbb{R}$ una función medible. Definamos la densidad de \mathbf{Y} condicional $\mathbf{X} = \mathbf{x}$ por

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})}.$$

Es fácil ver que para cada \mathbf{x} fijo con $f_{\mathbf{X}}(\mathbf{x}) > 0$, la función $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ es una densidad para el vector \mathbf{Y} . Es decir se tendrá

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) dy_1 \dots dy_j = 1.$$

El siguiente teorema es una versión para el caso continuo del Teorema 9.4.

Teorema 9.9 Sea $Z = h(\mathbf{X}, \mathbf{Y})$ una variable con esperanza finita, luego se tiene que

$$\begin{aligned} E(Z|\mathbf{X} = \mathbf{x}) &= g(\mathbf{x}) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h(\mathbf{x}, \mathbf{y}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) dy_1 \dots dy_j. \end{aligned}$$

Demostración. Para facilitar la notación en la demostración, supondremos que tanto X como Y son variables aleatorias en vez de vectores. Pero excepto por la notación más complicada, la demostración para vectores es similar, ya que solamente se deben reemplazar las integrales simples por integrales múltiples.

De acuerdo a (9.9) será suficiente probar que

$$E((h(X, Y) - g(X))t(X)) = 0,$$

o equivalentemente

$$E((h(X, Y)t(X))) = E(g(X)t(X)). \quad (9.12)$$

Por un lado tenemos que

$$E((h(X, Y)t(X))) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)t(x)f_{XY}(x, y)dx dy. \quad (9.13)$$

Además se tiene que

$$\begin{aligned} E(g(X)t(X)) &= \int_{-\infty}^{\infty} g(x)t(x)f_X(x)dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(x, y)f_{Y|X}(y|x)dy \right] t(x)f_X(x)dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)t(x)f_{XY}(x, y)dx dy. \end{aligned} \quad (9.14)$$

Las ecuaciones (9.13) y (9.14) prueban (9.12). \square

Definición 9.2 Sean dos vectores aleatorios \mathbf{X} e \mathbf{Y} de dimensiones k y j respectivamente. Luego dado $B \in \beta^j$ (conjunto Boreliano de dimensión j), la probabilidad de que $\mathbf{Y} \in B$, condicional $\mathbf{X} = \mathbf{x}$ que se denotará con $P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X} = \mathbf{x})$ está dado por

$$P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X} = \mathbf{x}) = E(I_B(\mathbf{Y})|\mathbf{X} = \mathbf{x}),$$

donde I_B es la función indicadora del conjunto B . La probabilidad de que $\mathbf{Y} \in B$, condicional \mathbf{X} que se denotará por $P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X})$ está dado por

$$P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X}) = E(I_B(\mathbf{Y})|\mathbf{X}).$$

La justificación de esta definición está dada por el hecho que

$$P_{\mathbf{Y}}(B) = E(I_B(\mathbf{Y})).$$

En efecto $I_B(\mathbf{Y})$ toma valor 1 con probabilidad $P_{\mathbf{Y}}(B)$ y 0 con probabilidad $1 - P_{\mathbf{Y}}(B)$. Luego $E(I_B(\mathbf{Y})) = 1P_{\mathbf{Y}}(B) + 0(1 - P_{\mathbf{Y}}(B)) = P_{\mathbf{Y}}(B)$.

En el caso discreto, de acuerdo al Teorema 9.4, se tendrá

$$\begin{aligned} P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X} = \mathbf{x}) &= E(I_B(\mathbf{Y})|\mathbf{X} = \mathbf{x}) \\ &= \sum_{\mathbf{y} \in R_{\mathbf{Y}}} I_B(\mathbf{y})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \\ &= \sum_{\mathbf{y} \in R_{\mathbf{Y}} \cap B} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}). \end{aligned}$$

En el caso absolutamente continuo, de acuerdo al Teorema 9.9 se tiene

$$\begin{aligned} P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X} = \mathbf{x}) &= E(I_B(\mathbf{Y})|\mathbf{X} = \mathbf{x}) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} I_B(\mathbf{y}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) d\mathbf{y} \\ &= \int_B f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) d\mathbf{y}. \end{aligned}$$

Obsevamos que $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ actua como una verdadera densidad, en el sentido de que para calcular la probabilidad condicional de un evento B hay que integrar esta función sobre ese conjunto.

De acuerdo al Teorema 9.7 se tendrá

$$E(P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X})) = P_{\mathbf{Y}}(B).$$

Para el caso discreto y continuo podemos definir la función de distribución de \mathbf{Y} condicional $\mathbf{X} = \mathbf{x}$, la cual se denotará por $F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ y estarán definidas respectivamente por

$$\begin{aligned} F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) &= P_{\mathbf{Y}|\mathbf{X}}\left(\prod_{i=1}^j (-\infty, y_i] | \mathbf{X} = \mathbf{x}\right) \\ &= \sum_{\mathbf{z} \in R_{\mathbf{Y}} \cap \{z_1 \leq y_1\} \dots \cap \{z_j \leq y_j\}} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{z}|\mathbf{x}). \end{aligned}$$

$$\begin{aligned} F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) &= P_{\mathbf{Y}|\mathbf{X}}\left(\prod_{i=1}^j (-\infty, y_i] | \mathbf{X} = \mathbf{x}\right) \\ &= \int_{-\infty}^{y_j} \dots \int_{-\infty}^{y_1} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) d\mathbf{y}. \end{aligned}$$

Es fácil ver que para cada \mathbf{x} fijo $F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ es una verdadera función de distribución del vector \mathbf{Y} , en el sentido que cumple con las propiedades que caracterizan a una función de distribución.

9.4. Varianza condicional

Definición 9.3 Sea $\mathbf{X} = (X_1, \dots, X_k)$ un vector aleatorio e Y una variable aleatoria con varianza finita. Entonces la varianza de Y condicional $\mathbf{X} = \mathbf{x}$ se define como

$$\text{Var}(Y|\mathbf{X} = \mathbf{x}) = E((Y - E(Y|\mathbf{X} = \mathbf{x}))^2 | \mathbf{X} = \mathbf{x}),$$

y esta varianza puede considerarse como la varianza de variable \mathbf{X} una vez que se conoce que $\mathbf{X} = \mathbf{x}$. Denotemos por $q(\mathbf{x}) = \text{Var}(Y|\mathbf{X} = \mathbf{x})$, luego

$q : \mathbb{R}^k \rightarrow \mathbb{R}$. Llamaremos *varianza condicional de Y condicional \mathbf{X} a la variable aleatoria*

$$\text{Var}(Y|\mathbf{X}) = q(\mathbf{X}) = E((Y - E(Y|\mathbf{X}))^2|\mathbf{X}). \quad (9.15)$$

Desarrollando el cuadrado en (9.15) y utilizando la Propiedad 9.10 se obtiene

$$\begin{aligned} \text{Var}(Y|\mathbf{X}) &= E([Y^2 + E^2(Y|\mathbf{X}) - 2YE(Y|\mathbf{X})]|\mathbf{X}) \\ &= E(Y^2|\mathbf{X}) + E^2(Y|\mathbf{X}) - 2E(Y|\mathbf{X})E(Y|\mathbf{X}) \\ &= E(Y^2|\mathbf{X}) - E^2(Y|\mathbf{X}). \end{aligned}$$

El siguiente Teorema vincula la varianza condicional con el error cuadrático medio del predictor óptimo no lineal $\hat{Y}_{O,NL} = E(Y|\mathbf{X})$.

Teorema 9.10 *Supongamos que Y es una variable aleatoria con varianza finita, \mathbf{X} un vector aleatorio, y sea $\hat{Y}_{O,NL} = E(Y|\mathbf{X})$, el mejor predictor no lineal de Y basado en \mathbf{X} . Luego se tiene*

- (i) $ECM(\hat{Y}_{O,NL}, Y) = E(\text{Var}(Y|\mathbf{X}))$.
- (ii) $E(\text{Var}(Y|\mathbf{X})) \leq \text{Var}(Y)$.
- (iii) $E(\text{Var}(Y|\mathbf{X})) = \text{Var}(Y)$ si y sólo si $P(E(Y|\mathbf{X}) = E(Y)) = 1$.

Demostración. Aplicando el Teorema 9.7 y utilizando la definición (9.15) se tiene

$$\begin{aligned} ECM(\hat{Y}_{O,NL}, Y) &= E((Y - E(Y|\mathbf{X}))^2) \\ &= E(E((Y - E(Y|\mathbf{X}))^2|\mathbf{X})) \\ &= E(\text{Var}(Y|\mathbf{X})), \end{aligned}$$

y por lo tanto queda demostrado parte (i) del Teorema.

Como $\hat{Y}_{O,NL}$ es el predictor con menor error cuadrático medio en la clase de predictores $\mathcal{P} = \{\hat{Y} : \hat{Y} = t(\mathbf{X}), \text{Var}(t(\mathbf{X})) < \infty\}$, y como el predictor óptimo constante $\hat{Y}_{O,C} = E(Y) \in \mathcal{P}$, se tiene

$$\begin{aligned} E(\text{Var}(Y|\mathbf{X})) &= ECM(\hat{Y}_{O,NL}, Y) \\ &\leq ECM(\hat{Y}_{O,C}, Y) \\ &= E((Y - E(Y))^2) \\ &= \text{Var}(Y) \end{aligned}$$

y por un Teorema anterior la igualdad vale si y solo si $P(\hat{Y}_{O,NL} = \hat{Y}_{O,C}) = 1$. \square

Capítulo 10

Convergencia de Variables Aleatorias.

10.1. Convergencia de funciones.

Comenzaremos recordando algunos tipos de convergencia en espacios de funciones.

Definición 10.1 Sea $\{f_n\}_{n \in \mathbb{N}}$ una sucesión de funciones definidas sobre un conjunto Ω y que toman valores reales. Se dice que f_n converge puntualmente a otra función $f : \Omega \rightarrow \mathbb{R}$ si para todo $\omega \in \Omega$ y para todo $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ dependiendo de ε y de ω tal que si $n \geq n_0$ entonces $|f_n(\omega) - f(\omega)| < \varepsilon$.

En general n_0 depende de ε y ω , es decir $n_0 = n_0(\omega, \varepsilon)$. Cuando la elección de n_0 puede hacerse con independencia de ω , se tiene la siguiente noción de convergencia.

Definición 10.2 Sea $\{f_n\}_{n \in \mathbb{N}}$ una sucesión de funciones definidas sobre un conjunto Ω y que toma valores reales. Se dice que f_n converge uniformemente en Ω a otra función f si para todo $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces $|f_n(\omega) - f(\omega)| < \varepsilon$ para todo $\omega \in A$.

Observación. Es inmediato ver que si $\{f_n\}_{n \in \mathbb{N}}$ converge uniformemente en Ω entonces $\{f_n\}_{n \in \mathbb{N}}$ converge puntualmente. La recíproca es falsa. Por ejemplo, si definimos $f_n(\omega) = \omega^n$ para $\omega \in [0, 1]$ entonces la sucesión converge puntualmente a la función

$$f(\omega) = \begin{cases} 0 & \text{si } 0 \leq \omega < 1 \\ 1 & \text{si } \omega = 1 \end{cases}$$

para todo $\omega \in [0, 1]$ pero no converge uniformemente en $[0, 1]$.

Veremos ahora algunos tipos de convergencia para variables aleatorias que hacen uso de la estructura del espacio de probabilidades.

Existen varios tipos de convergencia, pero en este curso consideraremos sólo dos: la convergencia casi segura y la convergencia en probabilidad.

10.2. Convergencia casi segura y en probabilidad.

Consideremos un espacio de probabilidades (Ω, \mathcal{A}, P) . Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias definidas sobre este espacio y X otra variable aleatoria también definida sobre el mismo espacio.

Definición 10.3 Diremos que una sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ converge casi seguramente a otra variable aleatoria X ($X_n \rightarrow X$ c.s.) si

$$P(\{\omega : X_n(\omega) \rightarrow X(\omega)\}) = 1. \quad (10.1)$$

Observación. En teoría de la medida, este tipo de convergencia se denomina convergencia en casi todo punto y se la nota $X_n \rightarrow X$ p.p. o bien $X_n \rightarrow X$ c.t.p.

Definición 10.4 Diremos que una sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ converge en probabilidad a otra variable aleatoria X si para todo $\varepsilon > 0$ se tiene

$$\lim_{n \rightarrow +\infty} P(\{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0. \quad (10.2)$$

Notación. Si la sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ converge en probabilidad a la variable aleatoria X escribiremos $X_n \xrightarrow{P} X$.

Observaciones.

1. Equivalentemente, (10.2) puede reescribirse como

$$\lim_{n \rightarrow +\infty} P(\{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\}) = 1.$$

2. La convergencia en probabilidad significa que fijado $\varepsilon > 0$ hay un subconjunto de Ω de probabilidad tan cercana a uno como se quiera en el que la distancia entre X_n y X se puede hacer menor que ε con tal de tomar n suficientemente grande.
3. En teoría de la medida la convergencia en probabilidad se denomina convergencia en medida.

Teorema 10.1 Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias definidas sobre un espacio de probabilidad (Ω, \mathcal{A}, P) y X otra variable aleatoria definida sobre el mismo espacio. Son equivalentes:

(i) $X_n \xrightarrow{P} X$.

(ii) Para todo $\varepsilon > 0$ y todo $\delta > 0$ existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces

$$P(|X_n - X| \geq \varepsilon) \leq \delta.$$

(iii) Para todo $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces

$$P(|X_n - X| \geq \varepsilon) \leq \varepsilon.$$

Demostración. (ii) es equivalente a (i) como consecuencia directa de la definición de convergencia en probabilidad. La equivalencia entre (ii) y (iii) se deja como ejercicio. \square

El siguiente teorema establece que la convergencia casi segura (10.1) implica la convergencia en probabilidad (10.2).

Teorema 10.2 Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias definidas sobre un espacio de probabilidad (Ω, \mathcal{A}, P) y X otra variable aleatoria definida sobre el mismo espacio. Entonces

(i) La sucesión X_n converge casi seguramente a X sii

$$\lim_{m \rightarrow \infty} P\left(\bigcup_{n=m}^{\infty} \{|X_n - X| \geq \varepsilon\}\right) = 0. \quad (10.3)$$

(ii) Si X_n converge casi seguramente a X entonces X_n converge en probabilidad a la variable aleatoria X .

Demostración.

(i) Llamemos A al conjunto de los puntos ω de Ω donde $X_n(\omega) \rightarrow X(\omega)$.
Luego

$$A = \{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}.$$

Decir que $\omega \in A$ es equivalente a decir que para todo $\varepsilon > 0$ existe $m \in \mathbb{N}$ tal que para todo $n \geq m$ se tiene $|X_n(\omega) - X(\omega)| < \varepsilon$, m dependerá de ω . Entonces, si para cada $\varepsilon > 0$ definimos

$$B_{n,\varepsilon} = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\}.$$

el conjunto A resulta

$$A = \bigcap_{\varepsilon > 0} \left(\bigcup_{m=1}^{\infty} \bigcap_{n \geq m} B_{n,\varepsilon} \right).$$

Sabemos que la convergencia casi segura se define por $P(A) = 1$ o equivalentemente por $P(A^c) = 0$. Pero para poder usar propiedades de probabilidad en el cálculo de $P(A)$ nos conviene tener escrito al conjunto A como una numerable cantidad de uniones e intersecciones de eventos. Por ello, como basta elegir ε tan chico como se quiera, nos podemos limitar a tomar $\varepsilon = 1/k$. Luego también tenemos

$$A = \bigcap_{k=1}^{\infty} \left(\bigcup_{m=1}^{\infty} \bigcap_{n \geq m} B_{n, \frac{1}{k}} \right).$$

Observemos que

$$A^c = \bigcup_{k=1}^{\infty} \left(\bigcap_{m=1}^{\infty} \bigcup_{n \geq m} B_{n, \frac{1}{k}}^c \right).$$

Luego, como A^c es una unión numerable, $P(A^c) = 0$ si y sólo si para todo $k \in \mathbb{N}$ se tiene

$$P \left(\bigcap_{m=1}^{\infty} \bigcup_{n \geq m} B_{n, \frac{1}{k}}^c \right) = 0.$$

En la notación del Capítulo 1 (Definición 1.3, página 15), esto es el límite inferior de los conjuntos $(B_{n, \frac{1}{k}}^c)_{n \in \mathbb{N}}$. Como $B_{n, \varepsilon}^c$ es ceciente con ε , esto es equivalente a que para todo $\varepsilon > 0$

$$P \left(\bigcap_{m=1}^{\infty} \bigcup_{n \geq m} B_{n, \varepsilon}^c \right) = 0. \quad (10.4)$$

Definamos

$$C_{m, \varepsilon} = \bigcup_{n \geq m} B_{n, \varepsilon}^c.$$

Claramente, para todo $\varepsilon > 0$ la sucesión $\{C_{m, \varepsilon}\}_{m \geq 1}$ es creciente (no necesariamente estrictamente creciente), de manera que

$$P \left(\bigcap_{m=1}^{\infty} \bigcup_{n \geq m} B_{n, \varepsilon}^c \right) = P \left(\bigcap_{m=1}^{\infty} C_{m, \varepsilon} \right) = \lim_{m \rightarrow \infty} P(C_{m, \varepsilon}).$$

Luego se tendrá que (10.4) es equivalente a

$$\lim_{m \rightarrow \infty} P(C_{m, \varepsilon}) = 0,$$

es decir,

$$\lim_{m \rightarrow \infty} P \left(\bigcup_{n \geq m} B_{n, \varepsilon}^c \right) = 0.$$

Pero como

$$B_{n,\varepsilon}^c = \{|X_n - X| \geq \varepsilon\},$$

(i) queda demostrado.

(ii) Supongamos que $X_n \rightarrow X$ c.s. Luego se cumple (10.3) y como

$$\{|X_m - X| \geq \varepsilon\} \subset \bigcup_{n=m}^{\infty} \{|X_n - X| \geq \varepsilon\},$$

por la monotonía de la probabilidad resulta

$$\lim_{m \rightarrow \infty} P(\{|X_m - X| \geq \varepsilon\}) = 0.$$

Por lo tanto $X_n \xrightarrow{P} 0$. \square

Observación. Notemos que en esta demostración hemos probado que

$$\begin{aligned} A &= \{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\} \\ &= \bigcap_{k=1}^{\infty} \left(\bigcup_{m=1}^{\infty} \bigcap_{n \geq m} B_{n, \frac{1}{k}} \right) \\ &= \bigcap_{k=1}^{\infty} \liminf_{n \rightarrow \infty} B_{n, \frac{1}{k}} \end{aligned}$$

o, equivalentemente

$$\begin{aligned} A^c &= \{\omega \in \Omega : X_n(\omega) \not\rightarrow X(\omega)\} \\ &= \bigcup_{k=1}^{\infty} \left(\bigcap_{m=1}^{\infty} \bigcup_{n \geq m} B_{n, \frac{1}{k}}^c \right) \\ &= \bigcup_{k=1}^{\infty} \limsup_{n \rightarrow \infty} B_{n, \frac{1}{k}}. \end{aligned}$$

Veremos que la recíproca de la parte (ii) de este teorema es falsa. Incluso puede ocurrir que exista convergencia en probabilidad, pero que el conjunto de los puntos donde haya convergencia puntual sea vacío.

10.3. Preservación de la convergencia por funciones continuas.

Los siguientes dos teoremas muestran que las funciones continuas preservan los dos tipos de convergencia que hemos definido: convergencia en probabilidad y convergencia casi segura.

Teorema 10.3 Sea $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ continua y supongamos que las sucesiones de variables aleatorias $(X_n)_{n \geq 1}$, $(Y_n)_{n \geq 1}$ convergen casi seguramente a las variables aleatorias X e Y . Entonces $(g(X_n, Y_n))_{n \geq 1}$ converge casi seguramente a la variable aleatoria $g(X, Y)$.

