

Notas de Probabilidades y Estadística

Capítulos 1 al 6

Víctor J. Yohai

Basadas en apuntes de clase tomados por Alberto Déboli

Diciembre 2003

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 prever 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.

Ejemplos.

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

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 .

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. En este caso la frecuencia con que el resultado esté por ejemplo en un intervalo $[a, b]$ dependerá del atleta.

4. Se elige al azar un alumno de primer grado de un colegio y se anota su peso, x y la altura 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 llamara probabilidad de A y 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 no se podrá 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 están definida su probabilidad. Esto nos lleva al concepto de σ -álgebra.

1.2.1. σ -Algebras.

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

Definición. Una familia de subconjuntos de Ω , $\mathcal{A} \subset \mathcal{P}(\Omega)$, es una σ -álgebra sobre Ω si satisface las siguientes propiedades

1. $\Omega \in \mathcal{A}$.
2. Si $A \in \mathcal{A}$ entonces $A^c \in \mathcal{A}$.
3. Si A_1, \dots, A_n, \dots es una sucesión de elementos de \mathcal{A} entonces $A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Propiedades

P1. Por 1 y 2 es claro que $\emptyset \in \mathcal{A}$.

P2. Si A_1, \dots, A_n son elementos de \mathcal{A} entonces $\bigcup_{i=1}^n A_i \in \mathcal{A}$

Demostración n

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

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

P3. 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 $A = (\bigcup_{i=1}^{\infty} A_i^c)^c /$

P4. Si \mathcal{A} es una σ -álgebra, y A_1, \dots, A_n son elementos de \mathcal{A} entonces $A = \bigcap_{i=1}^n A_i \in \mathcal{A}$.

Se demuestra igual que P2.

P5. 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 $A_1 - A_2 = A_1 \cap A_2^c \in \mathcal{A}$.

P6. La σ -álgebra más chica posible es

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

y la más grande posible es

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

Luego si \mathcal{A} es una σ -álgebra sobre Ω entonces

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

Observación. En el contexto de la teoría de la medida, un elemento de la σ álgebra se llama un conjunto medible.

Retornando a la heurística estará definida la probabilidad para los elementos de una σ -álgebra.

1.2.2. Espacios de Probabilidad.

Definición. 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

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

Notación: \biguplus denota una unión disjunta.

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 1 al espacio total.

3. Si queremos formalizar la idea intuitiva de la probabilidad como límite de la frecuencia es importante observar que la "frecuencia" tiene la propiedad de σ -aditividad. En principio veamos que es *aditiva*.

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

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

La σ aditividad ahora se deduce pasando al límite.

Ejemplos:

1.- 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 esta concentrada en x_0 o bien que el único punto de densidad positiva es x_0 .

2.- Si $\Omega = \{x_1, x_2, \dots, x_n, \dots\}$ $\mathcal{A} = \mathcal{P}(X)$, $0 \leq a_i \leq 1$, $i = 1, 2, \dots$

Supongamos que

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

Definimos: $\forall A \subseteq \Omega$

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

donde $P(\{x_i\}) = a_i$.

En este caso P define una medida de probabilidad y está completamente determinada por el valor o el "peso" asignado a cada singleton $\{x_0\}$.

Propiedades.

P1. $P(\emptyset) = 0$.

Demostración

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

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

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

P2. Sean A_1, \dots, A_n eventos disjuntos. Luego $P(\biguplus_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.

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.

P3. Si $A \in \mathcal{A}$ entonces

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

Demostración. Esto sale teniendo en cuenta que

$$1 = P(\Omega) = P\left(A \biguplus A^c\right) = P(A) + P(A^c).$$

P4. 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

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

se obtiene

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

y de ahí el resultado \square .

5- Si A_1, A_2 son eventos y $A_1 \subset A_2$ entonces

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

y además

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

Demostración

Para ver esto escribimos como unión disjunta a A_2

$$A_2 = A_1 \bigsqcup (A_2 - A_1).$$

Entonces aplicando la aditividad se obtienen los dos resultados

$$P(A_2) = P(A_1) + \underbrace{P(A_2 - A_1)}_{\geq 0} \geq P(A_1). \quad \square$$

P6. 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

Escibimos la unión en forma disjunta

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

Entonces

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

P7. En particular se obtiene de P6 la subaditividad finita

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

Por inducción sale que si $A_i \in A$, $i = 1, 2, \dots, k$ entonces

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

P8.- σ -subaditividad. Sea $(A_n)_{n \geq 1} \subset A$ y $B = \bigcup_{n \geq 1}^\infty A_n$. Entonces

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

Demostración

Podemos disjuntar la sucesión sin alterar la unión. Este es un proceso típico.

Comenzamos definiendo

$$B_j = \bigcup_{i=1}^j A_i \quad j \geq 1.$$

Esta sucesión es creciente para todo $n \in \mathbb{N}$ tal que $B_{n-1} \subset B_n$.

Ahora expresamos la unión de manera disjunta definiendo

$$\begin{aligned} \tilde{B}_0 &= \emptyset, \\ \tilde{B}_1 &= B_1, \\ \tilde{B}_2 &= B_2 - B_1, \\ \tilde{B}_3 &= B_3 - B_2, \\ &\dots \\ &\dots \\ \tilde{B}_n &= B_n - B_{n-1}. \end{aligned}$$

Luego

$$A = \bigsqcup_{n \in \mathbb{N}} \tilde{B}_n$$

Por la σ -aditividad y el hecho de que $\tilde{B}_i \subset A_i$, resulta

$$\begin{aligned} P(A) &= \sum_{n=1}^{\infty} P(\tilde{B}_n) = \lim_{k \rightarrow \infty} \sum_{n=1}^k P(B_n - B_{n-1}) = \\ &= \lim_{k \rightarrow \infty} P(B_k) = \lim_{k \rightarrow \infty} P\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{n=1}^{\infty} P(A_i). \quad \square \end{aligned}$$

P9. Si $(A_n)_{n \geq 1}$ es una sucesión de eventos tal que $A_n \subset A_{n+1}$. Si definimos $A = \bigcup_{i=1}^{\infty} A_i$, entonces

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

Demostración.

Como la sucesión es creciente entonces podemos transformar la unión en una unión disjunta definiendo:

$$B_0 = A_0 = \emptyset, \quad B_1 = A_1 - A_0, \quad B_2 = A_2 - A_1, \dots, \quad B_k = A_k - A_{k-1}, \dots$$

Luego

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

y por lo tanto

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

P10. Si $(A_n)_{n \geq 1}$ es una sucesión de eventos tal que $A_n \subset A_{n+1}$. Si definimos $A = \bigcap_{i=1}^{\infty} A_i$, 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

$$1 - P(A) = P(A^c) = \lim_{n \rightarrow +\infty} P(B_n) = \lim_{n \rightarrow +\infty} (1 - P(A_n)) = 1 - \lim_{n \rightarrow +\infty} P(A_n).$$

de donde sale el resultado. \square

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

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

y límite inferior de la sucesión al conjunto

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

Obviamente

$$\underline{A} \subset \bar{A}.$$

Además

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

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

Proposición. Caracterización de los límites superiores e inferiores

(a)

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

(b)

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

Demostración

a) Supongamos que $\omega \in A^\infty$ entonces dado $k \in \mathbb{N}$ existe $n_0 \geq k \in \mathbb{N}$ tal 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 un número finito de conjuntos A_n . Supongamos que A_{n_h} sea el último en el que está, es decir si $n > n_h$ entonces $\omega \notin A_n$ de manera que $\omega \notin \bigcup_{n=n_h+1}^\infty A_n$ y así $\omega \notin \overline{A}$.

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

$$(\underline{A})^c = (\{\omega \in \Omega : \omega \text{ está en infinitos conjuntos } A_n^c\})^c = \\ = \{\omega \in \Omega : \omega \text{ pertenece sólo a un número finito de conjuntos } A_n^c\}.$$

Entonces

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

P11. Dada una sucesión de eventos $(A_n)_{n \geq 1}$, se tiene

(a) $P(\overline{A}) \geq \overline{\lim} P(A_n)$.

(b) $P(\underline{A}) \leq \underline{\lim} P(A_n)$.

(c) Además, en el caso de que $\underline{A} = \overline{A}$ entonces se dice que existe el límite de la sucesión $(A_n)_{n \geq 1}$ y en tal caso

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

a) 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

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

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

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

Luego

$$\begin{aligned} P(\bar{A}) &= \lim P(B_k) = \inf\{P(B_k) : k \geq 1\} \geq_{(*)} \\ &\geq \inf_{k \geq 1} \sup_{i \geq k} \{P(A_i)\} = \overline{\lim} P(A_i). \quad \square \end{aligned}$$

(b) Se deja como ejercicio.

(c) De (a) y (b) tenemos que

$$P(\underline{A}) \leq \underline{\lim} P(A_n) \leq \overline{\lim} P(A_n) \leq P(\bar{A}).$$

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

$$P(\underline{A}) = \underline{\lim} P(A_n) = \overline{\lim} P(A_n) = P(\bar{A}).$$

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

Observaciones

1. En general no se puede tomar como σ -álgebra \mathcal{A} a $\mathcal{P}(\otimes)$ para definir el espacio de de probabilidad. Se prueba en teoría de la medida que bajo ciertas condiciones todo conjunto de medida positiva contiene un subconjunto no medible.

2. Si el conjunto es a lo sumo numerable se puede tomar $\mathcal{A} = \mathcal{P}(\otimes)$.

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

Proposición. Dado un conjunto Ω y una familia \mathfrak{S} de subconjuntos de Ω , existe una σ -álgebra \mathcal{A}^* tal que (i) $\mathfrak{S} \subset \mathcal{A}^*$ y (ii) Si \mathcal{A} es otra σ -álgebra \mathcal{A} tal que $\mathfrak{S} \subset \mathcal{A}$, entonces $\mathcal{A}^* \subset \mathcal{A}$. Es decir \mathcal{A}^* es la menor σ -álgebra que contiene a \mathfrak{S} .

Demostración

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

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

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

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

Primero mostraremos que \mathcal{A}^* es un σ -álgebra. En efecto, $\emptyset \in \mathcal{A}$, para toda $\mathcal{A} \in \mathcal{R}$, luego $\emptyset \in \mathcal{A}^*$.

Además sea una sucesión numerable de eventos $A_1, A_2, \dots, A_n, \dots$ que están en \mathcal{A}^* . Luego cada $A_i \in \mathcal{A}$ para todo $\mathcal{A} \in \mathcal{R}$. Luego $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, para todo $\mathcal{A} \in \mathcal{R}$ y entonces $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}^*$. Esto prueba que \mathcal{A}^* es un σ -álgebra. Por otro lado si \mathcal{A} es una σ -álgebra y $\mathcal{A} \supset \mathfrak{F}$, entonces $\mathcal{A} \in \mathcal{R}$, y esto implica que $\mathcal{A}^* \subset \mathcal{A}$.

Definición del σ -álgebra de Borel en los reales. Sea \mathbb{R} el conjunto de números reales. Luego el σ -álgebra de Borel que denotaremos por \mathcal{B} , es el σ -álgebra generada por todos los conjuntos de la forma $A_x = (-\infty, x]$, donde $x \in \mathbb{R}$. Un conjunto $B \in \mathcal{B}$ se denomina boreliano.

Propiedades.

P1. Todo intervalo $(a, b]$ es un boreliano. Para esto basta con observar que

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

P2. Dado $x \in \mathbb{R}$, $\{x\} \in \mathcal{B}$. Para esto se observa que $\forall n \in \mathbb{N}$

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

Puesto que $x - \frac{1}{n} \rightarrow x$ se obtiene que

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

P3. $(a, b) = (a, b] - \{b\} \in \mathcal{B}$.

P4. $[a, b] = \{a\} \cup (a, b] \in \mathcal{B}$.

P5. $[a, b) = \{a\} \cup (a, b) \in \mathcal{B}$.

P6. Todo abierto es un boreliano

Demostración.

Sea $G \subset \mathbb{R}$ un abierto. $\forall x \in G$ 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 $G \in \mathcal{B}$.

P7. Todo cerrado es un boreliano

Demostración.

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

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

Definición del σ -álgebra de Borel en \mathbb{R}^k . El σ -álgebra de Borel en \mathbb{R}^k es el σ -álgebra 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 a los borelianos en \mathbb{R} , la σ -álgebra de Borel coincide con la generada por los abiertos (cerrados).

P8. Cualquier rectángulo en \mathbb{R}^k es un boreliano.

P9. Todo abierto (cerrado) en \mathbb{R}^k es un boreliano.

Demostración: Se deja como ejercicio.

1.4. Espacios de probabilidad finitos o numerables.

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 . Definimos la *función de densidad* p , asociada a la probabilidad P por

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

de la siguiente manera

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

Se tendrá la siguiente propiedades

P1. La función de densidad determina la función de probabilidad:

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

Demostración. Si $A \subset \Omega$ entonces

$$A = \bigsqcup_{w \in A} \{w\},$$

de manera que

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

P2. Si Ω es finito o numerable, se cumple que

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

Demostración. En efecto por P1

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

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

$$\forall i, j : P(\{\omega_i\}) = P(\{\omega_j\}).$$

Observación.

Un espacio de probabilidad infinito no puede ser equiprobable puesto que en tal caso se escoge un conjunto infinito numerable. En efecto, supongamos que $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$, y $p(\omega) = c$. Luego por P2 se tendría

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

absurdo puesto que $\sum_{i=1}^{\infty} c = \infty$ o 0 según $c > 0$ o $c = 0$.

P3. Si Ω es un espacio de probabilidad finito equiprobable entonces, la probabilidad de cualquier evento A se calcula

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

donde $\#\Omega$ denota el cardinal de Ω .

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

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

y luego,

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

Además

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

Ejemplo.

Hallar la probabilidad de que dado un conjunto de n personas, dos personas cumplan años el mismo día. Suponemos que todos los años tienen 365

días y las probabilidades de nacimiento en cualquier fecha son equiprobables. 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

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

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

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

Por un lado

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

Por otro lado el cardinal de

$$\#A^c = C(365, n)n!.$$

Observación.

La importancia de la combinatoria se ve en este punto; es necesario contar con principios de enumeración.

En este caso, primero seleccionamos los m 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. $A, B \in \mathcal{A}$, tal que $P(B) \neq 0$.

Queremos estudiar la cómo se ve alterada la probabilidad de ocurrencia de A cuando se conoce que el evento B ha ocurrido. En este caso habrá que redefinir el espacio muestral considerando solamente a 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, resultará que la probabilidad 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 condicional 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. Sea (Ω, \mathcal{A}, P) un espacio de probabilidades $A, B \in \mathcal{A}$ tal

que $P(B) > 0$. Se define la *probabilidad condicional* de A dado B por

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

Proposición.

