Probabilidad

Universidad de Málaga

Grado en Matemáticas

Curso 2023-2024

Contenidos

TEMA 1

Espacios de probabilidad

1.1. Espacios muestrales y σ-álgebras

Definición 1. Al conjunto de todos los posibles resultados de un experimento aleatorio se le llama **espacio muestral**, y se denota por Ω.

De aquí en adelante, la letra Ω se referirá siempre a un espacio muestral, que no es más que un conjunto arbitrario no vacío.

Definición 2. Se dice que una familia de subconjuntos $\mathcal{A} \subset \mathcal{P}(\Omega)$ es una **σ-álgebra sobre Ω** cuando

(i) $\Omega \in \mathcal{A}$.

(ii) Para todo $A \in \mathcal{A}$ se tiene que $A^c \in \mathcal{A}$.

(iii) Para toda familia numerable $\{A_i\}_{i\in I}$ de elementos de $\mathcal{A}$ se tiene que $\bigcup_{i\in I} A_i \in \mathcal{A}$.

Un **suceso** no es más que un elemento de $\mathcal{A}$.

Definición 3. Dada una familia de subconjuntos $E \subset \mathcal{P}(\Omega)$, la menor σ-álgebra sobre Ω que contiene a E se denomina **σ-álgebra generada por E**, y se denota por $\sigma(E)$.

Proposición 1. *Se verifican las siguientes propiedades:*

(i) *Si $E_1, E_2 \subset \mathcal{P}(\Omega)$ son tales que $E_1 \subset E_2$, entonces $\sigma(E_1) \subset \sigma(E_2)$.*

(ii) *Dado $E \subset \mathcal{P}(\Omega)$, se tiene que $\sigma(E) = E$ si y solo si E es una σ-álgebra sobre Ω.*

Demostración. Se remite a la asignatura *Teoría de la Medida e Integración*. □

Definición 4. Sea (Ω, τ) un espacio topológico. La **σ-álgebra de Borel sobre Ω**, denotada por $\mathcal{B}(\Omega)$ o $\mathcal{B}_\Omega$, es la σ-álgebra generada por τ, es decir, $\mathcal{B}(\Omega) := \sigma(\tau)$.

Proposición 2. *La σ-álgebra de Borel de $\mathbb{R}^n$ junto con la topología usual está generada por cada una de las siguientes familias de intervalos:*

(i) $\mathcal{E}_1 = \{(a,b) = (a_1,b_1) \times \ldots \times (a_n,b_n) : a,b \in \mathbb{R}^n\}$.

(ii) $\mathcal{E}_2 = \{[a,b) = [a_1,b_1) \times \ldots \times [a_n,b_n) : a,b \in \mathbb{R}^n\}$.

(iii) $\mathcal{E}_3 = \{(a,b] = (a_1,b_1] \times \ldots \times (a_n,b_n] : a,b \in \mathbb{R}^n\}$.

(iv) $\mathcal{E}_4 = \{[a,b] = [a_1,b_1] \times \ldots \times [a_n,b_n] : a,b \in \mathbb{R}^n\}$.

(v) $\mathcal{E}_5 = \{(a,+\infty) = (a_1,+\infty) \times \ldots \times (a_n,+\infty) : a \in \mathbb{R}^n\}$.

(vi) $\mathcal{E}_6 = \{[a,+\infty) = [a_1,+\infty) \times \ldots \times [a_n,+\infty) : a \in \mathbb{R}^n\}$.

(vii) $\mathcal{E}_7 = \{(-\infty,b) = (-\infty,b_1) \times \ldots \times (-\infty,b_n) : b \in \mathbb{R}^n\}$.

(viii) $\mathcal{E}_8 = \{(-\infty,b] = (-\infty,b_1] \times \ldots \times (-\infty,b_n] : b \in \mathbb{R}^n\}$.

Demostración. Se remite a la asignatura *Teoría de la Medida e Integración*. □

Definición 5. Dada una σ-álgebra $\mathcal{A}$ sobre Ω, la dupla $(\Omega, \mathcal{A})$ se llama **espacio medible**, mientras que la dupla $(\Omega, \mathcal{B}(\Omega))$, o simplemente $(\Omega, \mathcal{B})$, se denomina **espacio medible-Borel**.

Si no se hace mención alguna sobre cuál es el espacio topológico del que procede una σ-álgebra de Borel $\mathcal{B}$, se supondrá siempre que se trata de $\mathbb{R}$ junto con la topología usual.