Observación. La propiedad vale en general para $g : \mathbb{R}^k \rightarrow \mathbb{R}$ continua. Si $(X_n^{(j)})_{n \geq 1} \rightarrow X^{(j)}$ c.s. para $j = 1, 2, \dots, k$ entonces

$$g(X_n^{(1)}, X_n^{(2)}, \dots, X_n^{(k)}) \rightarrow g(X^{(1)}, X^{(2)}, \dots, X^{(k)}) \text{ c.s.}$$

Demostración. Sean $A = \{\omega : X_n(\omega) \rightarrow X(\omega)\}$ y $B = \{\omega : Y_n(\omega) \rightarrow Y(\omega)\}$. Como $P(A) = P(B) = 1$, también se tendrá $P(A \cap B) = 1$. En efecto

$$0 \leq P((A \cap B)^c) = P(A^c \cup B^c) \leq P(A^c) + P(B^c) = 0.$$

Ahora si $\omega \in A \cap B$ entonces $X_n(\omega) \rightarrow X(\omega)$ e $Y_n(\omega) \rightarrow Y(\omega)$. Luego, por la continuidad de g se tiene

$$g(X_n(\omega), Y_n(\omega)) \rightarrow g(X(\omega), Y(\omega)).$$

Por lo tanto

$$A \cap B \subset \{\omega : g(X_n(\omega), Y_n(\omega)) \rightarrow g(X(\omega), Y(\omega))\},$$

y en consecuencia como

$$1 = P(A \cap B) \leq P(\{\omega : g(X_n(\omega), Y_n(\omega)) \rightarrow g(X(\omega), Y(\omega))\}) \leq 1,$$

el Teorema queda demostrado. \square

Teorema 10.4 (i) Si $Y_n \rightarrow Y$ c.s. y $X_n \rightarrow X$ c.s. entonces $X_n + Y_n \rightarrow X + Y$ c.s.

(ii) Si $Y_n \rightarrow Y$ c.s. y $X_n \rightarrow X$ c.s. entonces $X_n Y_n \rightarrow XY$ c.s.

(iii) Si $Y_n \rightarrow Y$ c.s. con $P(Y = 0) = 0$ y $X_n \rightarrow X$ c.s. entonces $\frac{X_n}{Y_n} \rightarrow \frac{X}{Y}$ c.s.

Demostración.

(i) y (ii) resultan de que las funciones $g(x, y) = x + y$ y $g(x, y) = xy$ son continuas y (iii) del hecho que $g(x, y) = x/y$ es continua si $y \neq 0$. \square

Para demostrar una propiedad similar para la convergencia en probabilidad necesitamos algunos resultados previos. Comenzamos probando que toda variable aleatoria es "acotada en probabilidad". Esto significa que X está dentro de un compacto, con probabilidad tan cercana a uno como se quiera.

Teorema 10.5 Sea X una variable aleatoria. Dado $\varepsilon > 0$ existe K tal que

$$P(|X| \geq K) < \varepsilon.$$

Demostración.

Consideremos la sucesión de conjuntos

$$A_n = \{|X| \geq n\}.$$

Esta sucesión es monótona decreciente, es decir, $A_{n+1} \subset A_n$ y además $\bigcap_{n=1}^{\infty} A_n = \emptyset$. Entonces

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

Luego, dado $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que $P(A_{n_0}) < \varepsilon$, es decir

$$P(A_{n_0}) = P(\{|X| \geq n_0\}) < \varepsilon.$$

Luego el Teorema es cierto tomando $K = n_0$. \square

Probaremos ahora un resultado más fuerte: sucesiones de variables que convergen en probabilidad están acotadas en probabilidad uniformemente.

Teorema 10.6 Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias que converge en probabilidad a la variable aleatoria X . Entonces dado $\varepsilon > 0$ existe K tal que $P(|X| \geq K) < \varepsilon$ y tal que para todo n

$$P(|X_n| \geq K) < \varepsilon.$$

Demostración.

En primer lugar podemos hallar, de acuerdo al Teorema 10.5, K_0 de forma tal que

$$P(|X| \geq K_0) < \frac{\varepsilon}{2}.$$

Teniendo en cuenta que

$$|X_n| \leq |X_n - X| + |X| \tag{10.5}$$

se prueba fácilmente que

$$\{|X_n| \geq K_0 + 1\} \subset \{|X_n - X| \geq 1\} \cup \{|X| \geq K_0\}. \tag{10.6}$$

En efecto, supongamos que

$$\omega \notin \{|X_n - X| \geq 1\} \cup \{|X| \geq K_0\}.$$

Luego $|X_n(\omega) - X(\omega)| < 1$ y $|X(\omega)| < K_0$ y por lo tanto por (10.5) resulta $|X_n(\omega)| < K_0 + 1$.

Debido a que $X_n \xrightarrow{P} X$ en probabilidad podemos encontrar n_0 tal que si $n \geq n_0$

$$P(|X_n - X| \geq 1) < \frac{\varepsilon}{2}.$$

Tomando probabilidades en ambos miembros de (10.6) obtenemos

$$\begin{aligned} P(\{|X_n| \geq K_0 + 1\}) &\leq P(\{|X_n - X| \geq 1\}) + P(\{|X| \geq K_0\}) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

para todo $n \geq n_0$. Además por el Teorema 10.5, para cada i tal que $1 \leq i \leq n_0$ podemos encontrar K_i tal que $P(|X_i| \geq K_i) \leq \varepsilon$. Luego tomando

$$K = \max \left\{ \max_{1 \leq i \leq n_0} \{K_i\}, K_0 + 1 \right\},$$

se obtiene la tesis. \square

Ahora estamos en condiciones de probar la propiedad de que las funciones continuas conservan la convergencia en probabilidad.

Teorema 10.7 *Sea $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ continua y supongamos que las sucesiones $(X_n)_{n \geq 1}$ e $(Y_n)_{n \geq 1}$ convergen en probabilidad a las variables aleatorias X e Y , respectivamente. Entonces $(g(X_n, Y_n))_{n \geq 1}$ converge en probabilidad a la variable aleatoria $g(X, Y)$.*

Observación. Vale la misma observación hecha para el caso de la convergencia casi segura en cuanto a que este teorema es válido para funciones continuas definidas en \mathbb{R}^k y vectores aleatorios k dimensionales.

Demostración.

Queremos probar que dado $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$

$$P(|g(X_n, Y_n) - g(X, Y)| \geq \varepsilon) < \varepsilon. \quad (10.7)$$

pues por el Teorema 10.1 esto garantiza la convergencia en probabilidad.

De acuerdo al Teorema 10.5 podemos hallar un K tal que simultáneamente

$$\begin{aligned} P(|X_n| \geq K) &< \frac{\varepsilon}{6} \quad \forall n \\ P(|X| \geq K) &< \frac{\varepsilon}{6} \\ P(|Y_n| \geq K) &< \frac{\varepsilon}{6} \quad \forall n \\ P(|Y| \geq K) &< \frac{\varepsilon}{6}. \end{aligned}$$

Esto puede lograrse considerando primero un K_1 que cumpla con las dos primeras desigualdades, después un K_2 que cumpla con las siguientes dos y tomando $K = \max\{K_1, K_2\}$.

Sea

$$C = [-K, K] \times [-K, K].$$

Como g es continua y C es compacto entonces g resulta uniformemente continua en C . Luego existe $\delta > 0$ tal que si $|x - x'| < \delta$, $|y - y'| < \delta$ y $\max\{|x|, |x'|, |y|, |y'|\} \leq K$ entonces

$$|g(x, y) - g(x', y')| < \varepsilon. \quad (10.8)$$

Por la convergencia en probabilidad existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces

$$P(|X_n - X| \geq \delta) < \frac{\varepsilon}{6} \quad (10.9)$$

$$P(|Y_n - Y| \geq \delta) < \frac{\varepsilon}{6}. \quad (10.10)$$

Esto se logra considerando un valor n_1 para la sucesión $(X_n)_{n \geq 1}$, un valor n_2 para la sucesión $(Y_n)_{n \geq 1}$ y luego tomando $n_0 = \max\{n_1, n_2\}$.

Ahora definimos los conjuntos

$$A_{1n} = \{|X_n - X| \geq \delta\}$$

$$A_{2n} = \{|Y_n - Y| \geq \delta\}$$

$$A_{3n} = \{|X_n| \geq K\}$$

$$A_{4n} = \{|Y_n| \geq K\}$$

$$A_{5n} = \{|X| \geq K\}$$

$$A_{6n} = \{|Y| \geq K\}.$$

Si bien A_{5n} , A_{6n} no dependen de n , usamos la notación por conveniencia. Vamos a mostrar que si llamamos

$$B_n = \bigcup_{i=1}^6 A_{in},$$

entonces

$$\{|g(X_n, Y_n) - g(X, Y)| \geq \varepsilon\} \subset B_n.$$

Para esto debemos mostrar que para todo $n \geq n_0$ en B_n^c se tiene

$$|g(X_n, Y_n) - g(X, Y)| < \varepsilon. \quad (10.11)$$

En efecto, como

$$B_n^c = \left(\bigcup_{i=1}^6 A_{in}\right)^c = \bigcap_{i=1}^6 A_{in}^c,$$

resulta que cuando B_n^c es cierto X_n, X, Y_n, Y están en el compacto C y además $|X_n - X| \leq \delta$ e $|Y_n - Y| \leq \delta$. Luego por (10.8) resulta (10.11). Luego para todo $n \geq n_0$

$$P(\{|g(X_n, Y_n) - g(X, Y)| \geq \varepsilon\}) \leq P(B_n) \leq \sum_{i=1}^6 P(A_{in}) < 6\frac{\varepsilon}{6} = \varepsilon,$$

y el Teorema queda demostrado. \square

Análogamente a lo observado para la convergencia casi segura se tienen los siguientes corolarios.

Teorema 10.8 (i) Si $Y_n \xrightarrow{P} Y$ y $X_n \xrightarrow{P} X$ entonces $X_n + Y_n \xrightarrow{P} X + Y$.

(ii) Si $Y_n \xrightarrow{P} Y$ y $X_n \xrightarrow{P} X$ c.s entonces $X_n Y_n \xrightarrow{P} XY$.

(iii) Si $Y_n \xrightarrow{P} Y$ con $P(Y = 0) = 0$ y $X_n \xrightarrow{P} X$ entonces $\frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{Y}$.

Demostración.

Similar a la demostración del Teorema 10.4. \square

10.4. Ley débil de los grandes números.

Teorema 10.9 (Desigualdad de Markov) Sea X una variable aleatoria y g una función par, no negativa y no decreciente en el módulo, esto es si $|x| > |y|$ entonces $g(x) \geq g(y)$. Supongamos además que $g(X)$ tiene esperanza finita, es decir que $E(g(X)) < \infty$. Entonces si $\varepsilon > 0$ es tal que $g(\varepsilon) > 0$, vale que

$$P(|X| \geq \varepsilon) \leq \frac{E(g(X))}{g(\varepsilon)}.$$

Demostración.

Consideremos el conjunto $A = \{\omega : |X(\omega)| \geq \varepsilon\}$. Entoces $\{A, A^c\}$ es una partición del espacio muestral Ω . Luego $I_A(x) + I_{A^c}(x) = 1$, y como todas las variables son no negativas y $g(x)$ es no decreciente en $|x|$, tenemos

$$\begin{aligned} g(X) &= g(X) I_A(X) + g(X) I_{A^c}(X) \\ &\geq g(X) I_A(X) \\ &\geq g(\varepsilon) I_A(X). \end{aligned}$$

Luego tomando esperanza obtenemos

$$E(g(X)) \geq g(\varepsilon) E(I_A) = g(\varepsilon) P(\{|X| \geq \varepsilon\}).$$

De esta desigualdad se obtiene inmediatamente el resultado buscado. \square

En particular tomando $g(x) = x^2$ se obtiene la siguiente versión de la *Desigualdad de Tchebichev*

$$P(\{|X| \geq \varepsilon\}) \leq \frac{E(X^2)}{\varepsilon^2}.$$

Por otro lado si consideramos la variable aleatoria $X - E(X)$ obtenemos la versión (clásica) de la desigualdad de Tchebichev

$$P(\{|X - E(X)| \geq \varepsilon\}) \leq \frac{E\left([X - E(X)]^2\right)}{\varepsilon^2} = \frac{\text{Var}(X)}{\varepsilon^2}.$$

Tomando complementos esta desigualdad puede escribirse como

$$P(\{|X - E(X)| < \varepsilon\}) \geq 1 - \frac{\text{Var}(X)}{\varepsilon^2}.$$

Luego si la $\text{Var}(X)$ es pequeña (o sea hay poca dispersión), la probabilidad de que la variable X tome valores en el intervalo $(E(X) - \varepsilon, E(X) + \varepsilon)$ será grande.

Ahora estamos en condiciones de estudiar la ley de los grandes números en sus dos versiones: débil y fuerte. La importancia de estas leyes, es que permite dar fundamento matemático a la argumentación heurística que interpreta la esperanza de una variable aleatoria como el valor al cual tiende el promedio de varias realizaciones de la variable correspondientes a la repetición de experimentos independientes. También permite fundamentar la noción heurística de la probabilidad de un evento como el valor límite de las frecuencias relativas con que ocurre el evento cuando se repiten muchos experimentos independientes. La ley débil expresa estos resultados en términos de convergencia en probabilidad y la ley fuerte en términos de convergencia casi segura.

Teorema 10.10 (Ley débil de los grandes números) *Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias no correlacionadas, es decir $\text{Cov}(X_i, X_j) = 0$ si $i \neq j$, tal que $E(X_i) = \mu_i$ y $\text{Var}(X_i) = \sigma_i^2$ para cada $i = 1, 2, \dots$. Consideramos la sucesión de variables aleatorias $(\bar{X}_n)_{n \geq 1}$ donde \bar{X}_n es el promedio de las primeras n variables. Luego*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

y sea $\bar{\mu}_n = E(\bar{X}_n)$ dada por

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n \mu_i.$$

Entonces si

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \right) = 0, \quad (10.12)$$

se tiene

$$\bar{X}_n - \bar{\mu}_n \xrightarrow{P} 0.$$

Demostración.

Se tiene que

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2,$$

y por Tchebichev

$$P(|\bar{X}_n - \bar{\mu}_n| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{1}{\varepsilon^2 n^2} \sum_{i=1}^n \sigma_i^2.$$

Tomando límite resulta que

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \bar{\mu}_n| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 = 0$$

y luego el Teorema queda demostrado. \square

Observaciones.

1. Si $(X_n)_{n \geq 1}$ es una sucesión de variables aleatorias independientes, entonces las variables X_n son no correlacionadas y el Teorema puede aplicarse.
2. Una condición suficiente para que se cumpla (10.12) es que $\{\sigma_i^2\}$ sea una sucesión acotada. En efecto, si $\sigma_i^2 \leq K$ para todo i , se obtiene

$$\frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \leq \frac{Kn}{n^2} = \frac{K}{n} \rightarrow 0.$$

En particular, esta condición se cumple si todas las variables tienen igual varianza.

3. Si todas las variables tienen igual media, digamos $\mu_i = \mu$, se tiene que $\bar{\mu}_n = \mu$, y entonces $\bar{X}_n - \mu \xrightarrow{P} 0$ o, lo que es equivalente,

$$\bar{X}_n \xrightarrow{P} \mu.$$

4. En particular si $(X_n)_{n \geq 1}$ es una sucesión de variables no correlacionadas igualmente distribuidas con $E(X_n) = \mu$ y $\text{Var}(X_n) = \sigma^2$, se tendrá $\bar{X}_n \xrightarrow{P} \mu$.

5. Veremos ahora como esta ley débil permite fundamentar el concepto de probabilidad de un evento. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y A un evento. Supongamos que realizamos n experimentos independientes y definimos

$$X_i(\omega) = \begin{cases} 1 & \text{si en el experimento } i, \omega \in A \\ 0 & \text{si en el experimento } i, \omega \notin A. \end{cases}$$

Definamos

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Se tiene

$$E(X_i) = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A),$$

y como $X_i^2 = X_i$

$$\begin{aligned} \text{Var}(X_i) &= E(X_i^2) - E(X_i)^2 \\ &= E(X_i) - E(X_i)^2 \\ &= P(A) - P(A)^2 \\ &= P(A)(1 - P(A)). \end{aligned}$$

Luego, como además las variables X_i son independientes, de acuerdo a la ley débil de los grandes números se tendrá

$$\bar{X}_n \xrightarrow{P} E(X_i) = P(A). \quad (10.13)$$

Obsérvese que \bar{X}_n es la frecuencia relativa de ocurrencia del evento A en n repeticiones independientes del experimento. Entonces (10.13) puede interpretarse como que la frecuencia relativa de ocurrencia del evento A tiende (en probabilidad) a su probabilidad.

10.5. Ley fuerte de los grandes números.

Para probar la ley fuerte de los grandes números necesitaremos algunos teoremas previos.

Teorema 10.11 (Desigualdad de Kolmogorov) Sean X_1, \dots, X_n variables independientes con $E(X_i) = 0$. Supongamos que $\sigma_i^2 = \text{Var}(X_i) < \infty$ y consideremos las sumas parciales $S_i = \sum_{j=1}^i X_j$. Entonces

$$P\left(\max_{1 \leq i \leq n} |S_i| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{i=1}^n \sigma_i^2. \quad (10.14)$$

Observación. Vamos a mostrar que la desigualdad de Kolmogorov es un refinamiento de la desigualdad de Tchebichev. Para ver esto, apliquemos la desigualdad de Tchebichev a la variable aleatoria S_n . Obtenemos

$$P(|S_n| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \text{Var}(S_n) = \frac{1}{\varepsilon^2} \sum_{i=1}^n \sigma_i^2. \quad (10.15)$$

Observemos que $|S_n| \leq \max_{1 \leq i \leq n} |S_i|$ de manera que

$$\{|S_n| \geq \varepsilon\} \subset \left\{ \max_{1 \leq i \leq n} |S_i| \geq \varepsilon \right\},$$

y por lo tanto

$$P(\{|S_n| \geq \varepsilon\}) \leq P\left(\max_{1 \leq i \leq n} |S_i| \geq \varepsilon\right).$$

Luego resulta que (10.14) implica (10.15).

Demostración. Sea

$$A = \left\{ \max_{1 \leq i \leq n} |S_i| \geq \varepsilon \right\},$$

y consideremos para cada i los conjuntos

$$A_i = \{|S_1| < \varepsilon, |S_2| < \varepsilon, \dots, |S_{i-1}| < \varepsilon, |S_i| \geq \varepsilon\}.$$

Estos eventos son disjuntos dos a dos y forman una partición de A . Luego

$$A = \bigcup_{i=1}^n A_i,$$

y por lo tanto se deduce que

$$I_A = \sum_{i=1}^n I_{A_i}.$$

Luego como $S_n^2 I_{A^c} \geq 0$ se deduce que

$$S_n^2 = S_n^2 I_A + S_n^2 I_{A^c} \geq S_n^2 I_A = S_n^2 \sum_{i=1}^n I_{A_i}.$$

Tomando esperanza en ambos miembros resulta

$$E(S_n^2) \geq \sum_{i=1}^n E(S_n^2 I_{A_i}). \quad (10.16)$$

Para cada término $S_n^2 I_{A_i}$ resulta

$$S_n^2 I_{A_i} = (S_i + T_i)^2 I_{A_i} = S_i^2 I_{A_i} + T_i^2 I_{A_i} + 2S_i T_i I_{A_i}, \quad (10.17)$$

donde

$$T_i = \sum_{j=i+1}^n X_j.$$

Ahora probaremos que $E(S_i T_i I_{A_i}) = 0$. Por un lado observamos que S_i depende sólo de X_1, \dots, X_i y lo mismo ocurre con I_{A_i} . Como T_i depende sólo de X_{i+1}, \dots, X_n , resulta que $S_i I_{A_i}$ es independiente de T_i . Luego como $E(T_i) = 0$ se obtiene

$$E(S_i T_i I_{A_i}) = E([S_i I_{A_i}] T_i) = E(S_i I_{A_i}) E(T_i) = 0. \quad (10.18)$$

Tomando esperanza en (10.17) y teniendo en cuenta (10.18) y el hecho de que en A_i se tiene $|S_i| \geq \varepsilon$

$$\begin{aligned} E(S_n^2 I_{A_i}) &= E(S_i^2 I_{A_i}) + E(T_i^2 I_{A_i}) \\ &\geq E(S_i^2 I_{A_i}) \\ &\geq \varepsilon E(I_{A_i}) \\ &= \varepsilon P(A_i). \end{aligned}$$

Luego por (10.16) resulta

$$\begin{aligned} E(S_n^2) &\geq \sum_{i=1}^n E(S_n^2 I_{A_i}) \\ &\geq \varepsilon^2 \sum_{i=1}^n P(A_i) \\ &= \varepsilon^2 P(A), \end{aligned}$$

o sea

$$\begin{aligned} P(A) &\leq \frac{E(S_n^2)}{\varepsilon^2} \\ &= \frac{1}{\varepsilon^2} \sum_{i=1}^n \sigma_i^2. \quad \square \end{aligned}$$

Para probar la ley fuerte de los grandes números necesitamos también el siguiente teorema.