Fijado el evento $B \in \Omega$ podemos definamos $\tilde{P} : \mathcal{A} \rightarrow [0, 1]$ $A \in \mathcal{A}$

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

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 $(A_n)_{n \geq 1} \in \mathcal{A}$ disjuntos

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

1.6. Independencia de eventos.

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

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

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

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

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

Definición. Se dice que 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 tres eventos A_1, A_2 y A_3 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 $\Omega = \{w_1, w_2, w_3, w_4\}$ espacio de probabilidad equiprobable, es decir $P(\{w_i\}) = \frac{1}{4}$

Entonces los conjuntos

$$\begin{aligned}A_1 &= \{w_1, w_2\} \\A_2 &= \{w_1, w_3\} \\A_3 &= \{w_2, w_3\}\end{aligned}$$

son independientes tomados de a dos pero no en forma conjunta.

Más precisamente, se cumple que

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

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

y

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

Pero

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

y por lo tanto

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

Proposición. A_1, \dots, A_k son eventos independientes sii 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

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

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

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

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 $i_1 \neq i_2$, pueden suceder que (a) $P(A_{i_2}) = 0$ o que (b)

$P(A_{i_2}) > 0$. En el caso (a) se tiene que como $A_{i_1} \cap A_{i_2} \subset A_{i_2}$, se tiene que $P(A_{i_1} \cap A_{i_2}) = 0$ y luego

$$P(A_{i_1} \cap A_{i_2}) = 0 = P(A_{i_1})P(A_{i_2}) \quad (1.3)$$

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 (1.3) también vale. 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) Consideremos el caso $P\left(\bigcap_{j=2}^{h+1} A_{i_j}\right) = 0$. En tal caso por la suposición que (1.2) vale para 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) Consideremos el caso $P\left(\bigcap_{j=2}^{h+1} A_{i_j}\right) > 0$. Entonces como estamos suponiendo que (1.1) vale para todo h

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

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. Sea I un conjunto finito o numerable, una sucesión $\{A_i\}_{i \in I}$ se dice una *partición* de Ω si

1.

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

2. Si $i \neq j$ entonces

$$A_i \cap A_j = \emptyset$$

Teorema de la Probabilidad Total. Sea (Ω, \mathcal{A}, P) espacio de probabilidad y $\{A_i\}_{i \in I} \subset \mathcal{A}$ una partición de Ω y $B \in \mathcal{A}$. Entonces

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

Demostración

Como

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

entonces

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

Teorema de Bayes. Sea (Ω, \mathcal{A}, P) e.p y $\{A_i\}_{1 \leq i \leq k}$ una partición de Ω y $B \in \mathcal{A}$.

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 $\{A_i\}_{1 \leq i \leq k}$ es una partición y aplicando la definición de probabilidad condicional se obtiene

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

Ejemplo de aplicación.

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 es sano”

A_2 : el evento “el paciente seleccionado es enfermo ”

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

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

Observación.

Para un test perfecto se pediría

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

Es decir supusimos la existencia de un test “no perfecto”.

Calculemos la probabilidad de que dado que el test detecta enfermedad el paciente sea efectivamente enfermo (esta probabilidad se denomina probabilidad a posteriori)

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

La conclusión es que no se puede tomar ninguna decisión con base en el resultado del test: no sirve.

Si logramos tener

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

la situación cambia; en tal caso resulta $P(A_2|T_+) = 0,91$, 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 dió el tiró tomando origen $(0, 0)$ el punto correspondiente al blanco. Aquí puede interesar solo la distancia al blanco es decir $(x^2 + y^2)^{1/2}$

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

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

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 función medible. Si $(\Omega, \mathcal{A}, \mu)$ es un espacio de medida y X un espacio topológico, $f : \mathcal{A} \rightarrow X$ se dice medible sii para todo $G \in \mathcal{A}$ vale que $f^{-1}(G) \in \mathcal{A}$.
3. Si \mathcal{A} es el conjunto de partes de Ω , como en el caso que Ω sea finito o numerable, la condición (2.1) se cumple trivialmente.

Teorema. Sea X una variable aleatoria sobre un espacio de probabilidad (Ω, \mathcal{A}, P) . Entonces vale $X^{-1}(B) \in \mathcal{A}$ para todo $B \in \mathcal{B}$. (\mathcal{B} es el conjunto de boreleanos 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. Luego se tendrá que $\beta \subset \Phi$ (puesto que la σ -álgebra de Borel es la más chica que contiene a las semirectas).

Veamos que

(a) $\mathbb{R} \in \mathcal{A}$ 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 un σ -álgebra

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

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

Observese que P está definido sobre $X^{-1}(B)$ ya que este conjunto 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* .

Proposición. 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

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

Consideraremos las funciones medibles $g : \mathbb{R} \rightarrow \mathbb{R}$ con el σ -álgebra de Borel

Definición. 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 a \mathbb{R} como un espacio muestral $\Omega = \mathbb{R}$ y $\mathcal{A} = \mathcal{B}$ se ve que “ g es medible Borel” es equivalente a que “ g es una variable aleatoria”

Ejercicio 1. Demostrar los siguientes resultados

1. Si $g : \mathbb{R} \rightarrow \mathbb{R}$ es continua entonces g es medible.
2. Si $g : \mathbb{R} \rightarrow \mathbb{R}$ es monótona entonces g es medible.
3. Si B es boreliano, su función característica I_B es medible
4. Si $\{f_n\}_{n \geq 1}$ es una sucesión de funciones medibles entonces

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

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

son medibles. También son medibles

$$f(x) = \overline{\lim} f_n(x),$$

$$f(x) = \underline{\lim} f_n(x).$$

En particular si existe el límite puntual

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

es medible.

Observación.

El siguiente teorema tiene importancia teórico y práctica. La composición de una variable aleatoria con una función medible es una variable aleatoria.

Teorema. 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 $B \in \mathcal{B}$

$$[g(X)]^{-1}(B) = X^{-1}\left(\underbrace{g^{-1}(B)}_{\in \mathcal{B}}\right)$$

Como consecuencia de este teorema si g es continua y X es una variable aleatoria resulta $g(X)$ también una variable aleatoria. Por ejemplo si X es una variable aleatoria, entonces $\sin(X)$, $\cos(X)$, $\exp(X)$ son variables aleatorias/

Proposición.

Si X, Y son variables aleatorias entonces

1. $X + Y$; XY son variables aleatorias

2 Si $P(Y \neq 0) = 1$ entonces X/Y es una v.a

La demostraciones de 1 y 2 se verán más adelante/

2.3. Función de distribución de una variable aleatoria.

Definición. Sea X una v.a. 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 $B \in \mathcal{B}$

Sea X una variable aleatoria sobre (Ω, \mathcal{A}, P) y sea F_X su función de distribución. Entonces se cumplen las siguientes propiedades

P1. F_X es monótona no decreciente, es decir $x_1 < x_2$ implica $F_X(x_1) \leq F_X(x_2)$.

P2. $\lim_{x \rightarrow \infty} F_X(x) = 1$.

P3. $\lim_{x \rightarrow -\infty} F_X(x) = 0$.

P4. F_X es continua a derecha en todo punto de \mathbb{R} .

Demostración.

P1. Si $x < x'$ entonces

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

y por lo tanto

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

P2. 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 a 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 ser una probabilidad evidentemente se cumple que para cualquier $\varepsilon > 0$, $F_X(x) < \varepsilon + 1$. Por lo tanto solo 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

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

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

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

P3. Se trata esencialmente de la misma demostración.

En primer lugar se ve 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 P1□

P4. 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: 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

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

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 $\delta < \frac{1}{n_0}$ entonces si $0 < x - x_0 < \delta$

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

Lema. Sea $(x_n)_{n \in \mathbb{N}}$ una sucesión de números reales estrictamente creciente que converge a x_0 .

Entonces

$$\lim_{n \rightarrow \infty} F_X(x_n) = F_X(x_0) - P_X(\{x_0\}).$$

Demostración. La sucesión de intervalos $A_n = (-\infty; x_n]$ es creciente y

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

Entonces

$$\begin{aligned} \lim_{n \rightarrow \infty} F_X(x_n) &= \lim_{n \rightarrow \infty} P_X(A_n) = P_X((-\infty, x_0)) \\ &= P_X((-\infty, x_0]) - P_X(\{x_0\}) \\ &= F_X(x_0) - P_X(\{x_0\}). \quad \square \end{aligned}$$

Teorema. 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.

Se aplica el lema anterior utilizando la misma técnica usada para probar P2 o P3.

Corolario.

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

Demostración

Inmediato a partir del teorema.

Teorema. Sea F_X la función de distribución de una v.a X . Entonces el conjunto de discontinuidades

$$\{x \in \mathbb{R} : F_X \text{ es discontinua en } x\}.$$

es a lo sumo numerable.

Demostración.

Consideremos

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

y para todo $k \in \mathbb{N}$ sea

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

Entonces es fácil ver 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 propiedades P1, P2, P3 y P4 define una función de distribución para cierta ariable aleatoria X (no única).

Teorema de Extensión. Sea $F : \mathbb{R} \rightarrow [0, 1]$ una función con las propiedades P1-P4. Luego existe una única probabilidad P sobre $(\mathbb{R}, \mathcal{B})$ tal que para todo $x \in \mathbb{R}$ se tiene

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

Demostración

No se demostrará en este curso. Hay que utilizar teoría de la medida.

Veremos ahora algunas consecuencias del Teorema de Extensión.

Corolario 1.

Si X y X^* son variables aleatorias con distribuciones F_X, F_{X^*} tal que $F_X = F_{X^*}$ entonces para todo $B \in \beta$ se tendrá

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

Demostración.

Es consecuencia de la unicidad del teorema de extensión.

Corolario 2.

Si F satisface P1, P2, P3 y P4 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}, β, 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

$$F_X(x) = P_X((-\infty, x]) = P(\{y : X(y) \leq x\}) = P((-\infty, x]) = F(x). \square$$

Capítulo 3

Variables aleatorias discretas y continuas.

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

3.1. Variables aleatorias discretas.

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

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

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

$$R_X \subset A.$$

Demostración

(i) Sea A un conjunto a lo sumo numerable tal que $P_X(A) = 1$

$$A = R_X \uplus (A - R_X).$$

Entonces

$$\begin{aligned} 1 = P_X(A) &= P_X\left(R_X \uplus (A - R_X)\right) = \\ &P_X(R_X) + P_X(A - R_X). \end{aligned}$$

Luego basta probar que

$$P_X(A - R_X) = 0. \quad (3.1)$$

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

$$B = \bigsqcup_{x \in B} \{x\},$$

resulta que $P(B) = \sum_{x \in B} P_X(\{x\}) = 0$. Luego hemos demostrado (3.1).

(ii) Supongamos que exista $x_0 \in R_X$ tal que $x_0 \notin A$ entonces consideramos $\tilde{A} = A \bigsqcup \{x_0\}$ 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 y que cada x en el que la probabilidad es mayor que cero es un átomo.

Corolario. Para todo $B \in \mathcal{B}$ se tiene

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

Demostración.

Escribiendo a

$$B = (R_X \cap B) \bigsqcup (B - R_X),$$

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

$$P(B - R_X) = 0,$$

y se obtiene el resultado.

Definición. Sea X una v.a.d. Se define la *función densidad de probabilidad asociada a la variable X* como la función

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

tal que

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

Observación.

La función de densidad satisface

$$p_X(x) > 0 \text{ si } x \in R_X$$

y determina totalmente la probabilidad P_X .

Para ver esto probaremos

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

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

Demostración

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

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.

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

El rango de esta variable es $R_X = \{0, 1, \dots, n\}$.

Obtendremos seguidamente la función de densidad de probabilidad de esta variable. 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

El espacio muestral Ω , es finito y por lo tanto podemos tomar la σ -álgebra como el conjunto de partes de Ω .

Con esta notación la v.a. se escribe

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

El evento que $\{X = x\}$ está dado por

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

En primer lugar determinamos la cantidad de elementos del conjunto A_x . Claramente se deben seleccionar x lugares entre los n posibles, para distribuir los unos, de manera que

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

Observación.

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

Para un experimento cualquiera se tiene que 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 sucesos y que la ocurrencia de $(\omega_1, \omega_2, \dots, \omega_n)$ involucra una intersección de eventos se tiene que

$$\begin{aligned} P((\omega_1, \omega_2, \dots, \omega_n)) &= \prod_{i=1}^n P(\omega_i) = \\ &= \prod_{i=1}^n \theta^{\omega_i} (1 - \theta)^{1-\omega_i} = \\ &= \theta^{\sum_{i=1}^n \omega_i} (1 - \theta)^{n - \sum_{i=1}^n \omega_i}. \end{aligned}$$

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 cualquier elemento 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

$$A_x = \bigsqcup_{\tilde{\omega} \in A_x} \{\tilde{\omega}\}$$

entonces

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

3.2.2. Binomial Negativa (o Distribución de Pascal)

Se tiene el experimento cuyo resultado es éxito con probabilidad θ como en el caso aso de la distribución binomial. Ahora se repite el experimento en forma independiente hasta que ocurran k éxitos. En este caso los parámetros 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 hubo que realizar para obtener los k éxitos. La distribución de esta variable se denomina binomial negatica o de Pascal y se la denotará con $BN(\theta, k)$. Ahora el rango es infinito numerable

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

Para determinar la distribución introducimos las variables aleatorias Z_i para todo $i \in \mathbb{N}$ definidas por

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

Estas variables se toman independientes. Luego definimos las variables a

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

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

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

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

Como los dos eventos del lado derecho de la última ecuación son independientes y usando el hecho que Y_{x-1} es $Bi(\theta, x - 1)$ resulta

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

3.2.3. Distribución geométrica.

Se llama distribución geométrica a la $BN(\theta, k)$, 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 $G(\theta)$.

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

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

Reemplazando $k = 1$ en la distribución BN se obtiene

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

Se verifica 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.

Podemos pensar en el esquema de una urna que contiene N bolillas de las cuales D son negras y $N - D$ blancas. Se extraen secuencialmente n bolillas y se define la variable X como el número de bolillas negras extraídas. Si la 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 una 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 $H(N, D, n)$.

Estudiemos su rango de esta distribución. Por un lado claramente se observa que no se puede sacar un número negativo de negras, ni tampoco más veces que la cantidad total de bolillas extraídas, por lo tanto:

$$0 \leq X \leq n.$$

Por otro lado, claramente a lo sumo se pueden extraer todas las negras

$$X \leq D.$$

También debemos observar que el número de blancas debe ser menor que su número total

$$n - X \leq N - D.$$

En definitiva

$$R_X = \{x \in \mathbb{N} : \max(0; n - N + D) \leq x \leq \min(n, D)\}.$$

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

$$\mathbb{I}_N = \{x \in \mathbb{N} : 1 \leq x \leq N\},$$

el conjunto de todos eventos posibles es

$$\Omega = \{A \subset \mathbb{I}_N : \#A = n\}.$$

Es decir que los posibles resultados del experimentos son todos los subconjuntos de \mathbb{I}_N con cardinal n . Como todos estos subconjuntos tendrán la misma probabilidad de ser extraídos, estaremos en un caso de resultados equiprobables. El cardinal de Ω es

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

Consideremos el evento $\{X = x\}$, es decir, los casos en los que de las n extracciones x bolillas sean negras.