1.2. Medidas

Definición 6. Dado un espacio medible $(\Omega, \mathcal{A})$, una **medida positiva** (o simplemente **medida**) es una aplicación $\mu\colon \mathcal{A} \to [0, \infty]$ verificando

(i) $\mu(\emptyset) = 0$.

(ii) Si $\{A_n\}_{n\in\mathbb{N}}$ es una familia de elementos de $\mathcal{A}$ disjuntos dos a dos, entonces

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

Esta propiedad suele ser denominada **σ-aditividad**.

Proposición 3. *Sea $\mu\colon \mathcal{A} \to [0, \infty]$ una medida positiva. Se verifican las siguientes propiedades:*

(i) *Si $A \subset B$, entonces $\mu(A) \leq \mu(B)$. Si además $\mu(A) < \infty$, entonces $\mu(B \setminus A) = \mu(B) - \mu(A)$.*

(ii) *$\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$.*

(iii) *Si $\{A_n\}_{n\in\mathbb{N}}$ es una familia de elementos de $\mathcal{A}$, entonces*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$$

Demostración. Solo se va a demostrar la última propiedad. Dada una familia $\{A_n\}_{n\in\mathbb{N}}$ de elementos de $\mathcal{A}$, se construye otra familia $\{B_n\}_{n\in\mathbb{N}}$, donde $B_n = A_n \setminus A_{n-1}$ para cada $n \geq 2$ y $B_1 = A_1$. Se tiene que los B_n son disjuntos dos a dos, que $B_n \subset A_n$ para cada $n \in \mathbb{N}$ y que

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$$

Consecuentemente,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n),$$

donde se han usado la σ-aditividad y el apartado primero. □

Definición 7. Una medida positiva μ en un espacio medible $(\Omega, \mathcal{A})$ se dice que es **finita** cuando $\mu(E) < \infty$ para todo $E \in \mathcal{A}$, o, equivalentemente, cuando $\mu(\Omega) < \infty$.

Definición 8. Una medida positiva finita P en un espacio medible $(\Omega, \mathcal{A})$ se dice que es una **medida de probabilidad** cuando $P(\Omega) = 1$.

Ejemplo. Sea $\omega \in \Omega$ un elemento fijo de un espacio muestral cualquiera, y considérese la función $\delta_\omega\colon \mathcal{A} \to [0, 1]$ dada por

$$\delta_\omega(A) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \in A^c \end{cases}$$

Se verifica que δ_ω es una medida de probabilidad en cualquier σ-álgebra $\mathcal{A}$ sobre Ω, y se denomina *delta de Dirac*.

Ejemplo. Sea Ω un espacio muestral finito. Se considera la función $\mu\colon \mathcal{A} \to [0,1]$ dada por

$$\mu(A) = \frac{\#A}{\#\Omega}$$

Resulta que μ es una medida de probabilidad en cualquier σ-álgebra $\mathcal{A}$ sobre Ω, denominada *medida uniforme*.

Ejemplo. Sea Ω un espacio muestral numerable y sea $\mathcal{A} = \mathcal{P}(\Omega)$. Sea $\{p_w\}_{w\in\Omega}$ una sucesión de números reales no negativos. Se define la función $\mu\colon \mathcal{A} \to [0,\infty]$ mediante

$$\mu(A) = \sum_{\omega\in A} p_\omega$$

Se verifica que μ es una medida positiva, pero podría no ser de probabilidad. Cuando se tenga $p_w = 1$ para todo $w \in \Omega$, se dirá que μ es la *medida de conteo*.

> **Definición 9.** Sea $\mathcal{A}$ una σ-álgebra sobre Ω y sea P una medida de probabilidad definida sobre $\mathcal{A}$. La terna $(\Omega, \mathcal{A}, P)$ se dice que es un **espacio de probabilidad**.

Funciones de distribución

2.1. Nociones básicas

Definición 10. Una función $F : \mathbb{R} \to [0,1]$ se dice que es una **función de distribución** si

(i) Es monótona creciente.

(ii) Es continua por la derecha.

(iii) $F(-\infty) \equiv \lim_{x\to-\infty} F(x) = 0$.

(iv) $F(+\infty) \equiv \lim_{x\to+\infty} F(x) = 1$.

(v) Existe $F(a^-) \equiv \lim_{x\to a^-} F(x)$ para todo $a \in \mathbb{R}$.

Proposición 4. *Si P es una medida de probabilidad en el espacio medible-Borel $(\mathbb{R}, \mathcal{B})$, entonces la función $F : \mathbb{R} \to [0,1]$ dada por $F(x) = P((-\infty, x])$ es una función de distribución.*

Demostración. Ya se vio en la asignatura *Introducción a la Probabilidad y a la Estadística*. □

Definición 11. Dada una medida de probabilidad P en $(\mathbb{R}, \mathcal{B})$, la función de distribución definida en la proposición anterior se denomina **función de distribución asociada a P**.