Teorema 10.12 *Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias. Una condición suficiente para que*

$$X_n \rightarrow X \text{ c.s.}$$

es que para todo $\varepsilon > 0$ exista una sucesión creciente de enteros positivos $r_1 < r_2 < \dots < r_n \dots$ que puede depender de ε tal que

$$\sum_{i=1}^{\infty} P\left(\bigcup_{n=r_i}^{r_{i+1}-1} B_{n\varepsilon}^c\right) < \infty, \quad (10.19)$$

donde $B_{n\varepsilon} = \{|X_n - X| < \varepsilon\}$.

Demostración. Recordemos el resultado ya probado en el Teorema 10.2 que establece que

$$X_n \rightarrow X \text{ c.s.}$$

si y sólo si

$$\lim_{m \rightarrow \infty} P \left(\bigcup_{n=m}^{\infty} B_{n\varepsilon}^c \right) = 0. \quad (10.20)$$

Supongamos que se cumple (10.19). Veremos que entonces se cumple (10.20). Sea $\varepsilon > 0$, entonces (10.19) implica que existe i_0 tal que

$$\sum_{i=i_0}^{\infty} P \left(\bigcup_{n=r_i}^{r_{i+1}-1} B_{n\varepsilon}^c \right) < \varepsilon.$$

Pero entonces

$$P \left(\bigcup_{n=r_{i_0}}^{\infty} B_{n\varepsilon}^c \right) = P \left(\bigcup_{i=i_0}^{\infty} \bigcup_{n=r_i}^{r_{i+1}-1} B_{n\varepsilon}^c \right) \leq \sum_{i=i_0}^{\infty} P \left(\bigcup_{n=r_i}^{r_{i+1}-1} B_{n\varepsilon}^c \right) < \varepsilon.$$

Esto implica que (10.20) se cumple. \square

Teorema 10.13 (Ley fuerte de los grandes números) *Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias independientes tal que $E(X_i) = \mu_i$ y $\text{Var}(X_i) = \sigma_i^2$ para cada $i \in \mathbb{N}$. Consideremos la sucesión de variables aleatorias $(\bar{X}_n)_{n \geq 1}$ definida por*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

y sus respectivas medias

$$\bar{\mu}_n = E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n \mu_i.$$

Entonces si

$$\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < \infty, \quad (10.21)$$

se tiene

$$\bar{X}_n - \bar{\mu}_n \rightarrow 0 \text{ c.s.}$$

Demostración. Basta probar el teorema suponiendo que para todo i , $\mu_i = 0$. Para ver esto, supongamos que el teorema fuera válido cuando para todo i , $\mu_i = 0$ y deduzcamos de esto el caso general, esto es, cuando para cada i la $E(X_i) = \mu_i$ arbitraria. Para ello, consideremos nuevas variables $Y_i = X_i - \mu_i$. Entonces $E(Y_i) = 0$ y $\text{Var}(Y_i) = \text{Var}(X_i) = \sigma_i^2$. Las variables Y_i

son independientes y luego se cumple $\bar{Y}_n \rightarrow 0$ c.s. Pero como $\bar{Y}_n = \bar{X}_n - \bar{\mu}_n$, resulta también $\bar{X}_n - \bar{\mu}_n \rightarrow 0$ c.s. Luego para demostrar el teorema podemos suponer que $\mu_i = 0$ para todo i .

Usaremos el Teorema 10.12, tomando $r_i = 2^{i-1}$. Luego si llamamos

$$\lambda_i = P \left(\bigcup_{n=2^{i-1}}^{2^i-1} B_{n\varepsilon}^c \right),$$

bastará demostrar que

$$\sum_{i=1}^{\infty} \lambda_i < \infty.$$

Si llamamos $S_n = \sum_{i=1}^n X_i$ tenemos que $\bar{X}_n = S_n/n$. Luego

$$\begin{aligned} \lambda_i &= P \left(\bigcup_{n=2^{i-1}}^{2^i-1} B_{n\varepsilon}^c \right) \\ &= P \left(\bigcup_{n=2^{i-1}}^{2^i-1} \{|\bar{X}_n| \geq \varepsilon\} \right) \\ &= P \left(\bigcup_{n=2^{i-1}}^{2^i-1} \{|S_n| \geq n\varepsilon\} \right) \\ &\leq P \left(\bigcup_{n=2^{i-1}}^{2^i-1} \{|S_n| \geq 2^{i-1}\varepsilon\} \right) \\ &\leq P \left(\bigcup_{n=1}^{2^i-1} \{|S_n| \geq 2^{i-1}\varepsilon\} \right). \end{aligned} \tag{10.22}$$

Usando la Desigualdad de Kolmogorov (Teorema 10.11) resulta

$$\begin{aligned} P \left(\bigcup_{n=1}^{2^i-1} \{|S_n| \geq 2^{i-1}\varepsilon\} \right) &= P \left(\max_{1 \leq n \leq 2^i-1} |S_n| \geq \varepsilon 2^{i-1} \right) \\ &\leq \frac{1}{4^{i-1}\varepsilon^2} \sum_{j=1}^{2^i-1} \text{Var}(X_j) \\ &\leq \frac{1}{4^{i-1}\varepsilon^2} \sum_{j=1}^{2^i-1} \sigma_j^2. \end{aligned} \tag{10.23}$$

Entonces de (10.22) y (10.23) obtenemos para cada i

$$\lambda_i \leq \frac{1}{4^{i-1}\varepsilon^2} \sum_{j=1}^{2^i-1} \sigma_j^2,$$

y cambiando el orden de sumación resulta

$$\begin{aligned}\sum_{i=1}^{\infty} \lambda_i &\leq \sum_{i=1}^{\infty} \frac{1}{4^{i-1} \varepsilon^2} \sum_{j=1}^{2^i-1} \sigma_j^2 \\ &= \frac{1}{\varepsilon^2} \sum_{j=1}^{\infty} \sigma_j^2 \sum_{i: 2^i-1 \geq j} \frac{1}{4^{i-1}}.\end{aligned}\quad (10.24)$$

La desigualdad $2^i - 1 \geq j$ es equivalente a

$$i \geq \frac{\ln(j+1)}{\ln(2)} = i_0(j),$$

y entonces podemos escribir

$$\begin{aligned}\sum_{i: 2^i-1 \geq j} \frac{1}{4^{i-1}} &= 4 \sum_{i \geq i_0(j)} \frac{1}{4^i} \\ &= 4a_0 \left(\frac{1}{1 - \frac{1}{4}} \right) \\ &= \frac{16}{3} a_0,\end{aligned}\quad (10.25)$$

donde a_0 es el primer término de la serie geométrica.

$$\sum_{i \geq i_0(j)} \frac{1}{4^i}.\quad (10.26)$$

Por otro lado $2^i - 1 \geq j$ implica que $4^i \geq j^2$, es decir para todos los términos de la serie geométrica (10.26) obtenemos

$$\frac{1}{4^i} \leq \frac{1}{j^2},$$

y en particular se tendrá

$$a_0 \leq \frac{1}{j^2}.\quad (10.27)$$

Entonces por (10.25 y (10.27) se tiene

$$\sum_{2^i-1 \geq j} \frac{1}{4^{i-1}} = \frac{16}{3} a_0 \leq \frac{16}{3} \frac{1}{j^2} = \frac{16}{3} \frac{1}{j^2},$$

y de acuerdo a (10.24) se tiene

$$\sum_{i=1}^{\infty} \lambda_i \leq \frac{16}{3\varepsilon^2} \sum_{j=1}^{\infty} \frac{\sigma_j^2}{j^2} < \infty.$$

Esto prueba la Ley Fuerte de los Grandes Números. \square

Observación. La condición (10.21) se cumple si todas las varianzas están acotadas. En efecto, si existe una constante K tal que para todo i , $\sigma_i^2 \leq K$ entonces como se tiene

$$\sum_{i=1}^{\infty} \frac{1}{i^2} < \infty,$$

resulta

$$\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} \leq K \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty.$$

Para el caso en que para todo i , $\mu_i = \mu$, $\sigma_i^2 = \sigma^2$ se cumple efectivamente que

$$\sigma^2 \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty,$$

y por lo tanto

$$\bar{X}_n - \mu \rightarrow 0 \text{ c.s.},$$

o equivalentemente

$$\bar{X}_n \rightarrow \mu \text{ c.s.}$$

Todas las consideraciones posteriores a la ley débil que discuten como ésta fundamenta las nociones heurísticas de esperanza de un variable aleatoria y de probabilidad de un evento siguen valiendo, reemplazando la convergencia en probabilidad por convergencia casi segura.

10.6. Teorema de la Convergencia Dominada

Ahora daremos una demostración del Teorema de la Convergencia Dominada (Lebesgue). Antes necesitamos el siguiente caso particular.

Teorema 10.14 Sean $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias no negativas y Z una variable aleatoria no negativa con $E(Z) < \infty$ que domina todos los términos de la sucesión, es decir $0 \leq X_n \leq Z$. Entonces si $X_n \xrightarrow{P} 0$ se tiene

$$E(X_n) \rightarrow 0.$$

Demostración. Recordemos que si $Z \geq 0$ la condición de $E(Z) < \infty$ es equivalente a $\int_0^{\infty} z dF_Z < \infty$ y esto es equivalente a $\lim_{k \rightarrow \infty} \int_k^k z dF_Z = 0$.

Vamos a demostrar que dado $\varepsilon > 0$ existe n_0 tal que si $n \geq n_0$ entonces $E(X_n) < \varepsilon$.

Dado $K > 0$ (arbitrario) particionamos al espacio de la siguiente manera

$$\Omega = \left\{ X_n \leq \frac{\varepsilon}{3} \right\} \cup \left\{ \frac{\varepsilon}{3} < X_n \leq K \right\} \cup \{ X_n > K \}.$$

Entonces

$$\begin{aligned} 0 \leq X_n &= X_n I_{\{X_n \leq \varepsilon/3\}} + X_n I_{\{\varepsilon/3 < X_n \leq K\}} + X_n I_{\{X_n > K\}} \\ &\leq \frac{\varepsilon}{3} + K I_{\{X_n > \varepsilon/3\}} + Z I_{\{Z > K\}}. \end{aligned} \quad (10.28)$$

Tomando esperanza en ambos miembros se tiene

$$E(X_n) \leq \frac{\varepsilon}{3} + KP \left(X_n > \frac{\varepsilon}{3} \right) + E(Z I_{\{Z > K\}}). \quad (10.29)$$

Sea $Y_K = Z I_{\{Z > K\}}$, luego

$$F_{Y_K}(y) = \begin{cases} 0 & \text{si } y < 0 \\ F_Z(K) & \text{si } 0 \leq y \leq K \\ F_Z(y) & \text{si } y > K, \end{cases}$$

y entonces

$$\begin{aligned} E(Z I_{\{Z > K\}}) &= E(Y_K) \\ &= \int_K^{+\infty} z dF_Z. \end{aligned}$$

Dado que $E(Z) < \infty$ existe K_0 tal que

$$E(Z I_{\{Z > K_0\}}) < \frac{\varepsilon}{3}. \quad (10.30)$$

Una vez elegido K_0 , usando que $X_n \xrightarrow{P} 0$, podemos encontrar n_0 tal que para todo $n \geq n_0$ se tiene

$$P \left(X_n > \frac{\varepsilon}{3} \right) < \frac{\varepsilon}{3K_0}. \quad (10.31)$$

Luego de (10.29), (10.30) y (10.31) resulta que para todo $n \geq n_0$

$$0 \leq E(X_n) \leq \frac{\varepsilon}{3} + K_0 \frac{\varepsilon}{3K_0} + \frac{\varepsilon}{3} = \varepsilon,$$

y el Teorema queda demostrado. \square

Ahora probaremos el Teorema de la Convergencia Dominada en el caso general.

Teorema 10.15 (Teorema de la Convergencia Dominada) *Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias tal que existe un variable $Z \geq 0$ con $E(Z) < \infty$ y $|X_n| \leq Z$ para todo n . Entonces si $X_n \xrightarrow{P} X$ se tendrá*

$$E(X_n) \rightarrow E(X).$$

Demostración. Debemos probar que

$$\lim_{n \rightarrow \infty} |E(X_n) - E(X)| = 0.$$

Ahora bien, por una propiedad de la esperanza

$$|E(X_n) - E(X)| = |E(X_n - X)| \leq E(|X_n - X|),$$

de manera que bastará con probar que

$$\lim_{n \rightarrow \infty} E(|X_n - X|) = 0. \quad (10.32)$$

Sea

$$Y_n = |X_n - X| \geq 0,$$

luego como $X_n \xrightarrow{P} X$ resulta $Y_n \xrightarrow{P} 0$.

Como

$$\{|X| > Z + 1\} \subset \{|X_n| > Z\} \cup \{|X_n - X| > 1\},$$

y dado $P(|X_n| > Z) = 0$ se tendrá para todo $\varepsilon > 0$,

$$P(|X| > Z + 1) \leq P(|X_n - X| > 1)$$

y por lo tanto como $X_n \xrightarrow{P} X$

$$P(|X| > Z + 1) \leq \lim_{n \rightarrow \infty} P(|X_n - X| > 1) = 0.$$

Esto muestra que para todo $\varepsilon > 0$ se tiene $P(|X| \leq Z + 1) = 1$.

Luego con probabilidad 1 se tiene $Y_n \leq |X_n| + |X| \leq 2Z + 1$, y estamos en la situación del Teorema 10.14. Por lo tanto podemos concluir que $E(Y_n) \rightarrow 0$. Luego (10.32) se cumple y el teorema queda demostrado. \square

Capítulo 11

Convergencia en Distribución.

11.1. Definición de convergencia en distribución.

Tanto la convergencia casi segura como la convergencia en probabilidad se basan en el concepto de “proximidad entre variables aleatorias”. Veremos ahora un tipo de convergencia que se basa en la proximidad entre las respectivas funciones de distribución.

Definición 11.1 Sea $(F_n)_{n \geq 1}$ una sucesión de funciones de distribución definidas sobre \mathbb{R} y F otra función de distribución. Diremos que la sucesión F_n converge débilmente a F si para todo punto x de continuidad de F , las F_n convergen puntualmente a F . Es decir, si para todo x tal que F es continua en x se tiene que

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

Notación. Si $\{F_n\}_{n \geq 1}$ converge débilmente en distribución a F escribiremos

$$F_n \xrightarrow{\mathcal{D}} F.$$

Observación. Recordemos que una función de distribución definida sobre \mathbb{R} se caracteriza por las propiedades P1, P2, P3 y P4 del teorema 2.5 y que el conjunto de puntos donde es discontinua es a lo sumo numerable.

Definición 11.2 Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias y F una función de distribución. Diremos que la sucesión X_n converge en distribución a F si $(F_{X_n})_{n \geq 1}$ converge débilmente a F .

Notación. Si $(X_n)_{n \geq 1}$ converge en distribución a F escribiremos

$$X_n \xrightarrow{\mathcal{D}} F.$$

Observación. Por extensión también diremos que $(X_n)_{n \geq 1}$ converge en distribución a X si $F_{X_n} \xrightarrow{\mathcal{D}} F_X$.

Al decir que $(X_n)_{n \geq 1}$ converge en distribución a X hay un abuso de lenguaje puesto que las variables X_n no se aproximan a X , sino que son las funciones de distribución de las X_n las que se aproximan a la función de distribución de X .

Consideremos el caso donde X e Y son dos variables independientes con distribución $N(0, 1)$. Definamos para todo n , $X_n = X$ entonces $X_n \xrightarrow{\mathcal{D}} Y$ y sin embargo como las variables X e Y son independientes, X no se “aproxima” a Y .

Veamos ahora la relación que existe entre la convergencia en probabilidad y la convergencia en distribución.

Teorema 11.1 Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias y X otra variable aleatoria. Entonces

$$X_n \xrightarrow{P} X$$

implica que

$$X_n \xrightarrow{\mathcal{D}} X.$$

Demostración. Sea F_X la función de distribución de X y x un punto de continuidad de F_X . Probemos primero que

$$\{X_n \leq x\} \subset \{X \leq x + \varepsilon\} \cup \{|X_n - X| \geq \varepsilon\}. \quad (11.1)$$

Para esto basta demostrar que si ω no está en ninguno de los dos conjuntos que forman la unión en el miembro derecho, entonces no está en $\{X_n \leq x\}$. Sea ω tal que $X(\omega) > x + \varepsilon$ y $|X_n(\omega) - X(\omega)| < \varepsilon$. Luego

$$\begin{aligned} X_n(\omega) &= X(\omega) + (X_n(\omega) - X(\omega)) \\ &\geq X(\omega) - |X_n(\omega) - X(\omega)| \\ &> x + \varepsilon - \varepsilon \\ &= x, \end{aligned}$$

probando (11.1). Tomado probabilidades en ambos miembros se obtiene

$$F_{X_n}(x) \leq F_X(x + \varepsilon) + P(|X_n - X| \geq \varepsilon).$$

Tomando límite superior en ambos miembros y teniendo en cuenta que

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0 \quad (11.2)$$

se obtiene

$$\overline{\lim}_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \varepsilon),$$

y haciendo que $\varepsilon \rightarrow 0$, en virtud de que las funciones de distribución son continuas a derecha se tiene que

$$\overline{\lim}_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x). \quad (11.3)$$

Ahora hacemos un razonamiento similar a izquierda de x . Consideramos la inclusión

$$\{X \leq x - \varepsilon\} \subset \{X_n \leq x\} \cup \{|X_n - X| \geq \varepsilon\}.$$

Tomado probabilidades en ambos miembros se obtiene

$$F_X(x - \varepsilon) \leq F_{X_n}(x) + P(|X_n - X| \geq \varepsilon).$$

Tomando límite inferior en ambos miembros y usando (11.2) se obtiene

$$F(x - \varepsilon) \leq \underline{\lim}_{n \rightarrow \infty} F_{X_n}(x),$$

y haciendo que $\varepsilon \rightarrow 0$, en virtud de la continuidad de F_X en x

$$F(x) \leq \underline{\lim}_{n \rightarrow \infty} F_{X_n}(x). \quad (11.4)$$

De (11.3) y (11.4) resulta

$$\overline{\lim}_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x) \leq \underline{\lim}_{n \rightarrow \infty} F_{X_n}(x),$$

y como

$$\underline{\lim}_{n \rightarrow \infty} F_{X_n}(x) \leq \overline{\lim}_{n \rightarrow \infty} F_{X_n}(x),$$

debe ser

$$\overline{\lim}_{n \rightarrow \infty} F_{X_n}(x) = \underline{\lim}_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

Luego existe el límite de (F_{X_n}) en el punto x y además

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F(x). \quad \square$$

Observación. La recíproca no vale en general. Pero sí es cierta en el caso en que $P(X = C) = 1$, donde C es una constante. Luego tenemos el siguiente teorema cuya demostración queda como ejercicio.

Teorema 11.2 *Supongamos que $X_n \xrightarrow{D} X$ y $P(X = C) = 1$. Entonces $X_n \xrightarrow{P} X$.*

11.2. Funciones características.

Una herramienta muy importante para la demostración del Teorema Central del Límite es la función característica asociada a una distribución. Para definirla necesitaremos presentar el concepto de variable aleatoria compleja.

11.2.1. Variables aleatorias complejas.

Definición 11.3 Sea (Ω, \mathcal{A}, P) un espacio de probabilidad. Se dice que X es una variable aleatoria compleja si $X : \Omega \rightarrow \mathbb{C}$ (\mathbb{C} indica el conjunto de números complejos) es de la forma $X = X_1 + iX_2$ con X_1 y X_2 variables aleatorias reales.

Definición 11.4 Sea la variable aleatoria compleja $X = X_1 + iX_2$, donde X_1 y X_2 tienen esperanza finita. Definimos la esperanza de X como $E(X) = E(X_1) + iE(X_2)$.

Observación. $E(X) \in \mathbb{C}$. La parte real e imaginaria de la esperanza son respectivamente $\operatorname{Re}(E(X)) = E(X_1)$ e $\operatorname{Im} E(X) = E(X_2)$.