Para obtener el cardinal de $\{X = x\}$ podemos pensar de la siguiente manera. En primer instancia, escogemos todos los subconjuntos de x bolas negras entre las D posibles, esto nos da

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

Para cada subconjunto de x 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.

En el contexto de las urnas de la situación anterior sea $n \in \mathbb{N}$ fijo y consideremos una sucesión de distribuciones hipergeométricas $H(N, D_N, n)$ tales que

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

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

$$\lim_{N \rightarrow \infty} p_H(x) = p_B(x).$$

Es decir para N suficientemente grande la distribución $H(N, D_N, n)$ se puede aproximar por la distribución $\text{Bi}(\theta, n)$. Heurísticamente, este resultado se debe a que cuando 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

Un proceso de Poisson se presenta cuando se trata de registrar el número de veces que ocurre cierto evento en un lapso determinado de tiempo. Por ejemplo

a) Se desea registrar el número de clientes que entran en un determinado banco a un determinado lapso de tiempo el mes de octubre del corriente año.

b) El número de accidentes automovilísticos que ocurren en la ciudad de Bs. As. en cada mes.

c) El número total de llamadas telefónicas que llegan a una central telefónica entre las 15hs y 16hs de los días hábiles.

Estos procesos se basan en un conjunto de supuestos que trataremos con mayor detalle, más adelante.

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

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

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 una función de distribución asociada a una variable aleatoria discreta

Supongamos que el rango de X sea finito $R_X = \{x_0, x_1, \dots, x_n\}$. 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

$$c_i = \sum_{j=0}^i p_X(x_j); \quad 0 \leq i \leq n.$$

Luego se tendrá

$$F_X(x) = \begin{cases} 0 & \text{si } x \in (-\infty, x_0) \\ c_i & \text{si } x \in [x_i, x_{i+1}) \text{ para } 1 \leq i \leq n-1 \\ 1 & \text{si } x \in [x_n, \infty) \end{cases}$$

Ejercicio graficar la situación.

3.3. Variables aleatorias absolutamente continuas.

Definición. Se dice que X es una *variable aleatoria 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. 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 (integrable Lebesgue) 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* .

Tendremos las siguientes propiedades.

P1. Si f_X es una función de densidad de probabilidad para una variable aleatoria X entonces

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

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

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

entonces definiendo

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

se obtiene una función que resulta ser de distribución para alguna variable aleatoria X . Esto se justifica puesto que una función integral de una función no negativa satisface las propiedades P1-P4 que caracterizan a una función de distribución (ver teorema de extensión).

P2. Supongamos que F_X es absolutamente continua. Entonces

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

P3. Si F_X es absolutamente continua entonces es continua.

Lo demostraremos para el caso que f_X es una función acotada, es decir existe $M > 0$ tal que para todo $x \in \mathbb{R}$ se tiene $f_X(x) \leq M$.

Sea $\varepsilon > 0$ y $x \in \mathbb{R}$ entonces

$$P_X(\{x\}) \leq P((x - \varepsilon, x + \varepsilon]) = \int_{x-\varepsilon}^{x+\varepsilon} f_X(t) dt \leq M2\varepsilon/$$

Ahora haciendo que $\varepsilon \rightarrow 0$ se obtiene $P_X(\{x\}) = 0$ y con ello que F_X es continua.

El mismo resultado se puede probar para una función de densidad f_X no acotada, integrable Riemann sobre \mathbb{R} .

El nombre densidad nos recuerda "la cantidad de masa por unidad de longitud, área o volumen" según el caso.

Por similitud 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. Sea f_X una función de densidad continua en x_0 entonces

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

Demostración

Sea

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

y

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

Por continuidad

$$f(x_0) = \lim_{h \rightarrow 0} M_h = \lim_{h \rightarrow 0} m_h. \quad (3.2)$$

Por otro lado valen las desigualdades

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

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

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

Luego, teniendo en cuenta (3.2) y pasando al límite cuando $h \rightarrow 0$ se obtiene

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

de donde se deduce el Teorema. \square

Teorema. 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

$$F'_X(x_0) = f(x_0)$$

Demostración.

Análoga a la anterior.

Comentarios vinculados a teoría de la medida.

1. Una definición alternativa de la absoluta continuidad, es esta:

Se dice que $F : \mathbb{R} \rightarrow \mathbb{R}$ es absolutamente continua si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que para toda sucesión

$$x_0 < y_0 < x_1 < y_1 < \dots < x_n < y_n$$

que cumple

$$\sum_{i=0}^n (y_i - x_i) < \delta,$$

se tiene que

$$\sum_{i=0}^n |F(y_i) - F(x_i)| < \varepsilon.$$

En este sentido toda función absolutamente continua es uniformemente continua y en consecuencia continua. Además resulta ser de variación acotada sobre todo intervalo acotado. Si f es integrable Lebesgue (Riemann) entonces

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

es absolutamente continua.

Es decir (con esta definición) resulta que si F es de variación acotada y continua a derecha entonces

F es absolutamente continua sii existe f integrable tal que

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

2. También se dice que la función integral es absolutamente continua como función de conjunto.

Dada f integrable definamos

$$T(A) = \int_A f(x) dx. \quad (3.3)$$

Se comprueba que la integral de Lebesgue (Riemann) satisface la siguiente propiedad.

Si f es una función integrable Lebesgue y sea μ la medida de Lebesgue sobre R entonces para todo $\varepsilon > 0$ existe $\delta > 0$ tal que si B es un boreliano que satisface $\mu(B) < \delta$ entonces

$$|T(B)| = \left| \int_B f(x) dx \right| < \varepsilon.$$

$T(B)$ define una probabilidad si f es una función de densidad, es decir si $f(x) \geq 0$ y $F(\mathbb{R}) = 1$.

3.4. Ejemplos de distribuciones continuas

3.4.1. Distribución uniforme en un intervalo $[a; b] : U(a, b)$.

Consideremos

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

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

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

La función distribución asociada es $F_X : \mathbb{R} \rightarrow [0, 1]$ es

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

Claramente no existe ninguna distribución que sea uniforme sobre toda la recta, puesto que al ser constante la densidad no resultaría integrable en \mathbb{R} .

En particular consideremos la distribución uniforme $U(0, 1)$

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

La función distribución asociada

$$F_X : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$$

es en este caso

$$F_X(x) = \begin{cases} 0 & \text{si } x \in (-\infty, 0] \\ x & \text{si } x \in (0, 1] \\ 1 & \text{si } x \in (1, \infty). \end{cases} \quad (3.4)$$

Observación.

1. Es claro que (3.4) vale puesto que si $x \in (0, 1)$

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

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 distribución nos da la medida del subconjunto, es decir su longitud.

3. De muchas maneras diferentes pueden generarse distribuciones uniformes. Por ejemplo recursivamente podemos considerar dos números A_1, A_2 de ocho dígitos, y definir A_3 por los últimos ocho dígitos de A_2A_1 . Luego el primer número con distribución $U(0, 1)$ sería

$$\tilde{A}_1 = A_3 10^{-8}.$$

Luego definimos A_4 por los últimos ocho dígitos de A_3A_2 y definimos el segundo número con distribución uniforme $U(0, 2)$ por

$$\tilde{A}_2 = A_4 10^{-8}.$$

En general si definimos A_k por las últimas ocho cifras de $A_{k-1}A_{k-2}$ el número $k - 2$ con distribución $U(0, 1)$ es

$$\tilde{A}_{k-2} = A_k 10^{-8}.$$

Se puede probar que los la distribución de los números se comporta como una variable $U(0,1)$

Generación de distribuciones a partir de la distribución uniforme $U(0, 1)$

Vamos a ver como a partir de una variable aleatoria con distribución $U(0, 1)$ 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 la 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 propiedades P1, P2, P3, y P4 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 preimagen via F del punto $y : F^{-1}(y) \neq \emptyset$. Si F es continua por Bolzano (generalizado) 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 que con las propiedades de la hipótesis constante en un intervalo. Para y igual a ese valor hay infinitas preimágenes.

Sugerencia: Hacer para cada una de las situaciones un gráfico adecuado.

Teorema. 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 P2 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 está acotado inferiormente. Por P3 se tiene que ,

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

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

$$F(-n_0) < y. \tag{3.5}$$

Ahora bien si $x \in A_y$ no puede ser que $-n_0 > x$ puesto que por monotonía (P1) se cumpliría

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

en contradicción con (3.5). En definitiva se tiene que si $x \in A_y$, entonces $n_0 \leq x$, y por lo tanto A_y está acotado inferiormente \square .

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

Definición. Dada

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

que satisface las propiedades de una función de distribución P1,P2,P3 y P4 se define $F^{-1} : (0, 1) \rightarrow \mathbb{R}$ por

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

Propiedades de la función F^{-1} .

P1. El ínfimo del conjunto A_y resulta ser el mínimo

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

Demostración.

Basta con probar que $F^{-1}(y)$ pertenece al conjunto A_y , lo cual significa que

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

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 P4 (continuidad a derecha)

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

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

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

de manera que por el teorema de la conservación del signo

$$F(F^{-1}(y)) \geq y, \tag{3.6}$$

y por lo tanto $F^{-1}(y) \in A_y$ \square .

P2. Se cumple que

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

Demostración. Se demostro durante la demostración de P1 en (3.6)
P3. 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 caracter de ínfimo del elemento $F^{-1}(y)$.
Tomemos un punto intermedio entre $F(F^{-1}(y))$ e y digamos y^*

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

Por ser F continua, el teorema de Bolzano (o el teorema de los valores intermedios) 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 (P1)

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

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

P4. En general vale que

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

Demostración.

Es claro que para todo $x : x \in A_{F(x)}$ puesto que naturalmente $F(x) \leq F(x)$.

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 $a = x \in A_{F(x)}$ resulta la propiedad P4. \square

Proposición (*Caracterización de A_y como semirecta*). Sea $y \in (0, 1)$ fijo. Los conjuntos

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

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

coinciden.

Demostración.

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

Sea $x \in B_y$, entonces como $x \geq F^{-1}(y)$, por la monotonía de F resulta $F(x) \geq F(F^{-1}(y)) \geq y$ y por lo tanto. Luego

$$B_y \subset A_y. \quad (3.7)$$

Sea ahora $x \notin B_y$, luego $x < F^{-1}(y)$. Como $F^{-1}(y)$ es el mínimo de A_y resulta $x \notin A_y$. Por lo tanto $B_y^C \subset A_y^C$ y luego

$$A_y \subset B_y. \quad (3.8)$$

(3.7) y (3.8) implican $A_y = B_y$.

Ejercicio 2. Probar que

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

Veremos ahora como a partir de una variable con distribución $U(0, 1)$ se puede generar otra variable que tenga función de distribución F , donde F es cualquier función que cumpla las propiedades P1-P4 que caracterizan a una función de distribución.

Teorema. Sea U una variable aleatoria con distribución $U(0, 1)$. Luego si F tiene las propiedades P1-P4 que caracterizan una función de distribución, se tiene que $X = F^{-1}(U)$ tiene función de distribución F

Demostración. Usando la proposición anterior y el hecho de que $F_U(u) = u$ se tiene

$$\begin{aligned} F_X(x) &= P_X((-\infty, x]) = P(\{F^{-1}(U) \leq x\}) = P(\{U \leq F(x)\}) = \\ &= F_U(F(x)) = F(x). \end{aligned}$$

Ejercicio 3.

Sea X una variable con rango $R_X = \mathbb{N}_{\geq 0}$ y sea

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

Verificar que F_X^{-1} es de la forma

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

Comprobar que el resultado anterior vale en este caso.

El siguiente lema de demostración inmediata es muy importante

Lema. Sean X y X^* dos variables aleatorias tales que $F_X = F_{X^*}$ (y por lo tanto $P_X = P_{X^*}$). Consideremos una función g medible y consideremos las variables aleatorias obtenidas componiendo

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

Entonces

$$F_Z = F_{Z^*},$$

y por lo tanto $P_Z = P_{Z^*}$.

Demostración.

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

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

Esto se deduce de

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

Ahora veamos en cierto sentido la reciproca de la observación 2.

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

Demostración.

Consideremos una variable aleatoria U con distribución $U(0, 1)$ y sea $X^* = F_X^{-1}(U)$. Sabemos que X^* tiene distribución F_X .

Luego por el Lema anterior 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 lo demostrado anteriormente se tiene $F_X(F_X^{-1}(U)) = U$. Luego Y^* tiene distribución $U(0, 1)$ y por lo tanto también Y \square .

3.4.2. 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 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, y que por tanto tiene distribución aproximadamente normal. La distribución depende de dos parámetros $\mu \in \mathbb{R}$ y $\sigma^2 \in \mathbb{R}_{>0}$.

La distribución normal esta Standarizada scuando los parámetros son $\mu = 0$ y $\sigma^2 = 1$. En este caso la función de densidad es

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

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 traves de un cambio de variable a corrdenadas polares. Optamos por la segunda forma

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

Ahora hacemos el cambio de variable

$$\begin{aligned} x(\rho, \phi) &= x = \rho \cos(\phi) \\ y(\rho, \phi) &= y = \rho \sin(\phi) \\ x^2 + y^2 &= \rho^2 \end{aligned}$$

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 la integral resulta ser

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

Haciendo el cambio de variable

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

se obtiene

$$I^2 = 2\pi \int_0^{+\infty} \exp(-u) du = 2\pi (-\exp(-u)|_0^{+\infty}) = 2\pi,$$

y por lo tanto

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

Luego

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

3.4.3. Distribución Exponencial

Esta distribución depende de un parámetro λ que puede tomar cualquier número 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

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

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

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0. \end{cases} \quad (3.9)$$

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

Esta distribución aparece generalmente cuando se trata de estudiar la “durabilidad de mecanismo bajo el supuesto de que el sistema no se desgasta significativamente a lo largo del tiempo”. Como ejemplos suelen citarse “la duración de una lámpara eléctrica”. 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 los mecanismos electrónicos, ya que no tienen prácticamente desgaste.

Para precisar el concepto de desgaste decimos que la distribución de X *no tiene desgaste* cuando

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

donde $0 < a, b$.

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 halla durado hasta el tiempo b . Es decir el proceso “no tiene memoria del tiempo que estuvo funcionando” (no recuerda que tan viejo es) y por tanto, mientras funciona ”lo hace como 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 del falta de desgaste caracteriza a la distribución exponencial. Esto significa que la 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

$$P(X \geq a + b) = P(X \geq b) P(X \geq a). \quad (3.10)$$

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 3.10, esta se puede escribir también como

$$G_X(a + b) = G_X(a) G_X(b) \quad (3.11)$$

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

En el caso de que X tenga distribución exponencial por (3.9) 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 las distribuciones exponenciales

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

Demostración. Supongamos primero que $G_X(x) = e^{-\lambda x}$. Probaremos que (3.11) 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 (3.11) 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 hip[otesis].

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 probaremos que para todo $a \geq 0$ vale que

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

La estrategia es primero probarlo para un entero no negativo, luego para los racionales y por último para un número real no negativo.