Definición 11.5 Diremos que dos variables aleatorias complejas $X = X_1 + iX_2$ e $Y = Y_1 + iY_2$ son independientes si el vector aleatorio $\mathbf{X} = (X_1, X_2)$ es independiente del vector aleatorio $\mathbf{Y} = (Y_1, Y_2)$.

Algunas propiedades

Veamos ahora algunas propiedades que cumplen las variables complejas, en analogía con las que ya probamos para variables aleatorias reales.

Propiedad 11.1 Sean $X = X_1 + iX_2$ e $Y = Y_1 + iY_2$ dos variables aleatorias complejas independientes. Entonces

$$E(XY) = E(X)E(Y).$$

Demostración. La demostración se basa en el cálculo directo usando la definición y la propiedad análoga para variables aleatorias reales independientes

$$\begin{aligned} E(XY) &= E[(X_1 + iX_2)(Y_1 + iY_2)] \\ &= E[(X_1Y_1 - X_2Y_2) + i(X_2Y_1 + Y_2X_1)] \\ &= E(X_1Y_1 - X_2Y_2) + iE(X_2Y_1 + Y_2X_1) = \\ &= E(X_1Y_1) - E(X_2Y_2) + iE(X_2Y_1) + iE(Y_2X_1) = \\ &= E(X_1)E(Y_1) - E(X_2)E(Y_2) + iE(X_2)E(Y_1) + iE(Y_2)E(X_1) \\ &= (E(X_1) + iE(X_2))(E(Y_1) + iE(Y_2)) \\ &= E(X)E(Y). \quad \square \end{aligned}$$

Propiedad 11.2 *Sea una variable compleja $X = X_1 + iX_2$. Entonces*

$$|E(X)| \leq E(|X|).$$

Demostración. Podemos suponer que $E(X) \neq 0$ pues en tal caso la desigualdad se cumple. Como $E(X) = E(X_1) + iE(X_2) \in \mathbb{C}$ podemos escribir

$$E(X) = re^{i\theta}$$

para cierto $r > 0, 0 \leq \theta < 2\pi$. Consideremos la variable aleatoria compleja $Y = e^{-i\theta}X$ y verifiquemos que su esperanza es real

$$\begin{aligned} E(Y) &= E(e^{-i\theta}X) \\ &= e^{-i\theta}E(X) \\ &= r > 0. \end{aligned}$$

Hemos probado con anterioridad que la propiedad se cumple para esperanzas de variables aleatorias reales. Luego

$$|E(Y)| \leq E(|Y|).$$

A partir de esto se deduce la tesis, pues

$$|E(X)| = r = E(Y) = |E(Y)| \leq E(|Y|) = E(|X|). \quad \square$$

11.2.2. Definición de función característica y propiedades.

Definición 11.6 *Sea X una variable aleatoria y F_X su función de distribución. Definimos a la función característica de X por la función $\phi_X : \mathbb{R} \rightarrow \mathbb{C}$ asociada a F_X de la siguiente manera*

$$\begin{aligned} \phi_X(t) &= E(\exp(itX)) \\ &= E(\cos(tX)) + iE(\sin(tX)). \end{aligned}$$

Observación. Como las variables $\cos(tX)$, $\sin(tX)$ son acotadas, las esperanzas de estas variables existen y son finitas.

El motivo de la introducción de la función característica es poder estudiar más fácilmente la distribución de la suma de variables aleatorias independientes. Mientras que la función de distribución de esta suma (que se obtiene por convoluciones) puede ser muy complicada, su función característica es muy simple, como se desprende de la Propiedad 11.3 que damos a continuación. Por otro lado, como veremos más adelante, hay una correspondencia biunívoca entre funciones de distribución y funciones características. Luego, conociendo la función característica de una variable aleatoria, también conocemos su función de distribución.

Propiedad 11.3 Sean X e Y dos variables aleatorias independientes. Entonces para todo $t \in \mathbb{R}$

$$\phi_{X+Y}(t) = \phi_X(t) \phi_Y(t).$$

Demostración. Observando que $\exp(itX)$, $\exp(itY)$ son variables aleatorias independientes se tiene

$$\begin{aligned} \phi_{X+Y}(t) &= E(\exp(it(X+Y))) \\ &= E(\exp(itX) \exp(itY)) \\ &= E(\exp(itX)) E(\exp(itY)) \\ &= \phi_X(t) \phi_Y(t). \quad \square \end{aligned}$$

Propiedad 11.4 Sea X una variable aleatoria. Entonces para todo $t \in \mathbb{R}$

$$|\phi_X(t)| \leq 1.$$

Demostración.

$$|\phi_X| = |E(\exp(itX))| \leq E(|\exp(itX)|) = E(1) = 1. \quad \square$$

Propiedad 11.5 $\phi_X(0) = E(1) = 1$.

Demostración. $\phi_X(0) = E(1) = 1$. \square

Ahora enunciamos dos teoremas muy importantes. Las demostraciones de estos teoremas se pueden encontrar en el libro de Barry R. James, "Probabilidad: un curso em nivel intermediário".

Teorema 11.3 Sean X e Y dos variables aleatorias. Entonces si

$$\phi_X = \phi_Y,$$

también se tiene

$$F_X = F_Y.$$

Teorema 11.4 (Teorema de Continuidad de Paul Levy) Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias, $(F_{X_n})_{n \geq 1}$ la correspondiente sucesión de funciones de distribución y $(\phi_{X_n})_{n \geq 1}$ la correspondiente sucesión de funciones características asociadas. Entonces

$$X_n \xrightarrow{\mathcal{D}} X$$

si y sólo si para todo $t \in \mathbb{R}$

$$\phi_{X_n}(t) \rightarrow \phi_X(t).$$

Teorema 11.5 *Sea X una variable aleatoria. Entonces ϕ_X es continua en todo punto.*

Demostración. Sea $t \in \mathbb{R}$ y consideremos una sucesión $(h_n)_{n \geq 1} \subset \mathbb{R}$ tal que $h_n \rightarrow 0$. Queremos probar que

$$\lim_{n \rightarrow \infty} \phi_X(t + h_n) = \phi_X(t).$$

Teniendo en cuenta que

$$\phi_X(t + h_n) = E(\cos((t + h_n)X)) + iE(\sin((t + h_n)X)),$$

bastará con probar que si $n \rightarrow +\infty$ entonces

$$E(\cos((t + h_n)X)) \rightarrow E(\cos(tX)),$$

y

$$E(\sin((t + h_n)X)) \rightarrow E(\sin(tX)).$$

Probaremos que $E(\cos((t + h_n)X)) \rightarrow E(\cos(tX))$ cuando $n \rightarrow +\infty$, la otra propiedad es análoga.

Consideremos la sucesión de variables aleatorias

$$Y_n = \cos((t + h_n)X).$$

Se comprueba fácilmente que Y_n está dominada por la variable aleatoria $Z = 1$, es decir para todo n

$$|Y_n| = |\cos((t + h_n)X)| \leq 1.$$

Además si $Y = \cos(tX)$, por la continuidad de la función coseno, se tiene convergencia puntual de Y_n a Y , es decir para todo $\omega \in \Omega$

$$Y_n(\omega) \rightarrow Y(\omega).$$

Luego, por el Teorema de Convergencia Dominada se obtiene

$$E(Y_n) \rightarrow E(Y). \quad \square$$

Observación. Se puede probar algo más fuerte: ϕ_X es uniformemente continua (ver el libro de Barry R. James).

Veamos como opera una función característica sobre una transformación afín de la variable aleatoria.

Propiedad 11.6 Sea X una variable aleatoria e $Y = aX + b$, con $a, b \in \mathbb{R}$.
Entonces para todo $t \in \mathbb{R}$

$$\phi_{aX+b}(t) = \exp(itb) \phi_X(at).$$

Demostración.

Para todo $t \in \mathbb{R}$ se tiene

$$\begin{aligned} \phi_Y(t) &= \phi_{aX+b}(t) \\ &= E(\exp(it(aX + b))) \\ &= E(\exp(it(aX)) \exp(itb)) \\ &= \exp(itb) E(\exp(it(aX))) \\ &= \exp(itb) \phi_X(at). \quad \square \end{aligned}$$

Ahora queremos caracterizar a las funciones características a valores reales. Para esto recordemos el concepto de variable aleatoria simétrica respecto del origen. La definición más general de simetría respecto de μ arbitrario está dada en la página 155.

Decimos que una variable aleatoria X es *simétrica respecto del origen* si y sólo si para todo $x \geq 0$ se tiene que

$$P(X \leq -x) = P(X \geq x). \quad (11.5)$$

El siguiente teorema permite dar una definición equivalente.

Teorema 11.6 ϕ_X es real sii X es simétrica respecto del origen. En este caso ϕ_X es par.

Demostración. Supongamos primero que X sea simétrica respecto del origen. Como para todo $t \in \mathbb{R}$

$$\phi_X(t) = E(\cos(tX)) + iE(\sin(tX)),$$

para mostrar que ϕ_X es real bastará ver que $E(\sin(tX)) = 0$.

Teniendo en cuenta que si X es simétrica se tiene que $F_X = F_{-X}$, de manera que $E(g(X)) = E(g(-X))$ para cualquier g medible, entonces si para cada $t \in \mathbb{R}$ se toma $g(x) = \sin(tx)$ se obtiene

$$E(\sin(tX)) = E(\sin(-tX)) = -E(\sin(tX)),$$

y por lo tanto $E(\sin(tX)) = 0$.

Además,

$$\begin{aligned}\phi_X(-t) &= E(\cos(X(-t))) \\ &= E(\cos(Xt)) \\ &= \phi_X(t).\end{aligned}$$

Luego ϕ_X es par.

Supongamos ahora que ϕ_X es real, esto es $E(\operatorname{sen}(tX)) = 0$. Entonces teniendo en cuenta que la función coseno es par y la función seno impar tendremos para todo $t \in \mathbb{R}$

$$\begin{aligned}\phi_X(t) &= E(\cos(tX)) + iE(\operatorname{sen}(tX)) \\ &= E(\cos(tX)),\end{aligned}$$

y

$$\begin{aligned}\phi_{-X}(t) &= E(\cos(t(-X))) + iE(\operatorname{sen}(t(-X))) \\ &= E(\cos(tX)) - iE(\operatorname{sen}(tX)) \\ &= E(\cos(tX))\end{aligned}$$

Luego $\phi_X(t) = \phi_{-X}(t)$ y entonces por el Teorema 11.3, se obtiene que $F_X = F_{-X}$ y por el Teorema 7.17 que X es simétrica respecto del origen. \square

Definición 11.7 (Momentos de orden k) Sea X una variable aleatoria. Definimos el momento de orden $k > 0$ de X como el número

$$\mu_k = E\left(X^k\right),$$

cuando este valor existe y el momento absoluto de orden $k > 0$ de X como el número

$$\mu_k^* = E\left(|X|^k\right).$$

Observación. Si k es par entonces $\mu_k = \mu_k^*$. Además siempre se tiene que $\mu_k < \infty$ sii $\mu_k^* < \infty$, es decir la integrabilidad absoluta de $|X|^k$ equivale a la de X^k . En particular $E(X) = \mu_1$ y $\operatorname{Var}(X) = \mu_2 - \mu_1^2$.

Teorema 11.7 Si $\mu_k^* < \infty$ entonces para todo $i < k$ se tiene $\mu_i^* < \infty$.

Demostración. Sea $i < k$. Se tiene

$$|X|^i = I_{\{|X| \leq 1\}} |X|^i + I_{\{|X| > 1\}} |X|^i.$$

Como

$$I_{\{|X|^i \leq 1\}} |X|^i \leq I_{\{|X| \leq 1\}}$$

y

$$I_{\{|X| > 1\}} |X|^i \leq I_{\{|X| > 1\}} |X|^k \leq |X|^k$$

obtenemos

$$|X|^i \leq I_{\{|X| \leq 1\}} + |X|^k.$$

Tomando esperanza en ambos miembros resulta

$$\mu_i^* \leq P(\{|X| \leq 1\}) + \mu_k^* < \infty,$$

y esto demuestra el teorema. \square

11.3. Momentos y función característica.

11.3.1. Derivación dentro del signo esperanza.

Para hacer un desarrollo de Taylor de la función característica, necesitaremos hallar sus derivadas. Como la función característica está definida como una esperanza, será conveniente encontrar condiciones bajo las cuales se pueda intercambiar el orden en el que se deriva y se toma esperanza.

Sea $g(x, t)$ una función de dos variables a valores reales, medible respecto de la primera variable y derivable respecto de la segunda variable. Sea g_2 definida por

$$g_2(x, t) = \frac{\partial g(x, t)}{\partial t}.$$

Sea X una variable aleatoria, entonces para cada t , $Y_t = g(X, t)$ es también una variable aleatoria. Supongamos que $E(|Y_t|) < \infty$ y consideremos la función $h(t) = E(Y_t) = E(g(X, t))$. El siguiente teorema nos da condiciones suficientes para que $h'(t) = E(g_2(X, t))$.

Teorema 11.8 *Supongamos que en $t = t_0$ se cumplen las siguientes condiciones:*

(i) *existe $\varepsilon > 0$ y Z variable aleatoria con $E(Z) < \infty$, tal que*

$$\sup_{|t-t_0| \leq \varepsilon} \{|g_2(X, t)|\} \leq Z,$$

(ii) *para todo x la función $g_2(x, t)$ es continua respecto a la segunda variable en $t = t_0$.*

Luego $h'(t_0) = E(g_2(X, t_0))$.

Demostración.

Sea $(r_n)_{n \geq 1}$ una sucesión de números reales no creciente que converge a 0 y tal que $|r_n| \leq \varepsilon$. Bastará demostrar que

$$\lim_{n \rightarrow +\infty} \frac{h(t_0 + r_n) - h(t_0)}{r_n} = E(g_2(X, t_0)).$$

Utilizando el teorema del valor medio existe $r_n^* = r_n^*(X)$ tal que $|r_n^*(X)| \leq r_n \leq \varepsilon$ y tal que

$$\frac{g(X, t_0 + r_n) - g(X, t_0)}{r_n} = g_2(X, t_0 + r_n^*(X)).$$

Luego

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{h(t_0 + r_n) - h(t_0)}{r_n} &= \lim_{n \rightarrow \infty} E \left(\frac{g(X, t_0 + r_n) - g(X, t_0)}{r_n} \right) \\ &= \lim_{n \rightarrow \infty} E(g_2(X, t_0 + r_n^*(X))). \end{aligned}$$

Por lo tanto bastará con mostrar que

$$\lim_{n \rightarrow \infty} E(g_2(X, t_0 + r_n^*(X))) = E(g_2(X, t_0)). \quad (11.6)$$

Ahora bien $r_n^*(X) \rightarrow 0$ y por la continuidad de g_2 en $t = t_0$, $(g_2(X, t_0 + r_n^*(X)))_{n \geq 1}$ converge puntualmente a la función $g_2(X, t_0)$. Además se cumple que

$$\sup_{n \in \mathbb{N}} |g_2(X, t_0 + r_n^*(X))| \leq Z,$$

con $E(Z) < \infty$. Luego aplicando el teorema de la convergencia dominada se obtiene (11.6). \square

11.3.2. Derivadas de la función característica y momentos.

Dada una variable aleatoria X , sabemos que $\phi_X(t) = E(\exp(itX))$. Procedamos de manera ingenua, sin preocuparnos por la justificación, y derivemos sucesivamente dentro del signo esperanza

$$\begin{aligned} \phi_X^{(1)}(t) &= E(iX \exp(itX)) = iE(X \exp(itX)) \\ \phi_X^{(2)}(t) &= E(i^2 X^2 \exp(itX)) = i^2 E(X^2 \exp(itX)) \\ &\vdots \\ \phi_X^{(n)}(t) &= E(i^n X^n \exp(itX)) = i^n E(X^n \exp(itX)). \end{aligned}$$

El siguiente teorema permite justificar estas expresiones.

Teorema 11.9 *Supongamos que $\mu_n^* < \infty$. Luego se cumple que*

$$\phi_X^{(n)}(t) = i^n E(X^n \exp(itX)). \quad (11.7)$$

Demostración. Demostraremos el teorema por inducción en n . Para $n = 0$ es cierto ya que $\phi_X(t) = E \exp(itX)$ por definición. Supongamos que el teorema es cierto para n . Vamos a demostrar que es cierto para $n + 1$. Supongamos que $\mu_{n+1}^* < \infty$, por el Teorema 11.7 resulta $\mu_n^* < \infty$ y luego la fórmula (11.7) es cierta para n . Entonces, tenemos que

$$\begin{aligned} \phi_X^{(n)}(t) &= i^n E(X^n \exp(itX)) \\ &= i^n (E(X^n \cos(tX)) + iE(X^n \sen(tX))). \end{aligned} \quad (11.8)$$

Sea $g(x, t) = x^n \cos(tx)$. Luego $g_2(x, t) = -x^{n+1} \sen(tx)$ es continua y $|g_2(X, t)| \leq |X|^{n+1}$. Como $E(|X|^{n+1}) < \infty$, por el Teorema 11.8 se tendrá que si $h(t) = E(X^n \cos(tx))$, entonces

$$\begin{aligned} h'(t) &= E(g_2(X, t)) \\ &= -E(X^{n+1} \sen(tX)). \end{aligned} \quad (11.9)$$

Similarmente si $h^*(t) = E(X^n \sen(tx))$, luego

$$h^{*'}(t) = E(X^{n+1} \cos(tX)). \quad (11.10)$$

Luego por (11.9), (11.10), derivando (11.8) se tendrá

$$\phi_X^{(n+1)}(t) = i^n (h'(t) + h^{*'}(t)) \quad (11.11)$$

$$= i^n (E(-X^{n+1} \sen(tX)) + iE(X^{n+1} \cos(tX))). \quad (11.12)$$

Multiplicando por i y dividiendo por i se obtiene

$$\phi_X^{(n+1)}(t) = i^{n+1} ((1/i)E(-X^{n+1} \sen(tX)) + E(X^{n+1} \cos(tX))),$$

y usando que $1/i = -i$

$$\begin{aligned} \phi_X^{(n+1)}(t) &= i^{n+1} (iE(X^{n+1} \sen(tX)) + E(X^{n+1} \cos(tX))) \\ &= i^{n+1} E(X^{n+1} \exp(itX)) \end{aligned}$$

y por lo tanto el teorema queda demostrado. \square .

Corolario 11.1 *Supongamos $\mu_n^* < \infty$. Entonces resulta que*

$$\phi_X^{(n)}(0) = i^n E(X^n).$$

Observemos entonces que de acuerdo al Teorema 11.9 si $\mu_n^* < \infty$ resulta

$$\begin{aligned}\phi_X^{(n)}(0) &= i^n E(X^n) \\ &= i^n \mu_n.\end{aligned}$$

En particular

$$\phi_X'(0) = i\mu_1 \quad (11.13)$$

y

$$\phi_X''(0) = -\mu_2. \quad (11.14)$$

Ahora estamos en condiciones de probar que la función característica de la distribución $X \sim N(0, 1)$ es su densidad, salvo una constante.

11.4. Función característica de una distribución normal.

Para la prueba del Teorema Central de Límite, necesitamos calcular la función característica de una distribución normal. Dado que si $X \sim N(\mu, \sigma^2)$ se puede escribir como $X = \sigma Y + \mu$, donde $Y \sim N(0, 1)$ de acuerdo a la Propiedad 11.6, sólo se necesitará calcular ϕ_X para el caso $\mu = 0$ y $\sigma^2 = 1$.

Teorema 11.10 *Sea $X \sim N(0, 1)$. La función característica de X es*

$$\phi^*(t) = \exp\left(-\frac{1}{2}t^2\right).$$

Demostración. Como X es simétrica respecto del origen, ϕ^* es real y par. Consideremos dos variables aleatorias independientes $X_1 \sim N(0, 1)$, $X_2 \sim N(0, 1)$ y definamos $Y = u_1 X_1 + u_2 X_2$ con $u_1 \geq 0$, $u_2 \geq 0$. Entonces $Y \sim N(0, u_1^2 + u_2^2)$.

Podemos expresar a Y como un múltiplo de una variable $N(0, 1)$. En efecto

$$\begin{aligned}Y &= \sqrt{u_1^2 + u_2^2} \frac{Y}{\sqrt{u_1^2 + u_2^2}} \\ &= \sqrt{u_1^2 + u_2^2} Z,\end{aligned}$$

donde

$$Z = \frac{Y}{\sqrt{u_1^2 + u_2^2}}$$

tiene distribución $N(0, 1)$.

Demostración. Calcularemos ϕ_Y de dos manera distintas. Por un lado, usando la Propiedad 11.6

$$\phi_Y(t) = \phi_{\sqrt{u_1^2 + u_2^2}Z}(t) \quad (11.15)$$

$$= \phi^* \left(\sqrt{u_1^2 + u_2^2}t \right). \quad (11.16)$$

Por otro lado siendo Y suma de variables aleatorias independientes, usando la Propiedad ?? y recordando que $u_1 \geq 0$ y $u_2 \geq 0$, se tiene que

$$\begin{aligned} \phi_Y(t) &= \phi_{u_1X_1 + u_2X_2}(t) \\ &= \phi_{u_1X_1}(t) \phi_{u_2X_2}(t) \end{aligned} \quad (11.17)$$

$$\begin{aligned} &= \phi^*(u_1t) \phi^*(u_2t) \\ &= \phi^* \left(\sqrt{u_1^2}t \right) \phi^* \left(\sqrt{u_2^2}t \right). \end{aligned} \quad (11.18)$$

De (11.15) y (11.18) se obtiene

$$\phi^* \left(\sqrt{u_1^2 + u_2^2}t \right) = \phi^* \left(\sqrt{u_1^2}t \right) \phi^* \left(\sqrt{u_2^2}t \right), \quad (11.19)$$

y haciendo $t = 1$

$$\phi^* \left(\sqrt{u_1^2 + u_2^2} \right) = \phi^* \left(\sqrt{u_1^2} \right) \phi^* \left(\sqrt{u_2^2} \right). \quad (11.20)$$

Definamos g^* como la composición de ϕ^* con la raíz cuadrada, es decir

$$g^*(u) = \phi^* \left(\sqrt{u} \right).$$

Luego por (11.20) se tiene

$$g^*(u_1^2 + u_2^2) = g^*(u_1^2) g^*(u_2^2).$$

Luego, si ponemos $v_1 = u_1^2$ y $v_2 = u_2^2$ entonces para todo $v_1, v_2 \geq 0$ obtenemos

$$g^*(v_1 + v_2) = g^*(v_1) g^*(v_2). \quad (11.21)$$

Entonces para todo $v \geq 0$ se tiene

$$g^*(v) = g^* \left(\frac{v}{2} + \frac{v}{2} \right) = \left(g^* \left(\frac{v}{2} \right) \right)^2 \geq 0.$$

Observación. La Ecuación (11.21) recuerda la caracterización de la distribución exponencial como una distribución con “falta de memoria”. Luego para caracterizar a g^* procederemos de igual manera.

Por inducción se puede probar que dados $v_1 \geq 0, v_2 \geq 0, \dots, v_n \geq 0$ entonces

$$g^* \left(\sum_{i=1}^n v_i \right) = \prod_{i=1}^n g^*(v_i). \quad (11.22)$$

Luego usando (11.22) se obtiene que para todo n natural

$$\begin{aligned} g^*(n) &= g^* \left(\underbrace{1 + 1 + \dots + 1}_{n \text{ veces}} \right) \\ &= [g^*(1)]^n. \end{aligned} \quad (11.23)$$

Usando (11.22) y (11.23) se obtiene que para todo m y n naturales

$$\begin{aligned} [g^*(1)]^n &= g^*(n) \\ &= g^* \left(m \frac{n}{m} \right) \\ &= g^* \left(\underbrace{\frac{n}{m} + \frac{n}{m} + \dots + \frac{n}{m}}_{m \text{ veces}} \right) \\ &= \left[g^* \left(\frac{n}{m} \right) \right]^m, \end{aligned}$$

y entonces

$$g^* \left(\frac{n}{m} \right) = [g^*(1)]^{\frac{n}{m}}.$$

Luego para todo $r \in \mathbb{Q}$ positivo se tiene

$$g^*(r) = [g^*(1)]^r.$$

Por la continuidad de g^* y la densidad de \mathbb{Q} en \mathbb{R} , se concluye que para todo $x \in \mathbb{R}_{\geq 0}$

$$g^*(x) = [g^*(1)]^x.$$

Ahora veamos que

$$0 < g^*(1) < 1. \quad (11.24)$$

Como $g^*(1)$ es real con $0 \leq g^*(1) \leq 1$ para demostrar (11.24) se deberá mostrar que $g^*(1) \neq 0$ y que $g^*(1) \neq 1$.

Supongamos que $g^*(1) = 0$. Entonces para todo $t \in \mathbb{R}_{\geq 0}$

$$\phi^*(\sqrt{t}) = g^*(t) = [g^*(1)]^t = 0.$$

Esto es absurdo, pues si $t = 0$ se tendría $\phi^*(0) = 0$ y según la Propiedad 11.5 resulta que $\phi^*(0) = 1$.

Supongamos que $g^*(1) = 1$ entonces

$$\phi^*(1) = \phi^*(\sqrt{1}) = g^*(1) = 1.$$

Ahora como ϕ^* es real, $\phi^*(1) = E(\cos(X))$. Entonces $g^*(1) = 1$ se puede escribir como

$$E(1) = E(\cos(X))$$

luego

$$E(1 - \cos(X)) = 0$$

Pero siendo la variable aleatoria $1 - \cos(X)$ no negativa se concluye que

$$P(\cos(X) = 1) = 1.$$

Esto no puede ser cierto puesto que $\{x \in R : \cos(x) = 1\}$ es un conjunto de puntos numerable, de manera que su probabilidad es cero puesto que la distribución normal es absolutamente continua.

Finalmente si ponemos $c = -\log(g^*(1))$ entonces, $c > 0$ y $g^*(1) = \exp(-c)$. Luego

$$\begin{aligned} g^*(t) &= [g^*(1)]^t \\ &= \exp(-ct), \forall t \geq 0. \end{aligned}$$

Además

$$\phi^*(t) = g^*(t^2) = \exp(-ct^2), \forall t \geq 0.$$

Como la función $\phi^*(t)$ es par se tendrá

$$\phi^*(t) = \exp(-ct^2), \forall t.$$

Derivando dos veces ϕ^*

$$(\phi^*)^{(1)}(t) = -2ct \exp(-ct^2),$$

$$\begin{aligned} (\phi^*)^{(2)}(t) &= -2c \exp(-ct^2) + 4c^2 t^2 \exp(-ct^2) \\ &= 2c \exp(-ct^2) (2ct^2 - 1), \end{aligned}$$

y evaluando en 0, de acuerdo a (11.14) se tendrá

$$\begin{aligned} -2c &= (\phi^*)^{(2)}(0) \\ &= -\mu_2 \\ &= -(\text{Var}(X) + E(X^2)) \\ &= -1. \end{aligned}$$

Por lo tanto obtenemos que $c = \frac{1}{2}$ y el Teorema queda demostrado. \square

11.5. Teorema Central del Límite.

El siguiente lema da el desarrollo de Taylor de la función característica de una variable aleatoria X con $E(X) = 0$ y $\text{Var}(X) = 1$.

Lema 11.1 *Sea X una variable aleatoria con $E(X) = 0$ y $\text{Var}(X) = 1$. Entonces*

$$\phi_X(t) = 1 - \frac{t^2}{2} + o_2(t^2),$$

donde $o_2(t^2)$ es una función tal que

$$\lim_{t \rightarrow 0} \frac{o_2(t^2)}{t^2} = 0. \quad (11.25)$$

Demostración. Sabemos que $\phi(0) = 1$ y por (11.13) y (11.14) se tiene $\phi'_X(0) = 0$ y $\phi''_X(0) = -1$. Luego usando un desarrollo de Taylor de grado 2 en $t = 0$ para ϕ_X se tiene

$$\begin{aligned} \phi_X(t) &= \phi_X(0) + \phi'_X(0)t + \phi''_X(0)\frac{t^2}{2} + o_2(t^2) \\ &= 1 - \frac{t^2}{2} + o_2(t^2). \end{aligned}$$

donde $o_2(t^2)$ satisface (11.25). Esto demuestra el lema. \square

11.5.1. Caso de variables independientes idénticamente distribuidas

Teorema 11.11 (Teorema Central del Límite) *Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias independientes idénticamente distribuidas (i.i.d.) con varianza finita. Llamemos $\mu = E(X_i)$ y $\sigma^2 = \text{Var}(X_i) > 0$. Sean las sumas parciales*

$$S_n = \sum_{i=1}^n X_i$$

y

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}}. \quad (11.26)$$

Entonces

$$Z_n \xrightarrow{\mathcal{D}} N(0, 1). \quad (11.27)$$

Observación. La expresión (11.26) puede reformularse escribiendo

$$Z_n = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{\text{Var}(\overline{X}_n)}},$$

donde

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

es la variable aleatoria “promedio aritmético”.

Demostración. En primer lugar veamos que basta con probar el teorema suponiendo que $\mu = 0$ y $\sigma^2 = 1$. Teniendo en cuenta la independencia de las X_i y la definición de S_n se tiene que

$$\begin{aligned} E(S_n) &= n\mu, \\ \text{Var}(S_n) &= n\sigma^2. \end{aligned}$$

Luego (11.26) se puede escribir como

$$\begin{aligned} Z_n &= \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right) \\ &= \frac{\sum_{i=1}^n X_i^*}{\sqrt{n}}, \end{aligned}$$

donde

$$X_i^* = \frac{X_i - \mu}{\sigma}$$

Claramente las variables X_i^* son i.i.d. con $E(X_i^*) = 0$ y $\text{Var}(X_i^*) = 1$. Luego si probamos que el teorema vale para $\mu = 0$ y $\sigma^2 = 1$ resulta válido para μ y σ^2 arbitrarios.

Supondremos entonces que $\mu = 0$ y $\sigma^2 = 1$. De acuerdo al teorema de continuidad de Levy y al Teorema 11.10, bastará probar que para todo $t \in \mathbb{R}$

$$\lim_{n \rightarrow +\infty} \phi_{Z_n}(t) = \exp\left(-\frac{t^2}{2}\right). \quad (11.28)$$

Sabemos que como $\mu = 0$ y $\sigma^2 = 1$, por el lema anterior para todo $i \in \mathbb{N}$ se tiene

$$\phi_{X_i}(t) = \phi_X(t) = 1 - \frac{t^2}{2} + o_2(t^2),$$

donde $o_2(t^2)$ es una función tal que

$$\lim_{t \rightarrow 0} \frac{o_2(t^2)}{t^2} = 0. \quad (11.29)$$

Como las variables X_i son independientes, podemos aplicar la Propiedad 11.3 de las funciones características y se tiene que para todo n

$$\phi_{S_n}(t) = \prod_{i=1}^n \phi_{X_i}(t) = \left(1 - \frac{t^2}{2} + o_2(t^2)\right)^n.$$

Finalmente teniendo en cuenta que $\mu = 0$ y $\sigma^2 = 1$, resulta $Z_n = S_n/\sqrt{n}$. Luego por la Propiedad 11.6 de las funciones características se obtiene

$$\begin{aligned}\phi_{Z_n}(t) &= \phi_{S_n}\left(\frac{t}{\sqrt{n}}\right) \\ &= \left(1 - \frac{t^2}{2n} + o_2\left(\frac{t^2}{n}\right)\right)^n.\end{aligned}$$

De acuerdo a (11.28), bastará ver que la sucesión de funciones ϕ_{Z_n} satisfacen

$$\lim_{n \rightarrow \infty} \left(1 - \frac{t^2}{2n} + o_2\left(\frac{t^2}{n}\right)\right)^n = \exp\left(-\frac{t^2}{2}\right). \quad (11.30)$$

Para ello escribamos la sucesión de características del siguiente modo

$$\phi_{Z_n}(t) = \left(1 - \frac{1}{n} \left[\frac{t^2}{2} - o_2\left(\frac{t^2}{n}\right)n\right]\right)^n,$$

y luego si llamamos

$$a_n = \left[\frac{t^2}{2} - o_2\left(\frac{t^2}{n}\right)n\right],$$

entonces resulta

$$\phi_{Z_n}(t) = \left(1 - \frac{a_n}{n}\right)^n.$$

Se conoce del cálculo elemental que si $a \rightarrow_{n \rightarrow \infty} L$ entonces

$$\left(1 - \frac{a_n}{n}\right)^n \rightarrow_{n \rightarrow \infty} \exp(-L).$$

Por lo tanto, para mostrar (11.30) bastará mostrar que en nuestro caso $L = t^2/2$. Equivalentemente bastará con mostrar que

$$\lim_{n \rightarrow \infty} o_2\left(\frac{t^2}{n}\right)n \rightarrow 0.$$

Pero esto resulta de escribir

$$o_2\left(\frac{t^2}{n}\right)n = \frac{o_2\left(\frac{t^2}{n}\right)}{\frac{t^2}{n}} t^2$$

y de observar que como $t^2/n \rightarrow 0$ cuando $n \rightarrow \infty$, de acuerdo a (11.29) se tiene

$$\lim_{n \rightarrow +\infty} \frac{o\left(\frac{t^2}{n}\right)}{\frac{t^2}{n}} = 0.$$

Esto prueba el teorema. \square

Observación. Teniendo en cuenta que

$$E(\bar{X}_n) = \frac{1}{n}n\mu = \mu$$

y

$$\text{Var}(\bar{X}_n) = n\frac{\sigma^2}{n} = \frac{\sigma^2}{n},$$

podemos escribir las variables Z_n de la siguiente manera

$$\begin{aligned} Z_n &= \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\text{Var}(\bar{X}_n)}} \\ &= \sqrt{n}\frac{(\bar{X}_n - \mu)}{\sigma}. \end{aligned}$$

Luego, de acuerdo a (11.27) tenemos

$$n^{\frac{1}{2}}\frac{(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0, 1). \quad (11.31)$$

De acuerdo a la Ley Fuerte de los Grandes Números $\bar{X}_n - \mu \rightarrow 0$ c.s., y por lo tanto también

$$W_n = (\bar{X}_n - \mu)/\sigma \rightarrow 0 \text{ c.s.}$$

Además, recordemos que convergencia casi segura implica convergencia en distribución. Al multiplicar W_n por el factor $n^{\frac{1}{2}}$, de acuerdo a (11.31) deja de tender a 0 y tampoco tiende infinito. Por eso se dice que la velocidad de convergencia de \bar{X}_n a μ es $n^{\frac{1}{2}}$. Se deja como ejercicio probar que si multiplicamos a W_n por $n^{\frac{1}{2}+\delta}$ la sucesión converge a $\pm\infty$ en probabilidad. Es decir que dado cualquier $K > 0$, tendremos

$$\lim_{n \rightarrow \infty} P(n^{\frac{1}{2}+\delta}|W_n| > K) = 1$$

También se deja como ejercicio probar que si multiplicamos a W_n por $n^{\frac{1}{2}-\delta}$ con $\delta > 0$ la sucesión $n^{\frac{1}{2}-\delta}W_n$ converge en probabilidad a 0. El exponente $\frac{1}{2}$ es el la potencia exacta de n por la que hay que multiplicar a W_n para que la sucesión $n^k W_n$ no converja ni a 0 ni a $\pm\infty$.

11.5.2. Teorema Central del Límite para variables no idénticamente distribuidas.

El Teorema Central del Límite sigue valiendo bajo condiciones menos restrictivas. Se puede suprimir la hipótesis de que las distribuciones sean idénticas y aún debilitar la hipótesis de la independencia.

El Teorema de Lindeberg o Teorema Central del Límite Fuerte da una condición suficiente para que una sucesión de variables aleatorias independientes no necesariamente idénticamente distribuidas converja en distribución a la normal estandarizada. Enunciamos este importante teorema sin demostración.

Teorema 11.12 (Teorema Central de Lindeberg) Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias independientes con $E(X_i) = \mu_i$ y $\text{Var}(X_i) = \sigma_i^2$ para todo $i \in \mathbb{N}$, donde $\sigma_i^2 < \infty$ y existe al menos un i_0 tal que $\sigma_{i_0}^2 > 0$. Sea como antes $S_n = \sum_{i=1}^n X_i$ y llamemos

$$s_n^2 = \sum_{i=1}^n \sigma_i^2 = \text{Var}(S_n).$$

Definamos las variables aleatorias centradas

$$Y_i = X_i - \mu_i.$$

Una condición suficiente para que

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} \xrightarrow{\mathcal{D}} N(0, 1)$$

es que para todo $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^n \int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i}}{s_n^2} = 0. \quad (11.32)$$

Demostración. Ver el libro citado de Barry R. James.

Observación. La condición (11.32) se llama *condición de Lindeberg*. Notemos que como $E(Y_i) = 0$ y $\text{Var}(Y_i) = \text{Var}(X_i) = \sigma_i^2$, se tiene

$$s_n^2 = \sum_{i=1}^n \sigma_i^2 \quad (11.33)$$

$$\begin{aligned} &= \sum_{i=1}^n \text{Var}(Y_i) \\ &= \sum_{i=1}^n \int_{-\infty}^{+\infty} y^2 dF_{Y_i} \\ &= \sum_{i=1}^n \int_{\{|y| < s_n \varepsilon\}} y^2 dF_{Y_i} + \sum_{i=1}^n \int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i}. \end{aligned} \quad (11.34)$$

Luego, la condición (11.32) es equivalente a que para todo $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \int_{\{|y| < s_n \varepsilon\}} y^2 dF_{Y_i}}{\sum_{i=1}^n \int_{-\infty}^{\infty} y^2 dF_{Y_i}} = 1, \quad (11.35)$$

lo cual se puede interpretar como que la condición de Lindeberg implica que la contribución de Y_i a la varianza de S_n proviene esencialmente de los valores donde $|Y_i|^2 \leq \varepsilon^2 s_n^2$. Si llamamos $S_n^* = \sum_{i=1}^n Y_i$ como $s_n^2 = \text{Var}(S_n) = \text{Var}(S_n^*)$ resulta que la contribución de Y_i^2 a la $\text{Var}(S_n^*)$ corresponde básicamente a los puntos donde $Y_i^2 < \varepsilon s_n^2$, es decir donde Y_i^2 es pequeña respecto a $E(S_n^{*2})$. Esto está diciendo que con alta probabilidad Y_i^2 es pequeño con respecto a S_n^{*2} . En particular de (11.32) se deduce que para todo $\varepsilon^2 > 0$, existe $n_0(\varepsilon)$ tal que para todo $n \geq n_0$

$$\int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i} < s_n^2 \varepsilon^2$$

para todo $1 \leq i \leq n$. Por otro lado para todo $1 \leq i \leq n$,

$$\int_{\{|y| < s_n \varepsilon\}} y^2 dF_{Y_i} \leq s_n^2 \varepsilon^2.$$

Luego para todo $1 \leq i \leq n$ y $n \geq n_0$ se tiene

$$\sigma_i^2 = \int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i} + \int_{\{|y| < s_n \varepsilon\}} y^2 dF_{Y_i} < 2s_n^2 \varepsilon^2,$$

y por lo tanto, para todo $n \geq n_0$ resulta

$$\frac{\max_{1 \leq i \leq n} \sigma_i^2}{\sum_{i=1}^n \sigma_i^2} < 2\varepsilon^2.$$

Luego

$$\lim_{n \rightarrow \infty} \frac{\max_{1 \leq i \leq n} \sigma_i^2}{\sum_{i=1}^n \sigma_i^2} = 0.$$

Es decir que la varianza de cada variable, sobre la suma de las varianzas tiende a 0.

Del teorema central del límite de Lindeberg se deduce la siguiente versión del Teorema Central del Límite.