Sea $n \in \mathbb{N}$ entonces

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

Ahora sea $\frac{r}{n} \in \mathbb{Q}$ entonces

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

Entonces

$$G_X\left(\frac{r}{n}\right) = [G_X\left(\frac{r}{n}\right)]^{\frac{1}{n}} = [(G_X(1))^r]^{\frac{1}{n}} = [G_X(1)]^{\frac{r}{n}}.$$

Por ultimo 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

$$G_X(a) = \lim_{n \rightarrow \infty} G_X(r_n) = \lim_{n \rightarrow \infty} (G_X(1))^{r_n} = (G_X(1))^{\lim_{n \rightarrow \infty} r_n} = [G_X(1)]^a.$$

Ahora pongamos

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

de manera que

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

En definitiva podemos escribir

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

y el teorema queda probado.

Capítulo 4

Vectores aleatorios.

Generalmente interesan más de una característica de una población. Por ejemplo podemos estar interesados en estudiar “el perfil biológico” de los alumnos de 10 años de una determinada escuela. Podríamos considerar que el perfil se compone de la talla, el peso, presión sanguínea, frecuencia cardíaca, capacidad respiratoria. Están en juego cinco variables aleatorias, que deberían tratarse simultáneamente. Esto motiva la siguiente definición un vector aleatorio.

Definición. 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 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 para todo i se tiene que $X_i^{-1}((-\infty, x_i]) \in \mathcal{A}$ y \mathcal{A} es un σ -álgebra se concluye que $\mathbf{X}^{-1}(B) \in \mathcal{A}$.

Observaciones.

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

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

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

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

ya que se pueden escribir de la siguiente forma

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

y que diferencias de conjuntos de un σ -álgebra son conjuntos del σ -álgebra.

Es interesante también 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 4. Probar la siguiente

Teorema . Sea \mathbf{X} es un vector aleatorio de dimensión k . Entonces si $B \in \mathcal{B}^k$ se tiene que $\mathbf{X}^{-1}(B) \in \mathcal{A}$.

4.1. Espacio de probabilidad inducido.

Definición. 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 5. Probar la siguiente.

Proposición. $P_{\mathbf{X}}$ 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} .

4.2. Función de distribución conjunta de un vector aleatorio.

Dado un vector aleatorio $\mathbf{X} = (X_1, \dots, X_k)$, se define la *función distribución conjunta del vector \mathbf{X}* como función $F_{\mathbf{X}} : \mathbb{R}^k \rightarrow [0; 1]$ definida 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 $F_{\mathbf{X}}$.

P1. $F_{\mathbf{X}}$ 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

$$F_{\mathbf{X}}((x_1, \dots, x_i, \dots, x_n)) \leq F_{\mathbf{X}}((x_1, \dots, x'_i, \dots, x_n)).$$

P2. 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

$$\begin{aligned} \lim_{i \rightarrow +\infty} F_{\mathbf{X}}(x_{1i}, x_{2i}, \dots, x_{ki}) &= \lim_{i \rightarrow \infty} P_{\mathbf{X}}((-\infty, x_{1i}] \times (-\infty, x_{2i}] \times \dots \times (-\infty, x_{ki})) = \\ &= P_{\mathbf{X}}\left(\bigcup_{i \in \mathbb{N}} C_i\right) = P_{\mathbf{X}}(\mathbb{R}^k) = 1. \quad \square \end{aligned}$$

P3. 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. Para este caso consideremos una sucesión monótona no creciente tal que $\{x_{ij}\}_{j \in \mathbb{N}} \downarrow -\infty$ (i fijo).

Entonces si definimos $\{C_i\}_{i \in \mathbb{N}}$ como en (4.5) tendremos para todo $i \in \mathbb{N}$ se tiene $C_{i+1} \subset C_i$ y además

$$\bigcap_{i \in \mathbb{N}} C_i = \emptyset.$$

Por lo tanto

$$\begin{aligned} \lim_{j \rightarrow -\infty} F_{\mathbf{X}}(x_{1i}, \dots, x_{ij}, \dots, x_{ki}) &= \lim_{i \rightarrow \infty} P_{\mathbf{X}}((-\infty, x_{1i}] \times \dots \times (-\infty, x_{ij}] \times \dots \times (-\infty, x_{ki})) = \\ &= P_{\mathbf{X}}\left(\bigcap_{j \in \mathbb{N}} C_j\right) = P_{\mathbf{X}}(\emptyset) = 0. \quad \square \end{aligned}$$

P4. $F_{\mathbf{X}}$ es continua a derecha.

Demostración. Sea $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ y consideremos sucesiones monótonas no crecientes 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

$$C_i = (-\infty, x_{1i}] \times (-\infty, x_{2i}] \times \dots \times (-\infty, x_{ki}].$$

Entonces

$$C_{i+1} \subset C_i$$

y

$$\bigcap_{i \in \mathbb{N}} C_i = A_{(x_1, \dots, x_k)}.$$

Luego

$$\lim_{i \rightarrow +\infty} F_{\mathbf{X}}(x_{1i}, x_{2i}, \dots, x_{ki}) = F_{\mathbf{X}}((x_1, x_2, \dots, x_k)). \quad \square$$

Observación.

Estas cuatro propiedades no caracterizan a una función de distribución conjunta como ocurría para el caso de una variable aleatoria.

Las propiedades P1, P2, P3, y P4 no alcanzan para caracterizar una función de distribución conjunta. Para fijar ideas pensemos en \mathbb{R}^2 .

Sea entonces un vector aleatorio en \mathbb{R}^2 $X = (X_1, X_2)$ y $F_{\mathbf{X}}$ su función de distribución conjunta. Sea $A_{x_1 x_2} = (-\infty, x_1] \times (-\infty, x_2]$ y $R = (a_1, b_1] \times (a_2, b_2]$.

El rectángulo R puede ser escrito de la siguiente manera

$$R = (A_{b_1 b_2} - A_{a_1 b_2}) - (A_{b_1 a_2} - A_{a_1 a_2}).$$

Teniendo en cuenta las inclusiones

$$A_{a_1 a_2} \subset A_{b_1 a_2},$$

$$A_{a_1 b_2} \subset A_{b_1 b_2}$$

y

$$(A_{b_1 a_2} - A_{a_1 a_2}) \subset (A_{b_1 b_2} - A_{a_1 b_2}),$$

resulta que

$$\begin{aligned} P_{\mathbf{X}}(R) &= P_{\mathbf{X}}(A_{b_1 b_2} - A_{a_1 b_2}) - P_{\mathbf{X}}(A_{b_1 a_2} - A_{a_1 a_2}) \\ &= P_{\mathbf{X}}(A_{b_1 b_2}) - P_{\mathbf{X}}(A_{a_1 b_2}) - P_{\mathbf{X}}(A_{b_1 a_2}) + P_{\mathbf{X}}(A_{a_1 a_2}). \end{aligned}$$

Como $P_{\mathbf{X}}(A_{x_1 x_2}) = F_{\mathbf{X}}(x_1, x_2)$, resulta

$$\begin{aligned} 0 &\leq P_{\mathbf{X}}(R) \\ &= F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(b_1, a_2) + F_{\mathbf{X}}(a_1, a_2). \end{aligned}$$

Observaciones.

1. Se sugiere hacer un dibujo.
2. Esto muestra que la probabilidad de el rectángulo R se determina por el valor de $F_{\mathbf{X}}$ sobre los vértices: 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 $F_{\mathbf{X}}$ para todo $a_1 < b_1$ y $a_2 < b_2$ se debería cumplir

$$F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(a_1, b_2) - F_{\mathbf{X}}(b_1, a_2) + F_{\mathbf{X}}(a_1, a_2) \geq 0. \quad (4.6)$$

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 (4.6).

Sea $F : \mathbb{R}^k \rightarrow [\mathbb{R}, \mathbb{R}]$ definida por

$$F(x_1, x_2) = \begin{cases} 1 & \text{si } x_1 + x_2 \geq 1, x_1 \geq 0, x_2 \geq 0 \\ 0 & \text{si no.} \end{cases}$$

Es fácil verificar que esta función es monótona no decreciente en cada variable,

$$\lim_{x_1 \rightarrow \infty, x_2 \rightarrow \infty} F(x_1, x_2) = 1,$$

$$\lim_{x_i \rightarrow -\infty} F(x_1, x_2) = 0 \text{ para cualquier } i = 1, 2,$$

y es continua a derecha. Pero si consideramos el rectángulo $R = (0, 1] \times (0, 1]$ entonces

$$P(R) = F(1, 1) + F(0, 0) - (F(0, 1) + F(1, 0)) = 1 - 2 = -1.$$

Esto muestra que F no puede ser la función de distribución de ningún vector aleatorio en \mathbb{R}^k .

Esto motiva la definición del operador diferencia.

Definición. Sea F una función de k variables. Si $a_i < b_i$ se define el *operador diferencia en la variable i* por

$$\Delta (x_i)_{a_i}^{b_i} F = F(x_1, x_2, \dots, x_{i-1}, b_i, x_{i+1}, \dots, x_k) - F(x_1, x_2, \dots, x_{i-1}, a_i, x_{i+1}, \dots, x_k).$$

Estos operadores se pueden aplicar en forma sucesiva. Por ejemplo

$$\begin{aligned} & \Delta (x_j)_{a_j}^{b_j} \Delta (x_i)_{a_i}^{b_i} F \\ &= \Delta (x_j)_{a_j}^{b_j} (F(x_1, \dots, x_{i-1}, b_i, x_{i+1}, \dots, x_k) \\ & \quad - F(x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots, x_k)) \\ &= \Delta (x_j)_{a_j}^{b_j} F(x_1, x_2, \dots, x_{i-1}, b_i, x_{j+1}, \dots, x_k) \\ & \quad - \Delta (x_j)_{a_j}^{b_j} F(x_1, x_2, \dots, x_{i-1}, a_i, x_{j+1}, \dots, x_k) \\ &= (F(x_1, \dots, x_{i-1}, b_i, x_{i+1}, \dots, x_{j-1}, b_j, x_{j+1}, \dots, x_k) \\ & \quad - F(x_1, \dots, x_{i-1}, b_i, x_{i+1}, \dots, x_{j-1}, a_j, x_{j+1}, \dots, x_k)) \\ & \quad - (F(x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots, x_{j-1}, b_j, x_{j+1}, \dots, x_k) \\ & \quad - F(x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots, x_{j-1}, a_j, x_{j+1}, \dots, x_k)). \end{aligned}$$

Más generalmente, si $a_1 < b_1, a_2 < b_2, \dots, a_k < b_k$ podemos conformar la diferencia sucesiva

$$\Delta (x_1)_{a_1}^{b_1} \dots \Delta (x_{k-1})_{a_{k-1}}^{b_{k-1}} \Delta (x_k)_{a_k}^{b_k} F.$$

Observación.

Podemos expresar la propiedad (4.6) en términos del operador diferencia como

$$\begin{aligned} P_{\mathbf{X}}(R) &= (F_{\mathbf{X}}(b_1, b_2) - F_{\mathbf{X}}(a_1, b_2)) - (F_{\mathbf{X}}(b_1, a_2) - F_{\mathbf{X}}(a_1, a_2)) \\ &= \Delta(x_1)_{a_1}^{b_1} F_{\mathbf{X}}(x_1, b_2) - \Delta(x_1)_{a_1}^{b_1} F_{\mathbf{X}}(x_1, a_2) \\ &= \Delta(x_2)_{a_2}^{b_2} \Delta(x_1)_{a_1}^{b_1} F_{\mathbf{X}}(x_1, x_2) \geq 0 \end{aligned}$$

En general se puede probar el siguiente Teorema

Teorema. Sea $F_{\mathbf{X}}$ e 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

$$\begin{aligned} P_{\mathbf{X}}((a_1, b_1] \times (a_2, b_2] \times \dots \times (a_k, b_k]) \\ = \Delta(x_1)_{a_1}^{b_1} \dots \Delta(x_{k-1})_{a_{k-1}}^{b_{k-1}} \Delta(x_k)_{a_k}^{b_k} F_{\mathbf{X}}(x_1, x_2, \dots, x_k) \geq 0. \end{aligned}$$

Demostración. Para probar el Teorema, consideremos para cada $h, 0 \leq h \leq k$ los conjuntos de la forma

$$C_h = (a_1, b_1] \times (a_2, b_2] \times \dots \times (a_h, b_h] \times (-\infty, x_{h+1}] \times \dots \times (-\infty, x_k].$$

Se e prueba por inducción que para todo h

$$P_{\mathbf{X}}(C_h) = \Delta(x_1)_{a_1}^{b_1} \dots \Delta(x_{h-1})_{a_{h-1}}^{b_{h-1}} \Delta(x_h)_{a_h}^{b_h} F(x_1, x_2, \dots, x_h, x_{h+1}, \dots, x_k).$$

Esta inducción se deja como ejercicio. Para $h = k$ se obtiene el Teorema.

Luego podemos enunciar una propiedad adicional que satisface una función de distribución conjunta

P5. 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

$$\Delta(x_1)_{a_1}^{b_1} \dots \Delta(x_{k-1})_{a_{k-1}}^{b_{k-1}} \Delta(x_k)_{a_k}^{b_k} F_{\mathbf{X}}(x_1, x_2, \dots, x_k) \geq 0.$$

4.2.1.

El siguiente Teorema generaliza el Teorema de extensión para variables aleatorias

Teorema. Sea $F : \mathbb{R}^k \rightarrow [0, 1]$ una función que satisface las propiedades P1, P2, P3, P4 y P5. 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.

Corolario. 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. Si F satisface P1, P2, P3, P4 y P5 entonces existe un vector aleatorio $\mathbf{X} = (X_1, \dots, X_k)$ tal que

$$F_{\mathbf{X}} = F.$$

Demostración.

Sea $(\mathbb{R}^k, \mathcal{B}^k, P)$ el espacio de probabilidad tal que P es la extensión de F . Luego para todo $(x_1, \dots, x_k) \in \mathbb{R}^k$

$$F(x_1, x_2, \dots, x_k) = P((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k])$$

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

$$X_i^{-1}(\{x_i\}) = R \times \dots \times R \times (-\infty, x_i] \times R \times \dots \times R,$$

y que

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2, \dots, x_k) &= P_{\mathbf{X}}((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]) \\ &= P\left(\bigcap_{i=1}^k \{(x_1, \dots, x_k) : X_i(x_1, \dots, x_k) \leq x_i\}\right) \\ &= P\left(\bigcap_{i=1}^k X_i^{-1}((-\infty, x_i])\right) \\ &= P((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]) \\ &= F(x_1, x_2, \dots, x_k). \end{aligned}$$

4.3. Algunas propiedades de vectores aleatorios

Ahora estamos en condiciones de demostrar algunas afirmaciones hechas con anterioridad.

Definición. Diremos que $g : \mathbb{R}^k \rightarrow \mathbb{R}$ es medible si para todo $x \in \mathbb{R}$ se tiene que $g^{-1}((-\infty, x]) \in \mathcal{B}^k$.

Observación. Una función medible puede interpretarse como una variable aleatoria en el espacio $(\mathbb{R}^k, \mathcal{B}^k)$.