Teorema 11.13 (Teorema Central del Límite de Liapunov) Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias independientes con $E(X_i) = \mu_i$ y varianza $\text{Var}(X_i) = \sigma_i^2 < \infty$ tal que para algún i_0 , $\sigma_{i_0}^2 > 0$. Llamemos $Y_i = X_i - \mu_i$ a las variable aleatoria centradas. Una condición suficiente para que

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} \xrightarrow{\mathcal{D}} \text{N}(0, 1)$$

es que exista $\delta > 0$ tal que

$$\lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^n E(|Y_i|^{2+\delta})}{s_n^{2+\delta}} = 0. \quad (11.36)$$

Demostración. Tenemos que

$$\begin{aligned} \int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i} &= \int_{\{|y| \geq s_n \varepsilon\}} \frac{|y|^{2+\delta}}{|y|^\delta} dF_{Y_i} \\ &\leq \frac{1}{\varepsilon^\delta s_n^\delta} \int_{\{|y| \geq s_n \varepsilon\}} |y|^{2+\delta} dF_{Y_i} \\ &\leq \frac{E(|Y_i|^{2+\delta})}{\varepsilon^\delta s_n^\delta} \end{aligned}$$

y luego

$$\sum_{i=1}^n \int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i} \leq \frac{1}{\varepsilon^\delta s_n^\delta} \sum_{i=1}^n E(|Y_i|^{2+\delta}).$$

Dividiendo por s_n^2 se tiene

$$\frac{\sum_{i=1}^n \int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i}}{s_n^2} \leq \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{i=1}^n E(|Y_i|^{2+\delta}),$$

y por lo tanto por la condición (11.36)

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i}}{s_n^2} = 0. \quad (11.37)$$

que es la condición de Lindeberg. \square

Esta condición es útil cuando las variables tienen momentos finitos de orden mayor que dos. La condición (11.36) se denomina la Condición de Liapunov.

Ejemplo. Consideremos ahora una sucesión de variables aleatorias $(Y_n)_{n \geq 1}$, donde Y_n tiene distribución $\text{Bi}(n, p)$. Podemos pensar a Y_n como el número de éxitos en n experimentos independientes realizados bajo las mismas condiciones, donde la probabilidad de éxito es p . Luego podemos escribir

$$Y_n = \sum_{i=1}^n X_i,$$

donde

$$X_i = \begin{cases} 1 & \text{si el } i\text{-ésimo experimento resulta éxito} \\ 0 & \text{si el } i\text{-ésimo experimento resulta fracaso.} \end{cases}$$

Claramente las variables X_i son independientes e idénticamente distribuidas. Sabemos que $P(X_i = 1) = p$ y $P(X_i = 0) = 1 - p$, $E(X_i) = p$ y $\text{Var}(Y_i) = p(1 - p)$. Luego, estamos en condiciones de aplicar el Teorema Central del Límite para variables i.i.d. Entonces

$$\frac{Y_n - E(Y_n)}{\sqrt{\text{Var}(Y_n)}} = \frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{D}} N(0, 1).$$

Se puede probar que para $n = 20$ la distribución normal es una buena aproximación de la binomial, de manera que a fines prácticos se pueden usar tablas normales para calcular probabilidades binomiales, si n es suficientemente grande.

11.5.3. Una Aplicación a la Binomial.

Se realiza una encuesta para determinar el porcentaje p de votantes que va a votar a un partido C determinado. Se toma una muestra al azar de n votantes y se los encuesta acerca de su intención de voto. Designemos mediante X_i a la variable que toma el valor 1, si la intención declarada del encuestado i es votar al partido C y $X_i = 0$ en caso contrario. Claramente $P(X_i = 1) = p$.

La variable

$$S_n = \sum_{i=1}^n X_i$$

da la cantidad de encuestados que dicen votar al partido C . La variable Y_n tiene distribución $\text{Bi}(n, p)$.

Como desconocemos el parámetro p , podemos estimarlo a partir del promedio

$$\hat{p}_n = \bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}.$$

Como $E(X_i) = p$, por la ley de los grandes números tendremos $\bar{X}_n \rightarrow p$ c.s. Lo que queremos saber es cuan grande tiene que ser n para lograr una precisión determinada en nuestra estimación de p con cierta probabilidad. Más precisamente fijemos una cota e para el error de estimación $E_n = \bar{X}_n - p$ (por ejemplo $e = 0,05$) y supongamos que queremos conocer aproximadamente la probabilidad de que $|E_n| \leq e$, es decir $P(|E_n| \leq e)$.

Sabemos que

$$\begin{aligned} Z_n &= \frac{S_n - np}{\sqrt{np(1-p)}} \\ &= \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \\ &= \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{D}} N(0, 1). \end{aligned}$$

Llamemos

$$a_n = \frac{\sqrt{ne}}{\sqrt{p(1-p)}}, \quad (11.38)$$

y Φ a la función de distribución de una variable $N(0,1)$. Luego, como la distribución de Z_n se comporta aproximadamente como la de una $N(0,1)$ para n grande, tenemos

$$\begin{aligned} P(|E_n| \leq e) &= P(|\bar{X}_n - p| \leq e) \\ &= P\left(\sqrt{n} \frac{|\bar{X}_n - p|}{\sqrt{p(1-p)}} \leq \frac{\sqrt{ne}}{\sqrt{p(1-p)}}\right) \\ &= P(|Z_n| \leq a_n) \\ &\cong \Phi(a_n) - \Phi(-a_n) \\ &= \Phi(a_n) - (1 - \Phi(a_n)) \\ &= 2\Phi(a_n) - 1, \end{aligned}$$

donde el signo \cong indica aproximadamente. Supongamos ahora que queremos saber qué tamaño de muestra se requiere para que $P(|E_n| \leq e)$ sea aproximadamente $1 - \alpha$, donde α es un número pequeño, por ejemplo 0,05. Entonces se requerirá un valor n tal que

$$2\Phi(a_n) - 1 = 1 - \alpha,$$

o equivalentemente

$$a_n = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Reemplazando a_n de acuerdo a (11.38) tendremos

$$\frac{\sqrt{ne}}{\sqrt{p(1-p)}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right),$$

o equivalentemente

$$n = \frac{p(1-p) \left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)^2}{e^2}.$$

Como p es desconocido podemos acotar la expresión de la derecha utilizando el valor de p más desfavorable. Hallemos dicho valor. Como n depende en forma creciente de $g(p) = p(1-p)$ deberíamos elegir el máximo de esta función para $0 \leq p \leq 1$. Observemos que $g'(p) = 1 - 2p$, de modo que el único punto crítico es $p = 1/2$, y como $g''(p) = -2 < 0$ corresponde a un máximo relativo. Como en los extremos $g(0) = g(1) = 0$ y $g(1/2) = 1/4$, resulta que el máximo absoluto de g se alcanza en $p = 1/2$ y vale $1/4$. Luego bastá tomar n igual a

$$n = \frac{\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)^2}{4e^2}.$$

Por ejemplo si $e = 0,05$ y $\alpha = 0,05$, buscando en la tabla normal se tendrá que $\Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0,975) = 1,96$, y luego

$$n = \frac{(\Phi^{-1}(1 - \frac{\alpha}{2}))^2}{4e^2} = 384,16.$$

Luego, como n tiene que ser entero, bastará tomar $n = 385$.

El valor n calculado nos asegura la probabilidad deseada, pero dado que se reemplazó $p(1 - p)$ por una cota superior, este valor de n hallado puede ser más grande que el estrictamente necesario. En la Sección siguiente veremos un teorema que nos permitirá reemplazar a $p(1 - p)$ por la estimación $\bar{X}_n(1 - \bar{X}_n)$.

11.6. Teorema de Slutsky.

El siguiente teorema tiene numerosas aplicaciones en Estadística.

Teorema 11.14 (Teorema de Slutsky) Sean $(X_n)_{n \geq 1}$ e $(Y_n)_{n \geq 1}$ dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} c$, donde X es una variable aleatoria y c una constante. Entonces se tiene

(i) $X_n + Y_n \xrightarrow{D} X + c$,

(ii) $X_n Y_n \xrightarrow{D} cX$,

(iii) Si $c \neq 0$ entonces,

$$\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{c}.$$

Para probar el el Teorema 11.14 necesitaremos probar previamente los Teoremas 11.15-11.20.

Teorema 11.15 Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias tales que $X_n \xrightarrow{D} X$ donde X es otra variable aleatoria. Entonces para toda constante $a \in \mathbb{R}$, se tiene $aX_n \xrightarrow{D} aX$.

Demostración. La demostración la haremos distinguiendo tres casos: (i) $a = 0$, (ii) $a > 0$ y (iii) $a < 0$.

(i) Si $a = 0$, entonces es claro que $aX = aX_n = 0$ y por lo tanto el teorema se cumple.

- (ii) Sea $a > 0$. Queremos probar que para todo punto x de continuidad de F_{aX} vale que

$$\lim_{n \rightarrow +\infty} F_{aX_n}(x) = F_{aX}(x).$$

Calculamos la función de distribución de aX_n

$$\begin{aligned} F_{aX_n}(x) &= P(aX_n \leq x) \\ &= P\left(X_n \leq \frac{x}{a}\right) \\ &= F_{X_n}\left(\frac{x}{a}\right), \end{aligned}$$

y de manera análoga, la función de distribución de aX

$$F_{aX}(x) = F_X\left(\frac{x}{a}\right).$$

Entonces x es un punto de continuidad de F_{aX} si y sólo si $\frac{x}{a}$ lo es de F_X . Ahora bien, como $X_n \xrightarrow{\mathcal{D}} X$ vale que para todo x punto de continuidad de F_X

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

En particular eso vale para $\frac{x}{a}$. Esto demuestra el caso (ii) $a > 0$.

- (iii) Sea $a < 0$. Este caso resulta más complicado de probar. Probaremos en primer lugar que vale para $a = -1$ y después pasaremos al caso general. Queremos probar que si $X_n \xrightarrow{\mathcal{D}} X$ entonces $-X_n \xrightarrow{\mathcal{D}} -X$.

En primer lugar es fácil ver que en general si X es una variable aleatoria $P(X < a) = F_X(a^-)$, donde $F_X(a^-)$ es el límite de $F_X(x)$, cuando x tiende a a por la izquierda. Para eso basta con observar que

$$\{X < a\} = \bigcup_{n=1}^{\infty} \left\{X \leq a - \frac{1}{n}\right\}.$$

La sucesión de conjuntos $C_n = \{X \leq a - \frac{1}{n}\}$ es monótona creciente y por lo tanto

$$\begin{aligned} P(X < a) &= \lim_{n \rightarrow \infty} P\left(X \leq a - \frac{1}{n}\right) \\ &= \lim_{n \rightarrow +\infty} F_X\left(a - \frac{1}{n}\right) \\ &= F_X(a^-). \end{aligned}$$

Calcularemos ahora F_{-X} y F_{-X_n} Por un lado

$$\begin{aligned} F_{-X}(x) &= P(-X \leq x) \\ &= P(X \geq -x) \\ &= 1 - P(X < -x) \\ &= 1 - F_X((-x)^-). \end{aligned}$$

Por otro lado y de manera análoga

$$F_{-X_n}(x) = 1 - F_{X_n}((-x)^-).$$

Entonces tenemos que probar que si x es un punto de continuidad de F_{-X} entonces

$$\lim_{n \rightarrow \infty} [1 - F_{X_n}((-x)^-)] = 1 - F_X((-x)^-),$$

o equivalentemente tenemos que probar que si x es un punto de continuidad de F_{-X} entonces

$$\lim_{n \rightarrow \infty} F_{X_n}((-x)^-) = F_X((-x)^-). \quad (11.39)$$

Como F_{-X} está definida como

$$F_{-X}(x) = 1 - F_X((-x)^-),$$

resulta que x es un punto de de continuidad de F_{-X} si y sólo si $-x$ lo es de F_X . Por lo tanto en los puntos donde F_{-X} es continua vale que $F_X((-x)^-) = F_X(-x)$. Por lo tanto (11.39) es equivalente a que

$$\lim_{n \rightarrow \infty} F_{X_n}((-x)^-) = F_X(-x), \quad (11.40)$$

en los puntos x para los cuales $-x$ es un punto de continuidad de F_X . Como x puede ser cualquiera, esto es equivalente a que

$$\lim_{n \rightarrow \infty} F_{X_n}(x^-) = F_X(x), \quad (11.41)$$

para todo punto x que sea de continuidad de F_X .

Por la monotonía de F_{X_n} se tiene que $F_{X_n}(x^-) \leq F_{X_n}(x)$. Entonces tomando límite superior en ambos miembros y recordando que la hipótesis de convergencia en distribución implica que $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ se obtiene

$$\begin{aligned} \overline{\lim} F_{X_n}(x^-) &\leq \overline{\lim} F_{X_n}(x) \\ &= \lim_{n \rightarrow \infty} F_{X_n}(x) \\ &= F_X(x). \end{aligned} \quad (11.42)$$

Observemos que como F_X es continua en x entonces dado $\varepsilon > 0$ existe $\delta > 0$ tal que $F_X(x) - \varepsilon < F_X(x - \delta)$. Como el conjunto de puntos de discontinuidad de F_X es a lo sumo numerable, podemos elegir $x - \delta$ de forma tal que F_X sea continua en $x - \delta$. Por la monotonía de F_{X_n} resulta

$$F_{X_n}(x^-) \geq F_{X_n}(x - \delta).$$

Tomando límite inferior y recordando que $x - \delta$ es un punto de continuidad de F_X se obtiene

$$\begin{aligned} \underline{\lim} F_{X_n}(x^-) &\geq \underline{\lim} F_{X_n}(x - \delta) \\ &= \lim_{n \rightarrow \infty} F_{X_n}(x - \delta) \\ &= F_X(x - \delta) \\ &> F_X(x) - \varepsilon. \end{aligned}$$

Ahora haciendo que $\varepsilon \rightarrow 0$ se tiene

$$\underline{\lim} F_{X_n}(x^-) \geq F_X(x). \quad (11.43)$$

Por lo tanto de (11.42) y (11.43) resulta

$$\overline{\lim} F_{X_n}(x^-) \leq F_X(x) \leq \underline{\lim} F_{X_n}(x^-).$$

Pero como siempre ocurre que $\underline{\lim} F_{X_n}(x^-) \leq \overline{\lim} F_{X_n}(x^-)$, resulta que

$$\overline{\lim} F_{X_n}(x^-) = F_X(x) = \underline{\lim} F_{X_n}(x^-),$$

y entonces necesariamente existe $\lim F_{X_n}(x^-)$ y además

$$\lim F_{X_n}(x^-) = F_X(x).$$

Esto demuestra (11.41).

Ahora probaremos el Teorema para cualquier $a < 0$. Para eso escribimos

$$aX_n = (-a)(-X_n).$$

Entonces por un lado como $X_n \xrightarrow{\mathcal{D}} X$ se tiene que $-X_n \xrightarrow{\mathcal{D}} -X$. Por otro lado si $a < 0$ entonces $-a > 0$ y por el caso (i) $aX_n = (-a)(-X_n) \xrightarrow{\mathcal{D}} (-a)(-X) = aX$. \square

Definición 11.8 Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias. Decimos que la sucesión está acotada uniformemente en probabilidad si dado $\varepsilon > 0$ existe $K > 0$ tal que

$$P(|X_n| \leq K) \geq 1 - \varepsilon.$$

Observación. Recordemos que hemos probado, en el Teorema 10.6 en la página 201 que si $X_n \xrightarrow{P} X$ entonces dado $\varepsilon > 0$ existe $K > 0$ tal que para todo $n \in \mathbb{N}$

$$P(|X_n| \leq K) \geq 1 - \varepsilon$$

y

$$P(|X| \leq K) \geq 1 - \varepsilon.$$

Esto significa que si una sucesión $(X_n)_{n \geq 1}$ converge en probabilidad está acotada uniformemente en probabilidad.

Para la convergencia en distribución se tiene un resultado análogo.

Teorema 11.16 *Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias y X otra variable aleatoria tal que $X_n \xrightarrow{D} X$. Entonces dado $\varepsilon > 0$ existe $K_0 > 0$ tal que para todo $n \in \mathbb{N}$*

$$P(|X_n| \leq K_0) \geq 1 - \varepsilon$$

y

$$P(|X| \leq K_0) \geq 1 - \varepsilon.$$

Demostración. Por el Teorema 10.5 sabemos que dado $\varepsilon > 0$ existe $K > 0$ tal que

$$P(|X| \leq K) \geq 1 - \frac{\varepsilon}{2}.$$

Observemos que si para cierto $K > 0$ vale la desigualdad, entonces también vale para cualquier $K_1 > K$. En efecto, como

$$\{|X| \leq K\} \subset \{|X| \leq K_1\},$$

tomando probabilidades se tiene

$$1 - \varepsilon \leq P(|X| \leq K) \leq P(|X| \leq K_1).$$

Luego, como el conjunto de puntos de discontinuidad de F_X es a lo sumo numerable, podemos elegir K de forma tal que F_X sea continua en K y en $-K$. Entonces

$$\begin{aligned} P(|X| \leq K) &= P(-K \leq X \leq K) \\ &= P(-K < X \leq K) \\ &= F_X(K) - F_X(-K) \end{aligned} \tag{11.44}$$

$$\geq 1 - \frac{\varepsilon}{2}. \tag{11.45}$$

Teniendo en cuenta la convergencia en distribución de X_n a X , resulta

$$\lim_{n \rightarrow \infty} F_{X_n}(K) = F_X(K),$$

y

$$\lim_{n \rightarrow \infty} F_{X_n}(-K) = F_X(-K).$$

Por definición de límite existe $n_1 \in \mathbb{N}$ tal que si $n \geq n_1$ entonces

$$F_{X_n}(K) > F_X(K) - \frac{\varepsilon}{4} \quad (11.46)$$

y también $n_2 \in \mathbb{N}$ tal que si $n \geq n_2$ entonces

$$F_{X_n}(-K) < F_X(-K) + \frac{\varepsilon}{4} \quad (11.47)$$

Luego tenemos

$$\begin{aligned} P(|X_n| \leq K) &= P(-K \leq X_n \leq K) \\ &\geq P(-K < X_n \leq K) \\ &= F_{X_n}(K) - F_{X_n}(-K). \end{aligned} \quad (11.48)$$

Sea $n_0 = \max\{n_1, n_2\}$. Luego de (11.44), (11.46), (11.47) y (11.48) resulta que si $n \geq n_0$ se tiene

$$\begin{aligned} P(|X_n| \leq K) &\geq F_{X_n}(K) - F_{X_n}(-K) \\ &> F_X(K) - \frac{\varepsilon}{4} - \left(F_X(-K) + \frac{\varepsilon}{4}\right) \\ &\geq F_X(K) - F_X(-K) - \frac{\varepsilon}{2} \\ &\geq 1 - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} = 1 - \varepsilon. \end{aligned}$$

Luego hemos conseguido la acotación requerida para X y X_n con $n \geq n_0$. Finalmente para cada $1 \leq j \leq n_0 - 1$, podemos encontrar un número $K_j > 0$ tal que $P(|X_j| \leq K_j) \geq 1 - \varepsilon$. Entonces si ponemos

$$K_0 = \max\{K, K_1, K_2, \dots, K_{n_0-1}\}$$

se cumple

$$P(|X_n| \leq K_0) \geq 1 - \varepsilon, \quad \forall n$$

y

$$P(|X| \leq K_0) \geq 1 - \varepsilon. \quad \square$$

Teorema 11.17 *Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias uniformemente acotada en probabilidad y supongamos que $Y_n \xrightarrow{P} 0$, entonces*

$$X_n Y_n \xrightarrow{P} 0.$$

Demostración. Utilizado las dos hipótesis dado $\varepsilon > 0$ existe $K > 0$ tal que

$$P(|X_n| \leq K) \geq 1 - \frac{\varepsilon}{2}$$

y $n_0 \in \mathbb{N}$ tal que para todo $n \geq n_0$ se tiene

$$P\left(|Y_n| \geq \frac{\varepsilon}{2K}\right) < \frac{\varepsilon}{2}.$$

Ahora observemos que

$$\{|X_n Y_n| > \varepsilon\} \subset \{|X_n| > K\} \cup \{|Y_n| \geq \frac{\varepsilon}{K}\},$$

ya que si $|X_n| \leq K$ y $|Y_n| < \varepsilon/K$ entonces $|X_n Y_n| \leq \varepsilon$.

Tomando probabilidades tenemos que para todo $n \geq n_0$ resulta

$$\begin{aligned} P(\{|X_n Y_n| > \varepsilon\}) &\leq P(\{|X_n| > K\}) + P\left(\{|Y_n| \geq \frac{\varepsilon}{K}\}\right) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Esto prueba el teorema. \square

Teorema 11.18 Sean $(X_n)_{n \geq 1}$ e $(Y_n)_{n \geq 1}$ dos sucesiones de variables aleatorias y X otra variable aleatoria tal que $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} 0$. Entonces

$$X_n + Y_n \xrightarrow{D} X.$$

Demostración.