En particular vale la siguiente

Proposición. 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 6. Probar el siguiente teorema

Teorema. 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. 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 $Z = \frac{X}{Y}$ es una variable aleatoria.

Demostración.

Se trata de escribir a Z como imagen de X e Y via 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.

La demostración de (ii) y (iii) se deja como ejercicio.

Definición. 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 si g_j es medible para cada $j = 1, 2, \dots, h$.

Proposición. 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.

4.4. Independencia de variables aleatorias.

4.4.1. 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.

Recordemos que un conjunto de eventos A_1, A_2, \dots, A_k son independientes si para toda elección $1 \leq i_1 < i_2 < \dots < i_h \leq k$

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_h}) = \prod_{j=1}^h P(A_{i_j}).$$

La independencia de eventos requiere algo más que la propiedad de que la intersección de todos los eventos sea el producto de sus probabilidades. Al respecto dimos un ejemplo.

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” la emisión monetaria” son independientes? ¿El peso de un individuo y su presión sanguínea son independientes? etc. Para definir el concepto de independencia de variables aleatorias utilizaremos la noción de independencia de eventos.

Definición. Sean X_1, X_2, \dots, X_k variables aleatorias, sobre un mismo espacio de probabilidad (Ω, \mathcal{A}, P) . Diremos que dichas variables son independientes si para cualquier conjunto $B_1, B_2, \dots, B_k \in \mathcal{B}$ (Boreleanos en \mathbb{R}), los eventos $X_j^{-1}(B_j)$, $j = 1, 2, \dots, k$ son independientes.

La siguientes dos proposiciones dan dos caracterizaciones de la propiedad de independencia de un conjunto variables aleatorias.

Proposición 1. Las variables aleatorias X_1, \dots, X_k son independientes si y sólo si para toda conjunto de borelianos B_1, B_2, \dots, B_k vale que

$$P\left(\bigcap_{j=1}^k X_j^{-1}(B_j)\right) = \prod_{j=1}^k P\left(X_j^{-1}(B_j)\right). \quad (4.7)$$

Demostración.

Primero mostraremos que (4.7) es una condición necesaria. Si X_1, \dots, X_k son independientes, (4.7) debe cumplirse por definición de independencia de eventos.

Ahora probaremos la suficiencia de (4.7)

Debemos probar que (4.7) implica para cualquier subconjunto de índices $i_1 < i_2 < \dots < i_h$, $h < k$ 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

$$P\left(\bigcap_{j=1}^h X_{i_j}^{-1}(B_{i_j})\right) = P\left(\bigcap_{i=1}^k X_i^{-1}(C_i)\right) = \blacktriangle$$

$$\prod_{i=1}^k P(X_i^{-1}(C_i)) = \prod_{j=1}^h P(X_{i_j}^{-1}(B_{i_j})) \quad \square$$

Hay que observar que en \blacktriangle aplicamos la hipótesis para los conjuntos $C_i, i = 1, 2, \dots, k$.

Ahora escribiremos la misma proposición de otra manera

Proposición 2. 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

$$P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_k) = \prod_{j=1}^k P_{X_j}(B_j),$$

donde $\mathbf{X} = (X_1, X_2, \dots, X_k)$.

Demostración.

Como $P_{X_j}(B_j) = P(X_j^{-1}(B_j))$ por la Proposición 1 bastará mostrar que

$$P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_k) = P\left(\bigcap_{j=1}^h X_{i_j}^{-1}(B_{i_j})\right).$$

Para eso observamos que

$$P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_k) = P_{\mathbf{X}}(\{\omega : \mathbf{X}(\omega) \in B_1 \times B_2 \times \dots \times B_k\})$$

$$= P_{\mathbf{X}}(\{\omega : (X_1(\omega), X_2(\omega), \dots, X_k(\omega)) \in B_1 \times B_2 \times \dots \times B_k\})$$

$$= P\left(\bigcap_{j=1}^k \{\omega : X_j(\omega) \in B_j\}\right) = P\left(\bigcap_{j=1}^h X_{i_j}^{-1}(B_{i_j})\right).$$

El siguiente teorema, cuya demostración es un poco más delicada da una condición necesaria y suficiente para la independencia de un conjunto de variables que es más simple de verificar

Teorema. 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 tiene

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_k) = F_{X_1}(x_1) F_{X_2}(x_2) \dots F_{X_k}(x_k), \quad (4.8)$$

donde $\mathbf{X} = (X_1, X_2, \dots, X_k)$.

Demostración.

Para ver que (4.8) es una condición necesaria para la independencia de X_1, \dots, X_k , basta aplicar la Proposición 2 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

$$B_1 \times B_2 \times \dots \times B_r \times (-\infty, x_{r+1}] \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k]$$

Probaremos por inducción sobre r que vale la siguiente propiedad que llamamos A_r :

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times (-\infty, x_{r+1}] \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k]) \\ = P_{X_1}(B_1) \dots P_{X_r}(B_r) P_{X_{r+1}}((-\infty, x_{r+1}]) \dots P_{X_k}((-\infty, x_k]). \end{aligned} \quad (4.9)$$

Para $r = 0$, la condición (4.9) vale, 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 (4.9) vale para r , también vale reemplazando $(-\infty, x_{r+1}]$ por \mathbb{R} , esto es

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times \mathbb{R} \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k]) \\ = P_{X_1}(B_1) \dots P_{X_r}(B_r) P_{X_{r+1}}(\mathbb{R}) \dots P_{X_k}((-\infty, x_k]) = \\ = P_{X_1}(B_1) \dots P_{X_r}(B_r) P_{X_{r+2}}((-\infty, x_{r+2}]) \dots P_{X_k}((-\infty, x_k]). \end{aligned} \quad (4.10)$$

Para mostrar esto podemos considerar una sucesión creciente de semirectas $C_n = (-\infty, n]$. Luego $\mathbb{R} = \bigcup_{n \in \mathbb{N}} C_n$ y la sucesión $\{B_1 \times B_2 \times \dots \times B_r \times C_n \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k)\}$ $n = 1, 2, \dots$ es monótona no decreciente en \mathbb{R}^1 y vale

$$\begin{aligned} \bigcup_{n \in \mathbb{N}} B_1 \times B_2 \times \dots \times B_r \times C_n \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k] \\ = B_1 \times B_2 \times \dots \times B_r \times \mathbb{R} \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k] \end{aligned}$$

Luego usando que vale A_r tenemos que

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times \mathbb{R} \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k]) \\ = \lim_{n \rightarrow \infty} P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times (-\infty, n] \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k]) \\ = \lim_{n \rightarrow \infty} P_{\mathbf{X}}(B_1) P_{\mathbf{X}}(B_2) \dots P_{\mathbf{X}}(B_r) P_{\mathbf{X}}((-\infty, n]) P_{\mathbf{X}}((-\infty, x_{r+2}]) \dots P_{\mathbf{X}}((-\infty, x_k]) \\ = P_{\mathbf{X}}(B_1) P_{\mathbf{X}}(B_2) \dots P_{\mathbf{X}}(B_r) P_{\mathbf{X}}(\mathbb{R}) P_{\mathbf{X}}((-\infty, x_{r+2}]) \dots P_{\mathbf{X}}((-\infty, x_k]). \end{aligned}$$

Ahora probaremos A_{r+1} . Es decir debemos probar que dados boreleanos B_1, \dots, B_r y reales x_{r+2}, \dots, x_k se tiene

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times B_{r+1} \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k]) \\ = P_{X_1}(B_1) \dots P_{X_r}(B_r) P_{X_{r+1}}(B_{r+1}) \dots P_{X_k}((-\infty, x_k]). \end{aligned} \quad (4.11)$$

Consideremos el conjunto

$$A = B_1 \times B_2 \times \dots \times B_r \times \mathbb{R} \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k],$$

y distinguimos dos casos: (a) $P_{\mathbf{X}}(A) = 0$, (b) $P_{\mathbf{X}}(A) > 0$

Consideremos primero el caso (a). Por (4.10)

$$\begin{aligned} 0 = P_{\mathbf{X}}(A) &= P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times \mathbb{R} \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k]) \\ &= P_{X_1}(B_1) \dots P_{X_r}(B_r) P_{X_{r+1}}(\mathbb{R}) \dots P_{X_k}((-\infty, x_k]) \end{aligned}$$

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 m\u00edembro derecho de (4.11) es 0

Supongamos que $P_{\mathbf{X}}(B_i) = 0$ podemos suponer que $i = 1$, para fijar ideas. Entonces teniendo en cuenta que

$$B_1 \times B_2 \times \dots \times B_r \times B_{r+1} \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k] \subset B_1 \times \mathbb{R} \times \dots \times R,$$

obtenemos que

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times B_{r+1} \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k]) \\ \leq P_{\mathbf{X}}(B_1 \times \mathbb{R} \times \dots \times R) = P_{X_1}(B_1) = 0, \end{aligned}$$

y luego el m\u00edembro izquierdo de (4.11) tambi\u00e9n es 0 y la igualdad se cumple..

Ahora si $P_{X_i}((-\infty, x_i]) = 0$, de nuevo podemos suponer que $i = 1$ y proceder de manera an\u00e1loga. Luego (4.11) vale para el caso (a).

Consideremos el caso (b), luego $P_{\mathbf{X}}(A) > 0$. Definimos un nuevo espacio de probabilidades $(\mathbb{R}, \mathcal{B}, \mathbb{P}^*)$ de la siguiente manera: Para todo $B \in \mathcal{B}$ definimos

$$P^*(B) = \frac{P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times B \times (-\infty, x_{r+2}] \times \dots \times (-\infty, x_k])}{P_{\mathbf{X}}(A)}.$$

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

$$\begin{aligned} & P^* \left(\bigsqcup_{n \in \mathbb{N}} C_n \right) \\ &= \frac{P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times \bigsqcup_{n \in \mathbb{N}} C_n \times (-\infty, x_{r+2}], \dots \times (-\infty, x_k])}{P_{\mathbf{X}}(A)} \\ &= \frac{P_{\mathbf{X}}(\bigsqcup_{n \in \mathbb{N}} (B_1 \times B_2 \times \dots \times B_r \times C_n \times (-\infty, x_{r+2}], \dots \times (-\infty, x_k]))}{P_{\mathbf{X}}(A)} \\ &= \frac{\sum_{n=1}^{\infty} P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times C_n \times (-\infty, x_{r+2}], \dots \times (-\infty, x_k])}{P_{\mathbf{X}}(A)} \\ &= \sum_{n=1}^{\infty} \frac{P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times C_n \times (-\infty, x_{r+2}], \dots \times (-\infty, x_k])}{P_{\mathbf{X}}(A)} \\ &= \sum_{n=1}^{\infty} P^*(C_n). \end{aligned}$$

Esto prueba que P^* es una probabilidad.

Observemos que en las anteriores igualdades se uso, además de que P es una probabilidad, una propiedad de la teoría de conjuntos, fácil de probar:

$$\begin{aligned} & B_1 \times B_2 \times \dots \times B_r \times \bigsqcup_{n \in \mathbb{N}} C_n \times (-\infty, x_{r+2}], \dots \times (-\infty, x_k] \\ &= \bigsqcup_{n \in \mathbb{N}} (B_1 \times B_2 \times \dots \times B_r \times C_n \times (-\infty, x_{r+2}], \dots \times (-\infty, x_k]). \end{aligned}$$

Ahora calcularemos el valor de P^* sobre una semirecta.

Dado que A_r es válida (hipótesis inductiva), si $x \in \mathbb{R}$ se tiene

$$\begin{aligned} & P^*((-\infty, x]) \\ &= \frac{P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times (-\infty, x] \times (-\infty, x_{r+2}], \dots \times (-\infty, x_k])}{P_{\mathbf{X}}(A)} \\ &= \frac{P_{X_1}(B_1) \dots P_{X_r}(B_r) P_{X_{r+1}}((-\infty, x]) \dots P_{X_k}((-\infty, x_k])}{P_{X_1}(B_1) \dots P_{X_r}(B_r) P_{X_{r+1}}(\mathbb{R}) \dots P_{X_k}((-\infty, x_k])} \\ &= P_{X_{r+1}}((-\infty, x]). \end{aligned}$$

Entonces por la unicidad de la extensión como $P_{X_{r+1}}$ y P^* coinciden en las semirectas $(-\infty, x]$ se tendrá 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

$$P_{X_{r+1}}(B_{r+1}) = \frac{P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times B_{r+1} \times (-\infty, x_{r+2}], \dots \times (-\infty, x_k])}{P_{X_1}(B_1) \dots P_{X_r}(B_r) P_{X_{r+1}}(\mathbb{R}) \dots P_{X_k}((-\infty, x_k])}.$$

Luego teniendo en cuenta que estamos suponiendo que A_r es cierta y usando que $P_{X_{r+1}}(\mathbb{R}) = 1$ obtenemos

$$\begin{aligned} & P_{\mathbf{X}}(B_1 \times B_2 \times \dots \times B_r \times B_{r+1} \times (-\infty, x_{r+2}], \dots \times (-\infty, x_k]) \\ &= P_{X_{r+1}}(B_{r+1}) P_{X_1}(B_1) \dots P_{X_r}(B_r) P_{X_{r+2}}(B_{r+2}) \dots P_{X_k}((-\infty, x_k]) \\ &= P_{X_1}(B_1) \dots P_{X_r}(B_r) P_{X_{r+1}}(B_{r+1}) \dots P_{X_k}((-\infty, x_k]), \end{aligned}$$

y luego o también vale $A_{r+1} \square$.

4.4.2. Presección de la independencia por transformaciones.

Teorema. 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 $X_j^{-1}(C_j)$. Luego Y_1, \dots, Y_h son independientes.

4.4.3. Independencia de vectores aleatorios.

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

$$\mathbf{X}_i : \Omega \rightarrow \mathbb{R}^{k_i}, \quad \square = \neq, \neq, \dots, \approx$$

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.

4.4.4.

Proposición 1. 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.

Proposición 2. 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 $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_h) \in \mathbb{R}^{l_1} \times \mathbb{R}^{l_2} \times \dots \times \mathbb{R}^{l_h}$ 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.

Proposición 3. 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, $g_i : \mathbb{R}^{k_i} \rightarrow \mathbb{R}^{l_i}$, $i = 1, 2, \dots, h$. 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.

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. Sea $\mathbf{X} = (X_1, X_2, \dots, X_h)$ un vector aleatorio. Se dice que \mathbf{X} es de tipo discreto o bien que tiene distribución discreta si para cada $i = 1, 2, \dots, h$ X_i es una 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 siguiente proposición 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 el siguiente lema

Lema. 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

$$P\left(\left(\bigcap_{i=1}^h A_i\right)\right) = 1 - P\left(\left(\bigcap_{i=1}^h A_i\right)^c\right) = 1. \square$$

Proposición. Sea $\mathbf{X} = (X_1, X_2, \dots, X_h)$ un vector aleatorio. Entonces el conjunto

$$R_{\mathbf{X}}^* = R_{X_1} \times \dots \times R_{X_h}$$

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

$$\{\omega : \mathbf{X}(\omega) \in R_{X_1} \times \dots \times R_{X_h}\} = \bigcap_{i=1}^h \{\omega : X_i(\omega) \in R_{X_i}\}.$$

Luego por el Lema anterior

$$\begin{aligned} P_{\mathbf{X}}(R_{\mathbf{X}}^*) &= P_{\mathbf{X}}(R_{X_1} \times \dots \times R_{X_h}) = P(\{\omega : \mathbf{X}(\omega) \in R_{X_1} \times \dots \times R_{X_h}\}) \\ &= P\left(\bigcap_{i=1}^h \{\omega : X_i(\omega) \in R_{X_i}\}\right) = 1, \end{aligned}$$

ya que $P(\{\omega : X_i(\omega) \in R_{X_i}\}) = P_{X_i}(R_{X_i}) = 1 \square$.