Queremos probar que si x es un punto de continuidad de F_X entonces

$$\lim_{n \rightarrow +\infty} F_{X_n + Y_n}(x) = F_X(x).$$

Sea $\varepsilon > 0$. Dado que el número de puntos de discontinuidad de F_X es a lo sumo numerable, siempre podemos elegir $0 < \varepsilon_1 < \varepsilon$ tal que $x + \varepsilon_1$ sea punto de continuidad de F_X . Luego tenemos

$$\{X_n + Y_n \leq x\} \subset \{X_n \leq x + \varepsilon_1\} \cup \{|Y_n| > \varepsilon_1\}$$

pues si $X_n > x + \varepsilon_1$ y $|Y_n| \leq \varepsilon_1$ entonces $X_n + Y_n > x$.

Tomando probabilidades en ambos miembros obtenemos

$$F_{X_n + Y_n}(x) \leq F_{X_n}(x + \varepsilon_1) + P(|Y_n| > \varepsilon_1). \quad (11.49)$$

Como

$$\lim_{n \rightarrow \infty} F_{X_n}(x + \varepsilon_1) = F_X(x + \varepsilon_1),$$

y

$$\lim_{n \rightarrow \infty} P(|Y_n| > \varepsilon_1) = 0,$$

tomando límite superior en (11.49) se obtiene

$$\begin{aligned} \overline{\lim} F_{X_n+Y_n}(x) &\leq \overline{\lim} [F_{X_n}(x + \varepsilon_1) + P(|Y_n| > \varepsilon_1)] \\ &= \lim_{n \rightarrow \infty} F_{X_n}(x + \varepsilon_1) + \lim_{n \rightarrow \infty} P(|Y_n| > \varepsilon_1) \\ &= F_X(x + \varepsilon_1) \\ &\leq F_X(x + \varepsilon). \end{aligned}$$

Haciendo $\varepsilon \rightarrow 0$ resulta

$$\overline{\lim} F_{X_n+Y_n}(x) \leq F_X(x). \quad (11.50)$$

Tomemos ahora $0 < \varepsilon_1 < \varepsilon$ y tal que $x - \varepsilon_1$ sea un punto de continuidad de F_X . Observemos que también vale

$$\{X_n \leq x - \varepsilon_1\} \subset \{X_n + Y_n \leq x\} \cup \{|Y_n| > \varepsilon_1\},$$

ya que $X_n + Y_n > x$ y $|Y_n| \leq \varepsilon$ equivale a $X_n + Y_n > x$ y $-\varepsilon \leq -Y_n \leq \varepsilon$ de manera que sumando obtenemos $X_n > x - \varepsilon$.

Tomando probabilidades resulta

$$F_{X_n}(x - \varepsilon_1) \leq F_{X_n+Y_n}(x) + P(|Y_n| > \varepsilon_1),$$

y pasando al límite inferior, como $x - \varepsilon_1$ es un punto de continuidad de F_X se obtiene

$$F_X(x - \varepsilon_1) \leq \underline{\lim} F_{X_n+Y_n}(x).$$

Además, como

$$F_X(x - \varepsilon) \leq F_X(x - \varepsilon_1),$$

resulta

$$F_X(x - \varepsilon) \leq \underline{\lim} F_{X_n+Y_n}(x).$$

Luego tomando límite cuando $\varepsilon \rightarrow 0$, y dado que F_X es continua en x , tenemos

$$F_X(x) \leq \underline{\lim} F_{X_n+Y_n}(x). \quad (11.51)$$

De (11.50) y (11.51) se obtiene

$$\overline{\lim} F_{X_n+Y_n}(x) \leq F_X(x) \leq \underline{\lim} F_{X_n+Y_n}(x),$$

y esto implica que

$$\lim_{n \rightarrow \infty} F_{X_n+Y_n}(x) = F_X(x). \quad \square$$

Teorema 11.19 Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias y X otra variable aleatoria tal que $X_n \xrightarrow{\mathcal{D}} X$. Si a es constante, entonces

$$X_n + a \xrightarrow{\mathcal{D}} X + a.$$

Demostración. Tenemos

$$\begin{aligned} F_{X_n+a}(x) &= P(X_n + a \leq x) \\ &= P(X_n \leq x - a) \\ &= F_{X_n}(x - a), \end{aligned}$$

y

$$\begin{aligned} F_{X+a}(x) &= P(X + a \leq x) \\ &= P(X \leq x - a) \\ &= F_X(x - a). \end{aligned}$$

Por lo tanto si x es un punto de continuidad de F_{X+a} entonces $x - a$ es un punto de continuidad de F_X de manera que aplicando la hipótesis y lo anterior resulta

$$\begin{aligned} \lim_{n \rightarrow +\infty} F_{X_n+a}(x) &= \lim_{n \rightarrow +\infty} F_{X_n}(x - a) \\ &= F_X(x - a) \\ &= F_{X+a}(x). \quad \square \end{aligned}$$

Teorema 11.20 Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias tal que $X_n \xrightarrow{P} c$, donde c es una constante. Luego si g es una función medible continua en c , se tiene

$$Y_n = g(X_n) \xrightarrow{P} g(c).$$

Demostración. Dado $\varepsilon > 0$ existe $\delta > 0$ tal que $|x - c| \leq \delta$ implica $|g(x) - g(c)| \leq \varepsilon$. Luego

$$\{|g(x) - g(c)| > \varepsilon\} \subset \{|x - c| > \delta\}.$$

En particular

$$\{|g(X_n) - g(c)| > \varepsilon\} \subset \{|X_n - c| > \delta\}.$$

y tomando probabilidades y límites obtenemos

$$\lim_{n \rightarrow \infty} P(|g(X_n) - g(c)| > \varepsilon) \leq \lim_{n \rightarrow \infty} P(|X_n - c| > \delta) = 0.$$

Luego

$$\lim_{n \rightarrow \infty} P(|g(X_n) - g(c)| > \varepsilon) = 0,$$

y el teorema queda probado. \square

Ahora estamos en condiciones de probar el Teorema de Slutsky, enunciado en la página 242.

Demostración.

(i) Podemos escribir

$$X_n + Y_n = (X_n + c) + (Y_n - c).$$

Sabemos por el Teorema 11.19 que

$$X_n + c \xrightarrow{\mathcal{D}} X + c,$$

e

$$Y_n - c \xrightarrow{P} 0.$$

y aplicando el Teorema 11.18

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

(ii) Escribimos el producto de la siguiente manera

$$X_n Y_n = cX_n + (Y_n - c)X_n.$$

Sean

$$Z_n = (Y_n - c)X_n,$$

y

$$U_n = cX_n.$$

Por un lado sabemos que $(Y_n - c) \xrightarrow{P} 0$ y que la sucesión $(X_n)_{n \geq 1}$ está uniformemente acotada en probabilidad, entonces aplicando el Teorema 11.17 se tiene que

$$Z_n \xrightarrow{P} 0,$$

y aplicando el Teorema 11.15

$$U_n \xrightarrow{\mathcal{D}} cX.$$

Finalmente, aplicando el Teorema 11.18

$$X_n Y_n = U_n + Z_n \xrightarrow{\mathcal{D}} cX.$$

(iii) Como $c \neq 0$ y la función $g(y) = 1/y$ es continua en $y = c$, resulta por el Teorema 11.20 que

$$\frac{1}{Y_n} \xrightarrow{P} \frac{1}{c}.$$

Luego como

$$\frac{X_n}{Y_n} = \left(\frac{1}{Y_n} \right) X_n.$$

(iii) resulta aplicando (ii). \square

Para ver cómo se usa el Teorema de Slutsky en casos particulares, retomemos la aplicación del Teorema Central del Límite a la binomial, presentada en la sección 11.5.3.

Sea

$$X_i = \begin{cases} 1 & \text{si el } i\text{ésimo encuestado declara votar al partido } C \\ 0 & \text{si el } i\text{ésimo encuestado declara no votar al partido } C \end{cases}$$

y sea $P(X_i = 1) = p$, siendo p el parámetro de interés que es desconocido. Luego habíamos demostrado que

$$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}} = \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{D}} N(0, 1),$$

donde

$$Y_n = \sum_{i=1}^n X_i, \quad \bar{X}_n = \frac{Y_n}{n}$$

Por la Ley Débil de los Grandes Números sabemos que

$$\bar{X}_n \xrightarrow{P} p.$$

Como la función $g(p) = p(1-p)$ es continua, por el Teorema 10.7 resulta que

$$\bar{X}_n(1 - \bar{X}_n) \xrightarrow{P} p(1-p).$$

Luego resulta que

$$\sqrt{n} \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1 - \bar{X}_n)}} \xrightarrow{\mathcal{D}} N(0, 1).$$

Ahora daremos una aplicación de estos conceptos resolviendo el siguiente problema de Estadística.

11.7. Aplicación a intervalos de confianza.

Problema: Sea X una variable aleatoria cuya función de distribución F desconocemos. Por ejemplo, puede tratarse del “peso de una lata de arvejas” que es una variable aleatoria que varía de lata en lata. La distribución de X no tiene por qué ser normal. Sean $\mu = E(X)$ y $\sigma^2 = \text{Var}(X)$ parámetros que dependen de F y que supondremos desconocidos. En estadística se los denomina parámetros poblacionales. Se toma una muestra aleatoria de tamaño n y se obtienen las variables aleatorias X_1, X_2, \dots, X_n . Estas variables serán independientes e idénticamente distribuidas con distribución F . El problema consiste en estimar el parámetro desconocido μ a partir de las variables observadas y precisar la magnitud del error que se puede estar cometiendo.

Como por la ley fuerte de los grandes números se tiene que $\bar{X}_n \rightarrow \mu$ c.s., podemos tomar como estimación del parámetro μ el promedio aritmético de la muestra, es decir, \bar{X}_n

$$\hat{\mu}_n = \bar{X}_n.$$

Para n grande este valor estará próximo a la media verdadera μ , y el error cometido en esta aproximación será

$$E_n = \bar{X}_n - \mu.$$

Así, el error resulta una variable aleatoria. Un problema natural es tratar de encontrar, para un valor de n determinado, una cota para el módulo del error, con una alta probabilidad.

Teniendo en cuenta que la varianza se define $\sigma^2 = E(X^2) - [E(X)]^2$ podemos estimar la varianza de la siguiente manera

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \frac{(\sum_{i=1}^n X_i)^2}{n}.$$

Usando la ley de los grandes números se tendrá que

$$\frac{\sum_{i=1}^n X_i^2}{n} \rightarrow E(X^2) \quad \text{c.s.},$$

y

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \rightarrow E(X) \quad \text{c.s.}$$

Luego como el cuadrado es una función continua se tendrá

$$\hat{\sigma}_n^2 \rightarrow E(X^2) - E^2(X) = \sigma^2 \quad \text{c.s.},$$

y por lo tanto,

$$\hat{\sigma}_n \rightarrow \sigma \quad \text{c.s.}$$

y

$$\hat{\sigma}_n \xrightarrow{P} \sigma.$$

Por el Teorema Central del Límite

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{D}} N(0, 1). \quad (11.52)$$

Como sabemos que $\hat{\sigma}_n \xrightarrow{P} \sigma$, se tendrá

$$\frac{\sigma}{\hat{\sigma}_n} \xrightarrow{P} 1. \quad (11.53)$$

Luego teniendo en cuenta (11.52) y (11.53), y aplicando el teorema de Slutsky resulta

$$Z_n = \sqrt{n} \frac{\bar{X}_n - \mu}{\hat{\sigma}_n} = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \frac{\sigma}{\hat{\sigma}_n} \xrightarrow{\mathcal{D}} N(0, 1).$$

Es decir, si se reemplaza a σ por su estimador $\hat{\sigma}_n$ en (11.52), la convergencia en distribución no cambia.

Ahora consideremos un valor α , $0 < \alpha < 1$ que en estadística recibe el nombre de nivel de significación, generalmente se toma $\alpha = 0,01$ o bien $\alpha = 0,05$. Buscamos en la tabla de la distribución normal un valor $z_{\alpha/2}$ tal que $P(Z > \alpha/2) = \alpha/2$ donde Z es una variable $N(0, 1)$. Luego por simetría también se tendrá $P(Z < -z_{\alpha/2}) = \frac{\alpha}{2}$.

Ahora bien si $Z_n \xrightarrow{\mathcal{D}} Z$ con $Z \sim N(0, 1)$ entonces también $-Z_n \xrightarrow{\mathcal{D}} -Z$. Como $-Z$ también es $N(0, 1)$ tenemos que para n grande

$$P(-z_{\alpha/2} \leq -Z_n \leq z_{\alpha/2}) \approx 1 - \alpha,$$

donde \approx indica aproximadamente es decir

$$P\left(-z_{\alpha/2} \leq \sqrt{n} \frac{\mu - \bar{X}_n}{\hat{\sigma}_n} \leq z_{\alpha/2}\right) \approx 1 - \alpha,$$

y despejando

$$P\left(\bar{X}_n - \frac{z_{\alpha/2} \hat{\sigma}_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{z_{\alpha/2} \hat{\sigma}_n}{\sqrt{n}}\right) \approx 1 - \alpha. \quad (11.54)$$

Luego fijando α se puede garantizar que la probabilidad de que μ se encuentre en el intervalo de extremos aleatorios

$$\left[\bar{X}_n - \frac{z_{\alpha/2} \hat{\sigma}_n}{\sqrt{n}}; \bar{X}_n + \frac{z_{\alpha/2} \hat{\sigma}_n}{\sqrt{n}}\right].$$

es aproximadamente $1 - \alpha$. Este intervalo se llama *intervalo de confianza para μ* . Obsérvese que hay dos parámetros que pueden variar, el nivel de

significación α y el tamaño de la muestra n . Cuando α decrece $z_{\alpha/2}$ aumenta y consecuentemente aumenta la longitud intervalo de confianza. Como contrapartida también aumenta la probabilidad que contenga a μ . En cambio cuando n crece y se mantiene el α constante, la longitud del intervalo decrece, tendiendo a 0 cuando n tiende a infinito.

Obsérvese que otra manera de escribir (11.54) es la siguiente

$$P\left(|E_n| \leq \frac{z_{\alpha/2}\hat{\sigma}_n}{\sqrt{n}}\right) \approx 1 - \alpha.$$

Es decir, tenemos acotado el error $|E_n|$ por $z_{\alpha/2}\hat{\sigma}_n/\sqrt{n}$ con probabilidad aproximada $1 - \alpha$.

11.8. Un teorema útil de Convergencia en Distribución

En primer lugar recordemos que si $(X_n)_{n \geq 1}$ es una sucesión de variables aleatorias i.i.d entonces

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{D}} N(0, 1),$$

o equivalentemente por el Teorema 11.15

$$\sqrt{n} (\bar{X}_n - \mu) \xrightarrow{\mathcal{D}} N(0, \sigma^2).$$

Sea g una función continua en μ . Parece natural preguntarse si $\sqrt{n}(g(\bar{X}_n) - g(\mu))$ converge en distribución y en caso de que así sea a qué distribución converge. El siguiente teorema responde esta pregunta.

Teorema 11.21 *Sea $(Y_n)_{n \geq 1}$ una sucesión de variables aleatorias y $(a_n)_{n \geq 1}$ una sucesión de números reales tal que $a_n \rightarrow \infty$. Supongamos que la sucesión de variables aleatorias $a_n(Y_n - \mu) \xrightarrow{\mathcal{D}} X$. Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función con derivada continua en un entorno de μ .*

(i) *Entonces*

$$W_n = a_n (g(Y_n) - g(\mu)) \xrightarrow{\mathcal{D}} g'(\mu) X.$$

(ii) *Si $X \sim N(0, \sigma^2)$ entonces $g'(\mu) X \sim N(0, [g'(\mu)]^2 \sigma^2)$. Este resultado vale aún cuando $g'(\mu) = 0$ si la distribución $N(0, 0)$ se interpreta como la distribución de la variable constantemente igual a cero.*

Demostración.

- (i) Por el Teorema 11.16, la sucesión $a_n (Y_n - \mu)$ está uniformemente acotada en probabilidad. Si consideramos la sucesión $(a_n)_{n \geq 1}$ de números reales como una sucesión de variables aleatorias constantes, es claro que

$$\frac{1}{a_n} \xrightarrow{P} 0.$$

Luego de acuerdo al Teorema 11.17 resulta

$$(Y_n - \mu) = \frac{1}{a_n} (a_n (Y_n - \mu)) \xrightarrow{P} 0,$$

o equivalentemente

$$Y_n \xrightarrow{P} \mu.$$

Como g es continua y derivable en un entorno de μ podemos aplicar el Teorema del Valor Medio y encontrar un punto intermedio ξ_n entre Y_n y μ tal que

$$W_n = a_n g'(\xi_n) (Y_n - \mu).$$

Además como $Y_n \xrightarrow{P} \mu$ resulta que la sucesión de variables aleatorias $(\xi_n)_{n \geq 1}$ también satisface $\xi_n \xrightarrow{P} \mu$. Por la continuidad de g' y el Teorema 11.20 se tiene

$$g'(\xi_n) \xrightarrow{P} g'(\mu).$$

Aplicando la parte (ii) del Teorema de Slutsky se obtiene

$$W_n = g'(\xi_n) Z_n \rightarrow g'(\mu) X.$$

- (ii) Se deduce de (i) pues si $X \sim N(0, \sigma^2)$, entonces $g'(\mu) X \sim N(0, [g'(\mu)]^2 \sigma^2)$. \square

Capítulo 12

Procesos de Poisson.

12.1. Procesos de punto.

Supongamos que se observan sucesos que ocurren a lo largo del tiempo en forma aleatoria. Por ejemplo, los sucesos pueden ser la llegada de clientes a un negocio, las llamadas telefónicas que llegan a una central, la emisión de partículas que realiza un cierto material radioactivo, etc.

Más formalmente, para cada valor $t \geq 0$, denominemos $X(t)$ la cantidad de sucesos que ocurrieron desde un instante inicial 0 hasta t . Luego, supondremos que para cada t , $X(t)$ es una variable aleatoria que toma valores enteros no negativos. Además tendremos naturalmente que $X(0) = 0$, y que si $t_1 < t_2$, entonces $X(t_1) \leq X(t_2)$. Todas las variables aleatorias $X(t)$, $t \geq 0$ estarán definidas sobre un mismo espacio de probabilidad (Ω, \mathcal{A}, P) , pero como la construcción de este espacio es sumamente complicada no daremos detalles sobre el mismo. Digamos solamente que un posible espacio muestral puede estar dado por

$$\Omega = \{\omega : \mathbb{R}_{\geq 0} \rightarrow \mathbb{N}_{\geq 0} : \omega \text{ es no decreciente y continua a derecha}\}.$$

Luego X puede pensarse entonces dependiendo de $t \in \mathbb{R}_{\geq 0}$ y $\omega \in \Omega$, $X(t) |_{\omega} = X(t, \omega) = \omega(t)$

Los procesos $X(t)$ que miden la cantidad de sucesos que ocurren hasta el tiempo t , se denominan *procesos de punto*.

12.2. Axiomática de los Procesos de Poisson

Los procesos de Poisson, son procesos de punto particulares que satisfacen los siguientes cuatro axiomas.

A1. Homogeneidad.

Supongamos que $0 \leq t_1 < t_2$, $0 \leq t_3 < t_4$ y además $t_4 - t_3 = t_2 - t_1$. Entonces las variables aleatorias

$$X(t_2) - X(t_1) \text{ y } X(t_4) - X(t_3)$$

tienen la misma distribución. Observando que $X(t_2) - X(t_1)$ es el número de sucesos que ocurrieron entre t_1 y t_2 , este axioma significa que la distribución del número de sucesos ocurridos en un período de tiempo, depende sólo de la longitud de ese período.

A2. Independencia.

Consideremos dos periodos de tiempo esencialmente disjuntos (a lo sumo pueden tener en común un punto) $[t_1, t_2]$, $[t_3, t_4]$, $t_1 < t_2 \leq t_3 < t_4$. Entonces las variables aleatorias

$$X(t_2) - X(t_1) \text{ y } X(t_4) - X(t_3)$$

son independientes. Esto significa que el número de sucesos que ocurre en un período de tiempo de tiempo $[t_1, t_2]$ es independiente del número de sucesos que ocurre en el período $[t_3, t_4]$, donde $t_3 \geq t_2$. Luego el hecho de tener información sobre el número de sucesos del período $[t_1, t_2]$ no aporta datos que ayuden a predecir el número de sucesos del período $[t_3, t_4]$. Los períodos considerados no tienen por qué ser de igual longitud.

Los axiomas A3 y A4 son de carácter más técnico que los anteriores.

A3. Sea

$$g_1(t) = P(X(t) = 1),$$

entonces

$$g_1'(0) = \lambda > 0,$$

es decir

$$\lim_{t \rightarrow 0} \frac{P(X(t) = 1)}{t} = \lambda > 0.$$

Esto es equivalente a que

$$P(X(t) = 1) = \lambda t + o_1(t), \quad (12.1)$$

donde

$$\lim_{t \rightarrow 0} \frac{o_1(t)}{t} = 0. \quad (12.2)$$

A4.

$$\lim_{t \rightarrow 0} \frac{P(X(t) > 1)}{t} = 0,$$

o equivalentemente existe $o_2(t)$ tal que

$$P(X(t) > 1) = o_2(t), \quad (12.3)$$

donde o_2 satisface

$$\lim_{t \rightarrow 0} \frac{o_2(t)}{t} = 0. \quad (12.4)$$

Para modelar un proceso real como un proceso de Poisson se requiere de la verificación de este conjunto de axiomas. Existen muchos procesos concretos que no responden a este modelo.

12.3. Distribución de un proceso de Poisson.

El siguiente teorema caracteriza la distribución de los procesos de Poisson.

Teorema 12.1 *Si $X(t)$ es un proceso de punto que satisface A1, A2, A3 y A4 entonces $X(t)$ tiene distribución de Poisson con parámetro λt , es decir $X(t) \sim P(\lambda t)$.*

Demostración. Para cada n dividimos el intervalo $[0, t]$ en n subintervalos de igual longitud que denominaremos I_i^n , $1 \leq i \leq n$. Más precisamente consideramos la partición regular del intervalo $[0, t]$ con $n + 1$ puntos

$$\pi^n = \left\{ 0, \frac{t}{n}, \frac{2t}{n}, \dots, \frac{(n-1)t}{n}, t \right\}.$$

Esta partición determina n subintervalos

$$I_i^n = \left[\frac{(i-1)t}{n}, \frac{it}{n} \right], 1 \leq i \leq n.$$

El número de sucesos que ocurre en I_i^n es

$$V_i^n = X\left(\frac{it}{n}\right) - X\left(\frac{(i-1)t}{n}\right).$$

Por A1, las variables V_i^n , $1 \leq i \leq n$, tienen la misma distribución que $X(t/n) = V_1^n$ y por el axioma A2 son independientes.

Para cada i definimos el vector aleatorio

$$\mathbf{Z}_i^n = (Z_{i1}^n, Z_{i2}^n, Z_{i3}^n)$$

donde

$$Z_{i1}^n = \begin{cases} 1 & \text{si } V_i^n = 0 \\ 0 & \text{si } V_i^n \neq 0, \end{cases}$$

$$Z_{i2}^n = \begin{cases} 1 & \text{si } V_i^n = 1 \\ 0 & \text{si } V_i^n \neq 1, \end{cases}$$

$$Z_{i3}^n = \begin{cases} 1 & \text{si } V_i^n > 1 \\ 0 & \text{si } V_i^n \leq 1. \end{cases}$$

El evento $Z_{i1}^n = 1$ indica que en el intervalo I_i^n no ocurrió ningún suceso, $Z_{i2}^n = 1$ que ocurrió sólo uno, y $Z_{i3}^n = 1$ que ocurrió más de uno. Es claro que siempre ocurre una y sólo una de esas tres posibilidades y por lo tanto

$$Z_{i1}^n + Z_{i2}^n + Z_{i3}^n = 1.$$

Por otro lado, la distribución del vector \mathbf{Z}_i^n es multinomial, digamos con parámetros de probabilidad p_{1n}, p_{2n}, p_{3n} y para una única repetición. Luego

$$\mathbf{Z}_i^n \sim M(p_{1n}, p_{2n}, p_{3n}, 1),$$

donde

$$p_{1n} = P\left(X\left(\frac{t}{n}\right) = 0\right),$$

$$p_{2n} = P\left(X\left(\frac{t}{n}\right) = 1\right),$$

$$p_{3n} = P\left(X\left(\frac{t}{n}\right) > 1\right).$$

Usando (12.2) y (12.3) resulta

$$p_{2n} = \lambda \frac{t}{n} + o_1\left(\frac{t}{n}\right), \quad (12.5)$$

y

$$p_{3n} = o_2\left(\frac{t}{n}\right). \quad (12.6)$$

Finalmente

$$p_{1n} = 1 - p_{2n} - p_{3n} \quad (12.7)$$

$$= 1 - \lambda \frac{t}{n} - o_1\left(\frac{t}{n}\right) - o_2\left(\frac{t}{n}\right)$$

$$= 1 - \lambda \frac{t}{n} - o_3\left(\frac{t}{n}\right), \quad (12.8)$$

donde

$$o_3(t) = o_1(t) + o_2(t).$$

Claramente, de (12.2) y (12.3) resulta

$$\lim_{t \rightarrow 0} \frac{o_3(t)}{t} = 0. \quad (12.9)$$

Como las variables V_i^n , $1 \leq i \leq n$ son independientes, y como el vector \mathbf{Z}_i^n depende solo de V_i^n , los vectores \mathbf{Z}_i^n , $1 \leq i \leq n$ también son independientes. Ahora definimos las variables

$$\begin{aligned} Y_1^n &= \sum_{i=1}^n Z_{i1}^n, \\ Y_2^n &= \sum_{i=1}^n Z_{i2}^n, \\ Y_3^n &= \sum_{i=1}^n Z_{i3}^n. \end{aligned}$$

Claramente Y_1^n es el número de intervalos en los que no ocurre ningún suceso, Y_2^n es el número de intervalos en los que ocurre exactamente uno e Y_3^n es la cantidad de intervalos en los que ocurre más de un suceso. Luego, la distribución del vector $\mathbf{Y}^n = (Y_1^n, Y_2^n, Y_3^n)$ es multinomial con parámetros de probabilidad p_{1n}, p_{2n}, p_{3n} y n repeticiones. Por lo tanto podemos escribir

$$\mathbf{Y}^n = (Y_1^n, Y_2^n, Y_3^n) \sim M(p_{1n}, p_{2n}, p_{3n}, n).$$

Sea A_n el evento “en ningún intervalo ocurre más de un suceso”. Es decir

$$A_n = \{Y_3^n = 0\}.$$

Veremos que

$$\lim_{n \rightarrow \infty} P(A_n) = 1.$$

o equivalentemente

$$\lim_{n \rightarrow \infty} P(A_n^c) = 0.$$

Observemos que

$$A_n^c = \bigcup_{i=1}^n \{Z_{i3}^n = 1\},$$

pues si en algún intervalo ocurre el suceso más de una vez entonces existe algún i tal que la variable $Z_{i3}^n = 1$ y recíprocamente.

Luego, como $P(Z_{i3}^n = 1) = p_{3n}$, usando (12.6) resulta

$$\begin{aligned} P(A_n^c) &= P\left(\bigcup_{i=1}^n \{Z_{i3}^n = 1\}\right) \\ &\leq \sum_{i=1}^n P(Z_{i3}^n = 1) = np_{3n} = no_2 \left(\frac{t}{n}\right). \end{aligned}$$

Como $t/n \rightarrow 0$ cuando $n \rightarrow \infty$, por (12.4) resulta

$$\lim_{n \rightarrow \infty} P(A_n^c) \leq \lim_{n \rightarrow \infty} \left(\frac{o_2\left(\frac{t}{n}\right)}{\frac{t}{n}} t\right) = t \lim_{n \rightarrow \infty} \left(\frac{o_2\left(\frac{t}{n}\right)}{\frac{t}{n}}\right) = 0. \quad (12.10)$$

Calculemos ahora la probabilidad de que hasta el momento t hayan ocurrido k sucesos. Tenemos

$$P(X(t) = k) = P(\{X(t) = k\} \cap A_n) + P(\{X(t) = k\} \cap A_n^c).$$

Pasando al límite y teniendo en cuenta (12.10) resulta

$$\lim_{n \rightarrow +\infty} P(\{X(t) = k\} \cap A_n^c) = 0,$$

y entonces

$$P(X(t) = k) = \lim_{n \rightarrow +\infty} P(\{X(t) = k\} \cap A_n).$$

Pero es claro que el evento $\{X(t) = k\} \cap A_n$ se caracteriza por

$$\{X(t) = k\} \cap A_n = \{Y_1^n = n - k, Y_2^n = k, Y_3^n = 0\},$$

y luego

$$P(X(t) = k) = \lim_{n \rightarrow +\infty} P(Y_1^n = n - k, Y_2^n = k, Y_3^n = 0).$$

Teniendo en cuenta que la distribución del vector \mathbf{Y}^n es $\mathcal{M}(p_{1n}, p_{2n}, p_{3n}, n)$, obtenemos

$$\begin{aligned} P(X(t) = k) &= \lim_{n \rightarrow +\infty} \frac{n!}{(n-k)!k!} p_{1n}^{n-k} p_{2n}^k p_{3n}^0 \\ &= \frac{1}{k!} \lim_{n \rightarrow +\infty} \left(\prod_{i=1}^k (n-i+1) \right) \\ &\quad \cdot \left(1 - \lambda \frac{t}{n} + o_3 \left(\frac{t}{n} \right) \right)^{n-k} \left(\lambda \frac{t}{n} + o_1 \left(\frac{t}{n} \right) \right)^k \end{aligned}$$

Como

$$\left(\lambda \frac{t}{n} + o_1 \left(\frac{t}{n} \right) \right)^k = \frac{1}{n^k} \left(\lambda t + n o_1 \left(\frac{t}{n} \right) \right)^k,$$

tenemos

$$\begin{aligned} P(X(t) = k) &= \frac{1}{k!} \lim_{n \rightarrow +\infty} \left(\prod_{i=1}^k \frac{(n-i+1)}{n} \right) \\ &\quad \cdot \left(1 - \lambda \frac{t}{n} + o_3 \left(\frac{t}{n} \right) \right)^{n-k} \left(\lambda t + n o_1 \left(\frac{t}{n} \right) \right)^k, \end{aligned}$$

o bien

$$P(X(t) = k) = \frac{1}{k!} \lim_{n \rightarrow \infty} B_n C_n D_n E_n, \quad (12.11)$$

donde

$$\begin{aligned}
 B_n &= \prod_{i=1}^k \frac{n-i+1}{n} \\
 C_n &= \left(1 - \lambda \frac{t}{n} + o_3\left(\frac{t}{n}\right)\right)^n \\
 D_n &= \left(1 - \lambda \frac{t}{n} + o_3\left(\frac{t}{n}\right)\right)^{-k} \\
 E_n &= \left(\lambda t + n o_1\left(\frac{t}{n}\right)\right)^k.
 \end{aligned}$$

Comencemos calculando el límite de B_n

$$\begin{aligned}
 \lim_{n \rightarrow +\infty} B_n &= \lim_{n \rightarrow +\infty} \prod_{i=1}^k \frac{n-i+1}{n} \\
 &= \prod_{i=1}^k \left(\lim_{n \rightarrow +\infty} \frac{n-i+1}{n} \right) \\
 &= \prod_{i=1}^k \left(1 - \lim_{n \rightarrow +\infty} \left(\frac{i-1}{n} \right) \right) \\
 &= 1^k = 1.
 \end{aligned} \tag{12.12}$$

El límite de C_n se puede calcular de la siguiente manera

$$\begin{aligned}
 \lim_{n \rightarrow +\infty} C_n &= \lim_{n \rightarrow +\infty} \left(1 - \lambda \frac{t}{n} + o_3\left(\frac{t}{n}\right) \right)^n \\
 &= \lim_{n \rightarrow +\infty} \left(1 - \frac{1}{n} \left(\lambda t - n o_3\left(\frac{t}{n}\right) \right) \right)^n \\
 &= \lim_{n \rightarrow +\infty} \left(1 - \frac{a_n}{n} \right)^n.
 \end{aligned}$$

donde

$$a_n = \lambda t - n o_3\left(\frac{t}{n}\right).$$

Como en (12.10) se puede demostrar que

$$\lim_{n \rightarrow \infty} n o_3\left(\frac{t}{n}\right) = 0,$$

y entonces resulta

$$\lim_{n \rightarrow +\infty} a_n = \lambda t.$$

Por lo tanto

$$\begin{aligned}\lim_{n \rightarrow +\infty} C_n &= \lim_{n \rightarrow +\infty} \left(1 - \frac{a_n}{n}\right)^n \\ &= \exp\left(-\lim_{n \rightarrow \infty} a_n\right) \\ &= \exp(-\lambda t).\end{aligned}\tag{12.13}$$

Por otro lado, como $t/n \rightarrow 0$ cuando $n \rightarrow \infty$, y $o_1(t/n) \rightarrow 0$, resulta

$$\lim_{n \rightarrow +\infty} D_n = \lim_{n \rightarrow +\infty} \left(1 - \lambda \frac{t}{n} + o_3\left(\frac{t}{n}\right)\right)^{-k} = 1^{-k} = 1.\tag{12.14}$$

Finalmente, como $\lim_{n \rightarrow +\infty} n o_1(t/n) = 0$, resulta

$$\begin{aligned}\lim_{n \rightarrow +\infty} E_n &= \lim_{n \rightarrow +\infty} \left(\lambda t + n o_1\left(\frac{t}{n}\right)\right)^k \\ &= (\lambda t)^k.\end{aligned}\tag{12.15}$$

Usando (12.11), (12.12), (12.13), (12.14) y (12.15) obtenemos

$$P(\{X(t) = k\}) = \exp(-\lambda t) \frac{(\lambda t)^k}{k!}.$$

Esto prueba que $X(t) \sim P(\lambda t)$. \square

12.4. Tiempos de espera

Sea T_1 la variable aleatoria definida como “el tiempo necesario hasta que ocurra el primer suceso”. Calcularemos ahora su distribución.

Teorema 12.2 T_1 tiene distribución exponencial con parámetro λ , es decir, $\mathcal{E}(\lambda)$.

Demostración.

$$\begin{aligned}F_{T_1}(t) &= P(T_1 \leq t) \\ &= P(X(t) > 0) \\ &= 1 - P(X(t) = 0) \\ &= 1 - \exp(-\lambda t).\end{aligned}$$

Luego $T_1 \sim \mathcal{E}(\lambda)$. \square

Otro problema de interés es la distribución de los tiempos sucesivos de ocurrencia de los sucesos. Definamos T_2 como “el tiempo de espera hasta que

ocurra el segundo suceso” entonces $T_2 - T_1$ tiene la misma distribución que T_1 . No daremos una demostración formal de este hecho. Heurísticamente, este resultado puede justificarse de la siguiente manera. $T_2 - T_1$ es el tiempo de espera para el primer suceso luego del instante T_1 . Como por A1 el proceso es homogéneo, este tiempo de espera debería tener la misma distribución que T_1 . Además como T_1 está determinado por $X(t)$ con $t \leq t_1$ y $T_2 - T_1$ por $X(t)$ con $t > T_1$, por A2, resulta que T_1 es independiente de $T_2 - T_1$.

Definamos ahora T_i como el tiempo de espera hasta que ocurran i sucesos. Luego, un argumento similar puede aplicarse, y tendremos el siguiente teorema que enunciaremos sin demostración.

Teorema 12.3 *Las variables aleatorias $T_1, T_2 - T_1, T_3 - T_2, \dots, T_i - T_{i-1}, \dots$ son i. i. d. con distribución $\mathcal{E}(\lambda)$.*

Corolario 12.1 *El tiempo de espera T_i tiene distribución $\Gamma(i, \lambda)$.*

Demostración. Podemos escribir a la variable T_i como una suma telescópica

$$T_i = T_1 + (T_2 - T_1) + (T_3 - T_2) + \dots + (T_i - T_{i-1}).$$

Recordando que $E(\lambda) = \Gamma(1, \lambda)$ y teniendo en cuenta que T_i una suma de variables independientes todas con distribución $\Gamma(1, \lambda)$ resulta que $T_i \sim \Gamma(i, \lambda)$. \square

12.5. Procesos de Poisson en el plano.

Los procesos de Poisson se pueden generalizar al plano. No vamos a describir estos procesos con detalle, pero daremos una breve presentación. Un ejemplo de este tipo de procesos podría ser los que representan la ubicación de los árboles en un bosque.

Consideramos ahora el plano en vez de la recta. Supongamos que en ciertos puntos del plano ocurren sucesos en forma aleatoria, como por ejemplo la presencia de un árbol. Luego para cada boreliano B del plano tendremos la variable aleatoria $X(B)$ que representa la cantidad de sucesos que han ocurrido en B (por ejemplo, la cantidad de árboles que se encuentran en la región B). Los axiomas de un proceso de Poisson en el plano son los siguientes:

AP1. Homogeneidad.

Dado un boreliano, notemos con A su área. Supongamos que $B_1, B_2 \in \mathcal{B}^2$ son boreleanos del plano tal que $A(B_1) = A(B_2)$ entonces las variables aleatorias

$$X(B_1) \text{ y } X(B_2)$$

tienen la misma distribución. Esto dice que la distribución del número de sucesos que ocurren en una región del plano sólo depende de su área.

AP2. Independencia.

Consideremos dos borelianos del plano esencialmente disjuntos $B_1, B_2 \in \mathcal{B}^2$, es decir tal que $A(B_1 \cap B_2) = 0$. Entonces las variables aleatorias $X(B_1)$ y $X(B_2)$ son independientes. Esto significa que cuando las regiones B_1 y B_2 tienen área en común igual a 0, entonces la información de lo que ocurre en una región B_1 no aporta ninguna información respecto de lo que ocurre en la región B_2 .

AP3.

$$\lim_{A(B) \rightarrow 0} \frac{P(X(B) = 1)}{A(B)} = \lambda > 0,$$

o bien

$$P(X(B) = 1) = \lambda A(B) + o_1(A(B)).$$

AP4.

$$\lim_{A(B) \rightarrow 0} \frac{P(\{X(B) > 1\})}{A(B)} = 0,$$

o equivalentemente existe $o_2(t)$ tal que

$$P(\{X(B) > 1\}) = o_2(A(B)).$$

El siguiente teorema se demuestra de manera totalmente análoga al correspondiente para procesos de Poisson en la recta.

Teorema 12.4 *Si $X(B)$ es un proceso que satisface AP1, AP2, AP3 y AP4 entonces la distribución de $X(B)$ es Poisson con parámetro $\lambda A(B)$, es decir $X(B) \sim P(\lambda A(B))$.*

Supongamos que se elija un punto cualquiera del plano (x_0, y_0) , y sea D_1 la distancia de este punto (x_0, y_0) al punto más cercano donde ocurre un suceso (en el ejemplo, D_1 sería la distancia al árbol más próximo), D_2 la distancia de (x_0, y_0) al punto donde ocurre el segundo suceso más próximo, ..., D_i la distancia de (x_0, y_0) al punto donde ocurre el i -ésimo suceso más próximo. El siguiente teorema nos da la distribución de D_1^2 .

Teorema 12.5 *La distribución de D_1^2 es $\mathcal{E}(\pi\lambda)$.*

Demostración. Sea $d > 0$ y sea C el círculo con centro en (x_0, y_0) y radio $d^{1/2}$. Decir que $D_1 \leq d^{1/2}$ es lo mismo que decir que en C ocurrió algún suceso. Luego

$$\begin{aligned} \{D_1^2 \leq d\} &= \{D_1 \leq d^{1/2}\} \\ &= \{X(C) > 0\} \\ &= \{X(C) = 0\}^c. \end{aligned}$$

Luego tomando probabilidades y teniendo en cuenta que $A(C) = \pi d$

$$\begin{aligned} P(D_1^2 \leq d) &= 1 - P(X(C) = 0) \\ &= 1 - \exp(-\lambda A(C)) \\ &= 1 - \exp(-\lambda \pi d) \end{aligned}$$

y por lo tanto D_1^2 tiene distribución $\mathcal{E}(\pi\lambda)$. \square

El siguiente teorema, del cual no se dará la demostración, es análogo al correspondiente teorema para Procesos de Poisson en la recta.

Teorema 12.6 *Las variables aleatorias $D_1^2, D_2^2 - D_1^2, D_3^2 - D_2^2, \dots, D_i^2 - D_{i-1}^2, \dots$ son i. i. d. con distribución $\mathcal{E}(\pi\lambda)$.*

Como corolario tendremos

Corolario 12.2 *La variable aleatoria D_i^2 tiene distribución $\Gamma(i, \pi\lambda)$.*