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.

Consideremos un vector aleatorio $\mathbf{X} = (X_1, X_2)$ que asume los valores $\{(0, 0), (1, 1)\}$ con equiprobabilidad $\frac{1}{2}$. De esto se deduce que las variables aleatorias X_1, X_2 a su vez asumen los valores 0 y 1 con probabilidad $\frac{1}{2}$ 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,

$$R_{\mathbf{X}} = \{\mathbf{x} : P_{\mathbf{X}}(\{\mathbf{x}\}) > 0\} \subset R_{X_1} \times \dots \times R_{X_h}.$$

Mostraremos que $P_{\mathbf{X}}(R_{\mathbf{X}}) = 1$ y que este conjunto es minimal con respecto a esta propiedad. es decir 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 la un vector aleatorio discreto vamos a mostrar que igual que en el caso de una variable aleatoria podemos determinar una función definida sobre \mathbb{R}^k que la determina totalmente $P_{\mathbf{X}}$.

Definición. 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.

1. 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 que si $B \in \mathcal{B}^k$ entonces

$$P_{\mathbf{X}}(B) = \sum_{\mathbf{x} \in B \cap R_{\mathbf{X}}} p_{\mathbf{X}}(\mathbf{x}).$$

2. 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 $R_{\mathbf{X}}$. Luego si $B = B_1 \times B_2 \times \dots \times B_k$, donde $B_1 \dots B_k$ son borelinanos en \mathbb{R} , podemos escribir

$$\begin{aligned} P_{\mathbf{X}}(B) &= \sum_{\mathbf{x} \in B \cap R_{\mathbf{X}}^*} p_{\mathbf{X}}(\mathbf{x}) = \sum_{\mathbf{x} \in B \cap R_{X_1} \times R_{X_2} \times \dots \times R_{X_k}} p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{x_k \in B_k \cap R_{X_k}} \sum_{x_{k-1} \in B_{k-1} \cap R_{X_{k-1}}} \dots \sum_{x_1 \in B_1 \cap R_{X_1}} p_{\mathbf{X}}(x_1, x_2, \dots, x_k). \end{aligned}$$

En particular si $B = \mathbb{R}^k$ obtenemos

$$\begin{aligned} 1 = P_{\mathbf{X}}(\mathbb{R}^k) &= \sum_{\mathbf{x} \in R_{\mathbf{X}}^*} p_{\mathbf{X}}(\mathbf{x}) = \sum_{\mathbf{x} \in R_{X_1} \times R_{X_2} \times \dots \times R_{X_k}} p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{x_k \in R_{X_k}} \sum_{x_{k-1} \in R_{X_{k-1}}} \dots \sum_{x_1 \in R_{X_1}} p_{\mathbf{X}}(\mathbf{x}). \end{aligned}$$

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 $\mathbf{X}^* = (X_1, X_2, \dots, X_h)$.

Proposición. La función de densidad marginal asociada al vector \mathbf{X}^* viene dada por la fórmula

$$p_{\mathbf{X}^*}(\mathbf{x}) = \sum_{x_{h+1} \in R_{X_{h+1}}} \sum_{x_{h+2} \in R_{X_{h+2}}} \dots \sum_{x_k \in R_{X_k}} p_{\mathbf{X}}(x_1, \dots, x_h, x_{h+1}, \dots, x_k).$$

Demostración.

Aplicando la definición de $p_{\mathbf{X}}$

$$\begin{aligned} p_{\mathbf{X}^*}((x_1, x_2, \dots, x_h)) &= P_{\mathbf{X}^*}(\{(x_1, x_2, \dots, x_h)\}) \\ &= P_{\mathbf{X}}(\{\{x_1\} \times \{x_2\} \times \dots \times \{x_h\} \times \mathbb{R} \times \dots \times \mathbb{R}\}). \end{aligned}$$

Entonces de acuerdo al resultado anterior

$$\begin{aligned} p_{\mathbf{X}^*}((x_1, x_2, \dots, x_h)) &= P_{\mathbf{X}}(\{\{x_1\} \times \{x_2\} \times \dots \times \{x_h\} \times \mathbb{R} \times \dots \times \mathbb{R}\}) \\ &= \sum_{x_k \in \mathbb{R} \cap R_{X_k}} \dots \sum_{x_{h+1} \in \mathbb{R} \cap R_{X_{h+1}}} p_{\mathbf{X}}(x_1, \dots, x_h, x_{h+1}, \dots, x_k) \\ &= \sum_{x_k \in R_{X_k}} \dots \sum_{x_{k+1} \in R_{X_{k+1}}} p_{\mathbf{X}}(x_1, \dots, x_h, x_{h+1}, \dots, x_k). \end{aligned}$$

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. 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 Boreleanos $B_j = \{x_j\}$, $j = 1, 2, \dots, h$ 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 Boreleanos, entonces

$$\begin{aligned} P_{\mathbf{X}}(B_1 \times B_2) &= \sum_{x_1 \in B_1 \cap R_{X_1}} \sum_{x_2 \in B_2 \cap R_{X_2}} p_{\mathbf{X}}(x_1, x_2) \\ &= \sum_{x_1 \in B_1 \cap R_{X_1}} \sum_{x_2 \in B_2 \cap R_{X_2}} p_{X_1}(x_1) \cdot p_{X_1}(x_2) \\ &= \left(\sum_{x_1 \in B_1 \cap R_{X_1}} p_{X_1}(x_1) \right) \left(\sum_{x_2 \in B_2 \cap R_{X_2}} p_{X_1}(x_2) \right). \end{aligned}$$

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 distribuciones discretas.

5.2.1. Multinomial (Binomial Multivariada o Generalizada) $\mathbf{M}(p_1, \dots, p_k, n)$.

Supongamos que un experimento que tiene a 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 $p_i = \frac{1}{6}$, $i = 1, \dots, 6$. 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 conformemos el vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_k)$.

Como espacio muestral consideremos

$$\Omega = \{(i_1, i_2, \dots, i_n) : i_j \in \mathbb{N}, 1 \leq i_j \leq k\}.$$

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 $X_j : \Omega \rightarrow \mathbb{N}$ está definida 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. Como suponemos independencia entre los experimentos, y el vector (i_1, i_2, \dots, i_n) indica la intersección de los n eventos

$$B_j = \{\text{en el experimento } j \text{ el resultado fue } i_j\}, \quad j = 1, \dots, n$$

y el evento B_j tiene probabilidad p_j , resulta

$$p_{\mathbf{X}}(i_1, i_2, \dots, i_n) = p_{i_1} p_{i_2} \dots p_{i_n} \quad (5.3)$$

Fijado $\mathbf{x} = (x_1, \dots, x_k)$ tal que $\sum_{i=1}^k x_i = n$, calcularemos la probabilidad del evento

$$\begin{aligned} A &= \{(i_1, i_2, \dots, i_n) \in \Omega : \mathbf{X}(i_1, i_2, \dots, i_n) \\ &= (x_1, x_2, \dots, x_k)\}, \end{aligned}$$

donde

$$\begin{aligned} \mathbf{X}(i_1, i_2, \dots, i_n) &= (X_1(i_1, i_2, \dots, i_n), X_2(i_1, i_2, \dots, i_n), \dots, X_k(i_1, i_2, \dots, i_n)) \\ &= (x_1, x_2, \dots, x_k), \end{aligned}$$

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.

Ahora podemos escribir usando (5.3), la probabilidad de cualquier elemento del evento Ω como

$$p_{\mathbf{X}}(i_1, i_2, \dots, i_n) = p_1^{X_1(i_1, i_2, \dots, i_n)} p_2^{X_2(i_1, i_2, \dots, i_n)} \dots p_k^{X_k(i_1, i_2, \dots, i_n)}.$$

En particular si $(i_1, i_2, \dots, i_n) \in A$, se tendrá

$$p_{\mathbf{X}}(i_1, i_2, \dots, i_n) = p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}.$$

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

$$\begin{aligned} \#A &= \binom{n}{x_1} \binom{n-x_1}{x_2} \binom{n-x_1-x_2}{x_3} \dots \binom{x_k}{x_k} = \\ &= \frac{n!}{(x_1)!(n-x_1)!(x_2)!(n-x_1-x_2)!(x_3)!(n-x_1-x_2-x_3)!} \dots 1 \\ &= \frac{n!}{(x_1)!(x_2)!(x_3)! \dots (x_k)!}. \end{aligned}$$

Luego tendremos

$$p_{\mathbf{X}}(x_1, x_2, \dots, x_k) = P_{\mathbf{X}}(A) = \frac{n!}{(x_1)!(x_2)!(x_3)! \dots (x_k)!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}.$$

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 , ..., D_k de la clase A_k . Desde luego $\sum_{i=1}^k D_i = N$. Supongamos que se realizan n extracciones y sea X_i la "cantidad de objetos de la clase i que se obtuvieron en las n extracciones". y 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 $p_i = \frac{D_i}{N}$ que será denotada por $M(p_1, \dots, p_n, n)$.

b) Las extracciones se hacen *sin reposición*. En este caso la distribución se denomina hipergeométrica multivariada y sera denotada por $HGM(D_1, \dots, D_k, n)$.

Calculemos el rango del vector \mathbf{X}

$$R_{\mathbf{X}} = \{(x_1, x_2, \dots, x_k) : 0 \leq x_i \leq D_i, x_1 + x_2 + \dots + x_k = n\}.$$

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

$$C = \{A : \mathbf{X}(A) = (x_1, x_2, \dots, x_k)\}.$$

C representan 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. 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

$$\begin{aligned} F_{\mathbf{X}}(x_1, x_2, \dots, x_k) &= \int_{-\infty}^{x_k} \int_{-\infty}^{x_{k-1}} \dots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_1 dt_2 \dots dt_k = \\ &= \int \dots \int_{(-\infty, x_k] \times (-\infty, x_{k-1}] \times \dots \times (-\infty, x_1]} f_{\mathbf{X}}(\mathbf{t}) d\mathbf{t}, \end{aligned}$$

donde $\mathbf{t} = (t_1, t_2, \dots, t_k)$ y $d\mathbf{t} = dt_1 dt_2 \dots dt_k$.

Usaremos esta última notación cuando convenga.

Observaciones.

1. Análogamente al caso univariado se tendrá

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(\mathbf{t}) d\mathbf{t} = P_{\mathbf{X}}(\mathbb{R}^k) = 1.$$

2. Supongamos que $a_1 < b_1$, $a_2 < b_2$, $a_3 < b_3$, ..., $a_k < b_k$ y consideremos el rectángulo $R = (a_1, b_1] \times (a_2, b_2] \times \dots \times (a_k, b_k]$. Luego la probabilidad de R se obtiene integrando la función densidad sobre R . Más explícitamente

$$\begin{aligned} P_{\mathbf{X}}((a_1, b_1] \times (a_2, b_2] \times \dots \times (a_k, b_k]) &= \Delta_{a_k}^{b_k}(x_k) \dots \Delta_{a_1}^{b_1}(x_1) F_{\mathbf{X}}(x_1, x_2, \dots, x_k) \\ &= \int_{a_k}^{b_k} \int_{a_{k-1}}^{b_{k-1}} \dots \int_{a_1}^{b_1} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_1 dt_2 \dots dt_k. \end{aligned}$$

3. Se puede probar, mediante teoría de la medida e integración que para todo Boreliano $B \in \mathcal{B}^k$

$$P_{\mathbf{X}}(B) = \int \dots \int_B f_{\mathbf{X}}(\mathbf{t}) d\mathbf{t}.$$

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. 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

$$\lim_{h \rightarrow 0} \frac{P_{\mathbf{X}}([x_{10} - h, x_{10} + h] \times \dots \times [x_{k0} - h, x_{k0} + h])}{(2h)^k} = f_{\mathbf{X}}(\mathbf{x}_0).$$

Demostración

Es análoga al caso univariado y se deja como ejercicio.

Observación. Los entornos cúbicos se pueden reemplazar por otro tipo de entornos, por ejemplo entornos esféricos.

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. Supongamos que $f_{\mathbf{X}}$ se continua en \mathbf{x}_0 . Entonces

$$f_{\mathbf{X}}(\mathbf{x}_0) = \left. \frac{\partial^K F_{\mathbf{X}}(x_1, x_2, \dots, x_k)}{\partial x_k \partial x_{k-1} \dots \partial x_1} \right|_{\mathbf{x}=\mathbf{x}_0}.$$

Demostración

Es consecuencia del teorema fundamental del cálculo para varias variables.

Por definición se tiene que

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_k) = \int_{-\infty}^{x_k} \int_{-\infty}^{x_{k-1}} \dots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, t_2, \dots, t_k) dt_1 dt_2 \dots dt_k.$$

Dejando fijas todas las variables excepto x_1 y aplicando el teorema fundamental del cálculo se obtiene

$$\frac{\partial F_{\mathbf{X}}(x_1, x_2, \dots, x_k)}{\partial x_1} = \int_{-\infty}^{x_k} \int_{-\infty}^{x_{k-1}} \dots \int_{-\infty}^{x_2} f_{\mathbf{X}}(x_1, t_2, \dots, t_k) dt_2 \dots dt_k.$$

Ahora manteniendo fijas todas las variables excepto x_2

$$\frac{\partial F_{\mathbf{X}}(x_1, x_2, \dots, x_k)}{\partial x_1} = \int_{-\infty}^{x_k} \int_{-\infty}^{x_{k-1}} \dots \int_{-\infty}^{x_3} f_{\mathbf{X}}(x_1, x_2, \dots, t_k) dt_3 \dots dt_k.$$

Derivando k veces se llega al resultado deseado

Definición. Dado un Boreliano $B \in \mathcal{B}^k$ se define su *volumen* de la siguiente manera

$$\text{Vol}(B) = \int \dots \int_B dx_1 dx_2 \dots dx_k = \int \dots \int_B d\mathbf{x}.$$

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 numerable de conjuntos de volumen cero tiene 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. 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 $B = (B \cap C_n) \uplus (B \cap C'_n)$ con lo cual se tiene

$$P_{\mathbf{X}}(B) = P_{\mathbf{X}}(B \cap C_n) + P_{\mathbf{X}}(B \cap C'_n).$$

Ahora calculamos $P_{\mathbf{X}}(B \cap C'_n)$. Para ello observemos que para todo $n \in \mathbb{N}$

$$\begin{aligned} P(B \cap C'_n) &= \int \cdots \int_{B \cap C'_n} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} \leq \\ & n \int \cdots \int_{B \cap C'_n} d\mathbf{x} \\ &= n \text{Vol}(B \cap C'_n) \\ &\leq n \text{Vol}(B) \\ &= 0. \end{aligned}$$

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 $P_{\mathbf{X}}(B) = 0$.

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. 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

$$f_{\mathbf{X}^*}(x_1, x_2, \dots, x_h) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x_1, x_2, \dots, x_k) dx_{h+1} dx_{h+2} \dots dx_k. \quad (5.4)$$

Observación.

Por comodidad hemos escogido las primeras h componentes pero lo mismo puede hacerse para una colección arbitraria. En el caso de una distribución bivariada $\mathbf{X} = (X_1, X_2)$ $\mathbf{X}^* = X_1$

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f_{\mathbf{X}}(x_1, x_2) dx_2.$$

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

$$F_{\mathbf{X}^*}(x_1, x_2, \dots, x_h) = \int_{-\infty}^{x_h} \dots \int_{-\infty}^{x_1} f_{X^*}(t_1, t_2, \dots, t_h) dt_1 \dots dt_h,$$

donde $f_{\mathbf{X}^*}$ está dada por (5.4). Esto prueba el Teorema. \square .

Capítulo 6

Transformaciones de variables y vectores aleatorios.

En esta sección estudiaremos como se obtienen las distribuciones de variables o vectores obtenidos a través de ciertas tipo de transformaciones.

6.1. Transformaciones 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 .

(1) 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 $b = \infty$.

Si $y \in (a, b)$ entonces teniendo en cuenta que es estricta monótona creciente se tiene

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

Luego es fácil ver que

$$F_Y(y) = \begin{cases} 0 & \text{si } y \leq a \\ F_X(g^{-1}(y)) & \text{si } y \in (a, b) \\ 1 & \text{si } y \geq b. \end{cases} \quad (6.1)$$

(2) Caso de g estrictamente decreciente.

Nuevamente la imagen de g es un abierto (a, b) de longitud finita o infinita. Pero ahora como g es estrictamente decreciente se tiene para $a < y < b$ que

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \geq g^{-1}(y)) = 1 - P(X < g^{-1}(y)).$$

Luego

$$F_Y(y) = \begin{cases} 0 & \text{si } y \leq a \\ 1 - P(X < g^{-1}(y)) & \text{si } y \in (a, b) \\ 1 & \text{si } y \geq b \end{cases} \quad (6.2)$$

Observación.

No suponemos de entrada que X sea continua. En el caso de que X sea continua se tiene que

$$P(X < g^{-1}(y)) = P(X \leq g^{-1}(y)) = 1 - F_X(g^{-1}(y)),$$

y por lo tanto

$$F_Y(y) = \begin{cases} 0 & \text{si } y \leq a \\ 1 - F_X(g^{-1}(y)) & \text{si } y \in (a, b) \\ 1 & \text{si } y \geq b. \end{cases} \quad (6.3)$$

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.

Distinguiamos dos casos como antes

(1) Supongamos que $g' > 0$. Derivando (6.1) obtenemos

$$f_Y(y) = \begin{cases} 0 & \text{si } y \leq a \\ \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} & \text{si } y \in (a, b) \\ 1 & \text{si } y \geq b. \end{cases} \quad (6.4)$$

(2) Supongamos $g' < 0$. Derivando (6.3) obtenemos

$$f_Y(y) = \begin{cases} 0 & \text{si } y \leq a \\ -\frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} & \text{si } y \in (a, b) \\ 1 & \text{si } y \geq b \end{cases}$$

Las fórmulas (6.4) y (6.5) pueden resumirse en

$$f_Y(y) = \begin{cases} 0 & \text{si } y \leq a \\ \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|} & \text{si } y \in (a, b) \\ 1 & \text{si } y \geq b. \end{cases} \quad (6.5)$$

Esto nos da la expresión para la densidad de una variable aleatoria inducida para g estrictamente monótona.

Un caso especial de interés es cuando la transformación es afín, es decir cuando $g(x) = cx + d$

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

$$f_X(y) = \frac{1}{|c|} f_X\left(\frac{y-d}{c}\right).$$

6.1.1. Distribución Normal

Hemos visto la distribución de una variable normal standarizada $X \approx N(0, 1)$ cuya función densidad es

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

El gráfico de esta función es la llamada campana (ojiba) de Gauss.

Consideremos ahora nuevos parámetros μ , llamada media y σ^2 llamado varianza

Observación. El cuadrado sólo recuerda que se trata de un número positivo.

Ahora podemos considerar la nueva variable $Y = \sigma X + \mu$ ($\sigma > 0$). Estamos en el caso anterior de una transformación afín y por lo tanto

$$\begin{aligned} f_Y(y) &= \frac{1}{\sigma} f_X\left(\frac{y-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y-\mu}{\sigma}\right)^2\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right). \end{aligned}$$

Observación.

Se puede interpretar el efecto de la transformación afín sobre la campana de Gauss. El factor μ representa un desplazamiento horizontal de la densidad e indica de alguna manera el centro alrededor del cual se concentra la densidad, de hecho es el punto de mayor densidad. El factor de proporción σ , afecta la curvatura de la campana y representa la dispersión respecto del centro. Un factor $\sigma \gg 1$ (grande) “achata” la curva hacia el eje x , de manera que en tal caso se presenta gran dispersión. Un factor $1 \gg \sigma$ (chico) hace que la probabilidad esta más concentrada cerca de μ . Más adelante daremos más precisiones sobre el significado de μ y σ^2 .

Ejercicio. Se deja como ejercicio mostrar que si Y tiene distribución $N(\mu, \sigma^2)$, entonces $X = (Y - \mu)/\sigma$ tiene distribución $N(0, 1)$.

6.2. Transformaciones 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.

Luego existe $g^{-1} : V \rightarrow U$. Supongamos que g es diferenciable en cada punto $x \in U$ de forma tal que su Jacobiano sea distinto de cero en cada punto. Es decir

$$J_g(\mathbf{x}) = \det \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_k} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_k} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial g_k}{\partial x_1} & \frac{\partial g_k}{\partial x_2} & \cdots & \frac{\partial g_k}{\partial x_k} \end{pmatrix} \neq 0$$

Entonces g^{-1} es diferenciable y

$$J_{g^{-1}}(\mathbf{y}) = \frac{1}{J_g(g^{-1}(\mathbf{y}))}.$$

El siguiente teorema permite un cambio de variables para integrales múltiples.

Teorema. Sea $A \subset U \subset \mathbb{R}^k$ y $f : U \rightarrow \mathbb{R}$ entonces

$$\int \cdots \int_A f(\mathbf{x}) d\mathbf{x} = \int \cdots \int_{g(A)} f(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| d\mathbf{y}.$$

donde como siempre $d\mathbf{x} = dx_1 dx_2, \dots, dx_k$ y $d\mathbf{y} = dy_1 dy_2, \dots, dy_k$.

Observación.

Se pide que A sea un dominio de integración, en nuestro caso bastará que sea un conjunto Boreliano.

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. Supongamos que U es un abierto de \mathbb{R}^k tal que $P_{\mathbf{X}}(U) = 1$, esto significa que la probabilidad se concentra en U , es decir con probabilidad 1 los valores del vector \mathbf{X} están en U y en consecuencia la densidad $f_{\mathbf{X}}$ fuera de U es cero.

Consideremos ahora $g : U \rightarrow \mathbb{R}^k$ inyectiva diferenciable tal que para todo $x \in U$ se tiene $J_g(\mathbf{x}) \neq 0$. El siguiente teorema permitirá encontrar la distribución del vector $\mathbf{Y} = g(\mathbf{X})$. Denotemos por I_V es la función característica o indicadora de un conjunto conjunto V)

$$I_V(\mathbf{y}) = \begin{cases} 1 & \text{si } \mathbf{y} \in V \\ 0 & \text{si } \mathbf{y} \notin V \end{cases}$$

Teorema. Supongamos dadas las condiciones anteriores y definamos $\mathbf{Y} = g(\mathbf{X})$. Luego la distribución de este vector resulta ser también absolutamente continua y se tendrá.

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| I_V(\mathbf{y}),$$

donde $V = g(U)$

Demostración. Para esto bastará demostrar que si $B \in \mathcal{B}^k$ entonces

$$P_{\mathbf{Y}}(B) = \int \cdots \int_B f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| I_V(\mathbf{y}) d\mathbf{y}. \quad (6.6)$$

Usando el Teorema de cambio de variable en integrales múltiples tenemos 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 razón 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

$$\begin{aligned} P_{\mathbf{Y}}(B) &= \int \cdots \int_{g(g^{-1}(B \cap V))} f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| d\mathbf{y} \\ &= \int \cdots \int_{B \cap V} f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| d\mathbf{y}. \end{aligned}$$

Esto muestra que vale (6.6). \square

El anterior resultado 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 podremos 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 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 $g_i : \mathbb{R}^k \rightarrow \mathbb{R}^k$, $i = j+1, \dots, k$ 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 $\tilde{g}(x_1, x_2) = (x_1 + x_2, x_2)$ y $\mathbf{Y} = (Y_1, Y_2)$ 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 su Jacobiano es

$$J_{\tilde{g}}(x_1, x_2) = \det \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = 1.$$

con inversa $\tilde{g}^{-1}(y_1, y_2) = (y_1 - y_2, y_2)$ diferenciable. También tenemos $X_1 = Y_1 - Y_2$ y $X_2 = Y_2$.

De acuerdo a lo que hemos visto

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(\tilde{g}^{-1}(\mathbf{y})) |J_{\tilde{g}^{-1}}(\mathbf{y})| I_V(\mathbf{y}) = \\ &= f_{\mathbf{X}}(y_1 - y_2, y_2) I_V(y_1, y_2) \end{aligned}$$

y

$$f_Y(y) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(y - y_2, y_2) I_V(y, y_2) dy_2.$$

6.2.1. Algunas aplicaciones.

1. Sea $\mathbf{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio tal que sus componentes son variables aleatorias independientes con idéntica distribución $N(0, 1)$. Entonces la función de densidad asociada al vector \mathbf{X} es

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{\sqrt{(2\pi)^k}} \prod_{i=1}^k \exp\left(-\frac{1}{2}x_i^2\right) = \\ &= \frac{1}{\sqrt{(2\pi)^k}} \exp\left(\sum_{i=1}^k x_i^2\right) = \\ &= \frac{1}{\sqrt{(2\pi)^k}} \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right) \end{aligned}$$

Ahora consideremos una matriz $A \in \mathbb{R}^{k \times k}$ ortogonal, esto es que $AA^t = I$.

Definimos la función $g : \mathbb{R}^k \rightarrow \mathbb{R}^k$ definida por $g(\mathbf{x}) = \mathbf{x}A$ y consideramos el vector aleatorio $\mathbf{Y} = \mathbf{X}A = g(\mathbf{X})$. Calculando el Jacobiano de g vemos que $J_g(\mathbf{x}) = \det A = \pm 1$, de manera que la densidad queda

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}(g^{-1}(\mathbf{y})) |J_{g^{-1}}(\mathbf{y})| I_{\mathbb{R}^k}(\mathbf{y}) \\ &= f_{\mathbf{X}}(g^{-1}(\mathbf{y})) = \frac{1}{\sqrt{(2\pi)^k}} \exp\left(-\frac{1}{2}\|\mathbf{y}\|^2\right). \end{aligned}$$

Es decir las transformaciones ortogonales preservan las distribuciones $N(0, 1)$ independientes.

2. Hemos visto anteriormente que si $X \sim N(0, 1)$ e $Y = \sigma X + \mu$ entonces $Y \sim N(\mu, \sigma^2)$. Y recíprocamente si $Y \sim N(\mu, \sigma^2)$ y $X = \frac{Y - \mu}{\sigma}$ entonces $X \sim N(0, 1)$.

Ahora supongamos que Y_1, Y_2, \dots, Y_k son variables aleatorias independientes cada una con distribución $N(\mu_i, \sigma_i)$ $i = 1, 2, \dots, k$ respectivamente. Queremos encontrar la distribución de una transformación afín de las variables Y_i . Es decir la distribución de una variable aleatoria de la forma $Z = \sum_{i=1}^k \alpha_i Y_i + \beta$.

Para esto necesitamos probar el siguiente lema

Lema. Sean X_1, X_2, \dots, X_k variables aleatorias independientes idénticamente distribuidas, con distribución $N(0, 1)$. Sean b_1, b_2, \dots, b_k números reales tales que $\sum_{i=1}^k b_i^2 = 1$. Luego la variable aleatoria $Z = \sum_{i=1}^k b_i X_i$ tiene distribución $N(0, 1)$.

Demostración

Sea $\mathbf{a}_1 = (b_1, b_2, \dots, b_k)'$, donde $'$ indica transpuesto. 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 unitarios tales que $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_k\}$ es una base ortonormal de \mathbb{R}^k .

Luego la matriz \mathbf{A} cuyas columnas son los vectores \mathbf{a}_j , $j = 1, 2, \dots, k$ es una matriz ortogonal. Definamos el vector aleatorio $\mathbf{Y} = \mathbf{X}\mathbf{A}$, y sea Y_i su componente i . 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 $N(0, 1)$ \square .

Ahora podemos probar el siguiente Teorema

Teorema. Sean Y_1, Y_2, \dots, Y_k variables aleatorias independientes cada una con distribución $N(\mu_i, \sigma_i)$, $i = 1, 2, \dots, k$ respectivamente. Entonces $Z = \sum_{i=1}^k \alpha_i Y_i + \beta$ tiene distribución

$$N\left(\sum_{i=1}^k \alpha_i \mu_i + \beta, \sum_{i=1}^k \alpha_i^2 \sigma_i^2\right).$$

Demostración. Escribimos

$$Z = \sum_{i=1}^k \alpha_i \frac{Y_i - \mu_i}{\sigma_i} \sigma_i + \beta + \sum \alpha_i \mu_i = \sum_{i=1}^k \alpha_i \sigma_i X_i + \delta,$$

donde $X_i = (Y_i - \mu_i)/\sigma_i$ y

$$\delta = \beta + \sum \alpha_i \mu_i \quad (6.7)$$

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

$$A = \left(\sum_{i=1}^k \alpha_i^2 \sigma_i^2 \right)^{\frac{1}{2}}. \quad (6.8)$$

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 al lema anterior se tendrá que

$$W = \sum_{i=1}^k b_i X_i$$

tiene distribución $N(0, 1)$. Por lo tanto como

$$Z = A \sum_{i=1}^k \frac{\alpha_i \sigma_i}{A} X_i + \delta = AW + \delta \sim N(\delta, A^2),$$

se tendrá que Z tiene distribución $N(\delta, A^2)$. Luego el teorema se deduce de (6.7) y (6.8). \square

6.2.2. Transformaciones no inyectivas

Sea $g : U \subset \mathbb{R}^k \rightarrow \mathbb{R}^k$ con U abierto y el vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_k)$ con distribución absolutamente continua. Supongamos que exten h abiertos U_1, U_2, \dots, U_h disjuntos de a dos, es decir $U_i \cap U_j = \emptyset$ si $i \neq j$, tales que $P\left(\bigsqcup_{i=1}^h U_i\right) = 1$.

Ahora supongamos también que

$$g : \bigsqcup_{i=1}^h U_i \rightarrow \mathbb{R}^k$$

es inyectiva y diferenciable en en cada U_i , con jacobiano no nulo. Llamemos $g_i = g|_{U_i}$, $V_i = g(U_i)$ y consideremos $\mathbf{Y} = g(\mathbf{X})$. En este caso existe la inversa de cada g_i que llamaremos $g_i^{-1} : V_i \rightarrow U_i$ y se tendrá

$$J_{g_i^{-1}}(\mathbf{y}) = J_{g_i}(g_i^{-1}(\mathbf{y})) \quad \forall \mathbf{y} \in V_i$$

Teorema. Bajo las condiciones anteriores la distribución de \mathbf{Y} es absolutamente continua y su densidad está dada por

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{i=1}^h f_{\mathbf{X}}(g_i^{-1}(\mathbf{y})) |J_{g_i^{-1}}(\mathbf{y})| I_{V_i}(\mathbf{y})$$

Demostración. Bastará probar probar que para todo $B \in \mathcal{B}^k$ se tiene

$$P_{\mathbf{Y}}(B) = \int \cdots \int_B \sum_{i=1}^h f_{\mathbf{X}}(g_i^{-1}(\mathbf{y})) |J_{g_i^{-1}}(\mathbf{y})| I_{V_i}(\mathbf{y}) d\mathbf{y} \quad (6.9)$$

Usando que $P\left(\bigsqcup_{i=1}^h U_i\right) = 1$ y

$$\{\mathbf{Y} \in B\} \cap \{\mathbf{X} \in U_i\} = \{\mathbf{Y} \in B \cap V_i\} \cap \{\mathbf{X} \in U_i\} = \{\mathbf{X} \in g_i^{-1}(B \cap V_i)\}$$

obtenemos

$$\begin{aligned} P_{\mathbf{Y}}(B) &= P(\mathbf{Y} \in B) \\ &= P\left(\bigsqcup_{i=1}^h \{\mathbf{Y} \in B\} \cap \{\mathbf{X} \in U_i\}\right) \\ &= \sum_{i=1}^h P(\{\mathbf{Y} \in B\} \cap \{\mathbf{X} \in U_i\}) \\ &= \sum_{i=1}^h P(\mathbf{X} \in g_i^{-1}(B \cap V_i)) \\ &= \sum_{i=1}^h P_{\mathbf{X}}(g_i^{-1}(B \cap V_i)) \\ &= \sum_{i=1}^h \int \cdots \int_{g_i^{-1}(B \cap V_i)} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \end{aligned}$$

Como las funciones g_i son biunívocas en cada U_i , usando la fórmula de cambio de variables en integrales múltiples se tiene

$$\begin{aligned} P_{\mathbf{Y}}(B) &= \sum_{i=1}^h \int \cdots \int_{g_i^{-1}(B \cap V_i)} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \sum_{i=1}^h \int \cdots \int_{B \cap V_i} f_{\mathbf{X}}(g_i^{-1}(\mathbf{y})) |J_{g_i^{-1}}(\mathbf{y})| d\mathbf{y} \\ &= \int \cdots \int_B \sum_{i=1}^h f_{\mathbf{X}}(g_i^{-1}(\mathbf{y})) |J_{g_i^{-1}}(\mathbf{y})| I_{V_i}(\mathbf{y}) d\mathbf{y}, \end{aligned}$$

y por lo tanto se cumple (6.9). \square

6.2.3. Distribución Chi-cuadrado con un grado de libertad.

Esta aplicación tiene gran importancia en estadística.

Sea $X \sim N(0, 1)$ y consideremos $g : \mathbb{R} \rightarrow \mathbb{R}$ $g(x) = x^2$. Definimos $Y = g(\mathbf{X}) = X^2$

Definimos $U_1 = \{x : x < 0\}$ 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

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) \frac{1}{2}y^{-\frac{1}{2}} I_{V_1}(y) + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) \frac{1}{2}y^{-\frac{1}{2}} I_{V_2}(y) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) y^{-\frac{1}{2}} I_{\{y: y>0\}}(y). \end{aligned}$$

6.3. Algunas distribuciones complementarias.

6.3.1. Distribución Gamma

En primer lugar introducimos la función Gamma, que resulta ser una extensión del concepto de factorial sobre los números naturales.

Definimos la función

$$\Gamma : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 0}$$

de la siguiente manera

$$\Gamma(\alpha) = \int_0^{+\infty} \exp(-x) x^{\alpha-1} dx.$$

Para probar la existencia de este integral la descomponemos como

$$\begin{aligned} \Gamma(\alpha) &= \int_0^1 \exp(-x) x^{\alpha-1} dx + \int_1^{+\infty} \exp(-x) x^{\alpha-1} dx \\ &= I + II. \end{aligned}$$

Es fácil ver que I es finita, teniendo en cuenta que $\exp(-x) \leq 1$ sobre $(0, 1)$

$$I = \int_0^1 \exp(-x) x^{\alpha-1} dx \leq \int_0^1 x^{\alpha-1} dx = \frac{x^\alpha}{\alpha} \Big|_0^1 = \frac{1}{\alpha}.$$

Para estudiar la convergencia de la segunda integral observemos que de acuerdo al desarrollo de Taylor

$$\exp\left(\frac{x}{2}\right) = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{x}{2}\right)^k$$

se tiene que por tratarse de una serie de términos positivos, para todo $k \in \mathbb{N}$

$$\exp\left(\frac{x}{2}\right) \geq \frac{1}{k!} \left(\frac{x}{2}\right)^k$$

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 $k > \alpha - 1$, luego se obtiene

$$\begin{aligned} II &= \int_1^{+\infty} \exp(-x) x^{\alpha-1} dx \\ &\leq \int_1^{+\infty} \exp(-x) C_k \exp\left(\frac{x}{2}\right) dx \\ &\leq C_k \int_1^{+\infty} \exp\left(\frac{-x}{2}\right) dx < \infty. \end{aligned}$$

Algunas propiedades.

P1. Si $\alpha > 0$ entonces $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$

Para ello 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

$$\begin{aligned}\Gamma(\alpha + 1) &= \int_0^{+\infty} \exp(-x) x^\alpha dx \\ &= \int_0^{+\infty} u dv \\ &= -x^\alpha \exp(-x) \Big|_0^\infty - \alpha \int_0^{+\infty} (-\exp(-x)) x^{\alpha-1} \\ &= -x^\alpha \exp(-x) \Big|_0^\infty + \alpha \int_0^{+\infty} \exp(-x) x^{\alpha-1}.\end{aligned}$$

Como $\lim_{x \rightarrow \infty} x^\alpha \exp(-x) = 0$, resulta que $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$.

P2. Γ es una extensión del factorial. Más precisamente para todo $n \in \mathbb{N}$ se tiene $\Gamma(n) = (n-1)!$. 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 P1 y la hipótesis inductiva tenemos $\Gamma(n+1) = n\Gamma(n) = n((n-1)!) = n!$, con lo cual la propiedad queda demostrada.

Definición. Se define la distribución Gamma con parámetros $\alpha > 0$ y 1 ($\Gamma(\alpha, 1)$) como la distribución absolutamente continua cuya función densidad es

$$f(x) = \frac{1}{\Gamma(\alpha)} \exp(-x) x^{\alpha-1} I_{[0, \infty)}.$$

Observaciones.

1. De acuerdo a la definición de la función Gamma es claro que f es una densidad ya que

$$\int_{-\infty}^{+\infty} f(x) = 1.$$

2. Una vez definida la distribución $\Gamma(\alpha, 1)$ definiremos la distribución $\Gamma(\alpha, \lambda)$ para todo $\lambda > 0$. Para ello consideremos $X \sim \Gamma(\alpha, 1)$. Entonces la distribución $\Gamma(\alpha, \lambda)$ se define como la distribución de $Y = X/\lambda$. Tomando $g(x) = x/\lambda$ y usando la fórmula para la transformación de densidades se tiene

$$\begin{aligned}f_Y(y) &= f_X(g^{-1}(y)) |J_{g^{-1}}(y)| I_{[0, \infty)}(y) = \\ &= \frac{1}{\Gamma(\alpha)} \exp(-\lambda y) (\lambda y)^{\alpha-1} |\lambda| I_{[0, \infty)}(y) = \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \exp(-\lambda y) y^{\alpha-1} I_{[0, \infty)}(y).\end{aligned}$$

3. Recordemos que si $X \sim N(0, 1)$ entonces $Y = X^2$ tiene una distribución chi-cuadrado con un grado de libertad. Más precisamente

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} \exp\left(-\frac{y}{2}\right) I_{[0, \infty)}(y). \quad (6.10)$$

Ahora bien si consideramos $Z \sim \Gamma(1/2, 1/2)$ entonces su densidad es

$$\begin{aligned} f_Z(z) &= \frac{\left(\frac{1}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)} \exp\left(-\frac{z}{2}\right) y^{-\frac{1}{2}} I_{[0, \infty)}(z) = \\ &= \frac{1}{\sqrt{2}\Gamma\left(\frac{1}{2}\right)} \exp\left(-\frac{z}{2}\right) y^{-\frac{1}{2}} I_{[0, \infty)}(z), \end{aligned} \quad (6.11)$$

Las densidades (6.10) y (6.11) difieren solo en una constante, luego deben ser iguales. Esto se muestra integrando las densidades sobre \mathbb{R} , ya que ambas integrales deben ser 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 lema

Lema. 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:

- (a) $f_{W_2} = g_2$ y por lo tanto g_2 es una función de densidad.
- (b) $f_{W_1} = g_1$
- (c) Además 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 (a). Para ver (b) se usa el mismo argumento . Como (a) y (b) implican que

$$f_{\mathbf{W}}(w_1, w_2) = f_{W_1}(w_1)f_{W_2}(w_2),$$

resulta que W_1 y W_2 son independientes.

Lema. 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

- (a) La distribución de W_1 es $W \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$
- (b) W_2 tiene densidad

$$\frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} w_2^{\alpha_1-1} (1 - w_2)^{\alpha_2-1} I_{[0,1]}(w_2).$$

- (c) W_1 y W_2 son independientes

Demostración. La demostración se basa en el teorema del cambio de variable. Sea el abierto $U \subset \mathbb{R}^{\neq}$ definido por $U = \{(y_1, y_2), y_1 > 0, y_2 > 0\}$. Luego $P_Y(U) = 1$.

Consideremos la transformacin $g : U \rightarrow \mathbb{R}^{\neq}$ 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 y 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 del cambio de variable y el hecho 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

$$\begin{aligned}
& f_{\mathbf{W}}(w_1, w_2) \\
&= \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} \exp(-\lambda w_1) (w_1 w_2)^{\alpha_1 - 1} (w_1 (1 - w_2))^{\alpha_2 - 1} w_1 I_V(w_1, w_2) \\
&= \left[\frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} w_1^{\alpha_1 + \alpha_2 - 1} \exp(-\lambda w_1) I_{(0, \infty)}(w_1) \right] \\
&\cdot \left[\frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} w_2^{\alpha_1 - 1} (1 - w_2)^{\alpha_2 - 1} I_{(0, 1)}(w_2) \right] \\
&g_1(w_1) g_2(w_2)
\end{aligned}$$

donde

$$g_1(w_1) = \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} w_1^{\alpha_1 + \alpha_2 - 1} \exp(-\lambda w_1) I_{(0, \infty)}$$

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 $\Gamma(\alpha_1 + \alpha_2, \lambda)$.

Del segundo factor no podemos a priori afirmar nada. Pero de acuerdo al lema previo resulta que W_1 tiene distribución $\Gamma(\alpha_1 + \alpha_2, \lambda)$ y

$$g_2(w_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} w_2^{\alpha_1 - 1} (1 - w_2)^{\alpha_2 - 1} I_{(0, 1)}(w_2)$$

es la función de densidad de W_2 . Finalmente tenemos que W_1 y W_2 son independientes. \square .

6.4. Distribución Beta:

Definición. Se define la distribución Beta con parámetros α_1, α_2 , que denotaremos por $\beta(\alpha_1, \alpha_2)$, como la distribución absolutamente continua cuya función de densidad es:

$$f(w) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} w^{\alpha_1 - 1} (1 - w)^{\alpha_2 - 1} I_{(0, 1)}(w)$$

Observación. Esta función es una densidad por el resultado anterior. 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 por lo tanto

$$\int_0^1 w^{\alpha_1-1} (1-w)^{\alpha_2-1} dw = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}.$$

Corolario. Sean Y_1, Y_2, \dots, Y_n variables 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 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.

6.5. 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 a los resultados anteriores, 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 $\Gamma(n/2, 1/2)$.

6.5.1. Distribución 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 Student con n grados de libertad, que simbolizaremos con t_n , como la distribución de

$$T = \frac{U}{\sqrt{V/n}}.$$

Se deja como ejercicio de la práctica mostrar que la densidad de T es

$$f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

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.

6.6. Otro tipo de variables aleatorias.

Además de las variables discretas y absolutamente continuas existen otros tipos de variables. Un estudio exhaustivo de los tipos de variables aleatorias o 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 continuas.

6.6.1. Variables aleatorias mixtas.

Definición. 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

$$F = (1 - \delta) F_1 + \delta F_2 \quad (6.12)$$

Observación.

1. Es fácil probar que F es una función de distribución. Es monótona no decreciente y continua a la derecha por ser suma de funciones de ese tipo. 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:

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

2. Veamos ahora que efectivamente una variable aleatoria mixta no es ni absolutamente continua ni discreta.

Sean P_i , la probabilidad asociada F_i , $i = 1, 2$ y R_2 el rango de una variable con distribución F_2 . Luego si P es la probabilidad asociada a F , es fácil ver que

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

Por lo tanto R_2 es numerable y $P_2(R_2) = 1$. Por otro, como R_2 es numerable lado $P_1(R_2) = 0$ $P_{F_1}(R_{F_2}) = 0$. Luego, teniendo en cuenta que $P_F = (1 - \delta) P_{F_1} + \delta P_{F_2}$

$$P(R_2) = (1 - \delta) P_1(R_1) + \delta P_2(R_2) = \delta > 0$$

Por lo que se deduce que F no corresponde a una distribución absolutamente continua.

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

$$P(A) = (1 - \delta) P_1(A) + \delta P_2(A) \leq \delta < 1.$$

Siendo A arbitrario, F no puede ser discreta.

Ejemplo.

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

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.

La combinación convexa en este caso es

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

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

2. Una manera de obtener la distribución mixta (6.12) es la siguiente. 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

$$U = \begin{cases} 0 & \text{con probabilidad } 1 - \delta \\ 1 & \text{con probabilidad } \delta \end{cases}$$

Definimos la variable

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

Veamos que la distribución de X es una combinación convexa de X_1 y X_2 .

Teniendo en cuenta la independencia de las variables

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