Proposición 5. *Sea $F : \mathbb{R} \to [0,1]$ la función de distribución asociada a una medida de probabilidad P. Para cada $a \in \mathbb{R}$, se verifican las siguientes propiedades:*

(i) $P((-\infty, a)) = F(a^-)$.

(ii) $P((a, +\infty)) = 1 - F(a)$.

(iii) $P([a, +\infty)) = 1 - F(a^-)$.

(iv) $p(a) \equiv P(\{a\}) = F(a) - F(a^-)$.

(v) *F es continua en a si y solo si $p(a) = 0$.*

Demostración. Ya se vio en la asignatura *Introducción a la Probabilidad y a la Estadística*. □

Obsérvese que el apartado quinto de la proposición anterior permite escribir el conjunto de puntos de discontinuidad de una función de distribución F asociada a una medida de probabilidad como

$$D_F := \{x \in \mathbb{R} : p(x) > 0\}$$

Proposición 6. *Dada una función de distribución F asociada a una medida de probabilidad, el conjunto D_F es finito o infinito numerable.*

Demostración. Sea $x \in D_F$. Entonces se tiene que $F(x^-) < F(x)$, y por la densidad de $\mathbb{Q}$ en $\mathbb{R}$, existe $q_x \in \mathbb{Q}$ tal que $F(x^-) < q_x < F(x)$. Sea $y \in D_F$ con $y \neq x$ y veamos que $q_y \neq q_x$. Usando que $x, y \in D_F$ y que F es creciente, se tiene que $F(x^-) < q_x < F(x) \leq F(y^-) < q_y < F(y)$, así que es claro que $q_x \neq q_y$. En consecuencia, la función dada por $x \mapsto q_x$ es inyectiva, y como $\mathbb{Q}$ es numerable, entonces D_F también lo es. □

Corolario 1. *El conjunto de puntos de continuidad de una función de distribución F asociada a una medida de probabilidad es denso en $\mathbb{R}$.*

Demostración. Sea C_F el conjunto de puntos de continuidad de F. Hay que demostrar que todo abierto no vacío de $\mathbb{R}$ interseca a C_F. Sea I un abierto no vacío de $\mathbb{R}$. Como D_F es numerable e I es no numerable, entonces $I \setminus D_F = I \cap D_F^c = I \cap C_F$ es no numerable, luego $I \cap C_F \neq \emptyset$ y por tanto C_F es denso en $\mathbb{R}$. □

Definición 12. Dada una función de distribución F asociada a una medida de probabilidad P, el **soporte de F** se define como

$$S_F = \{x \in \mathbb{R} \colon F(x+\varepsilon) - F(x-\varepsilon) > 0 \text{ para todo } \varepsilon > 0\}$$

Nótese que un punto no se encuentra en el soporte de F si y solo si F es constante en algún entorno de dicho punto, lo que permite calcular fácilmente S_F observando la gráfica de F.

2.2. Funciones de distribución discretas

Definición 13. Una función de distribución F asociada a una medida de probabilidad P se dice que es **discreta** cuando

$$\sum_{x \in \mathbb{R}} p(x) = 1$$

Proposición 7. *Sea F una función de distribución discreta asociada a una medida de probabilidad P. Se verifican las siguientes propiedades:*

(i) $P(D_F) = 1$.

(ii) $S_F = D_F$.

(iii) *Si $B \in \mathcal{B}$, entonces*

$$P(B) = \sum_{x \in D_F \cap B} p(x)$$

Demostración. Ejercicio. □

2.3. Funciones de distribución continuas

Definición 14. Una función de distribución F asociada a una medida de probabilidad P se dice que es **continua** cuando

$$\sum_{x \in \mathbb{R}} p(x) = 0$$

Definición 15. Sea F una función de distribución continua asociada a una medida de probabilidad P. La **función de pseudodensidad de F** es la función $\tilde{f} \colon \mathbb{R} \to \mathbb{R}$ definida por

$$\tilde{f}(x) = \begin{cases} F'(x) & \text{si } x \in A \\ 0 & \text{si } x \in A^c, \end{cases}$$

donde A es el conjunto de puntos donde F es derivable.

Proposición 8. *Sea F una función de distribución continua asociada a una medida de probabilidad P y con función de pseudodensidad $\tilde{f}$. Se verifican las siguientes propiedades:*

(i) $D_F = \emptyset$.

(ii) $S_F = \{x \in \mathbb{R} \colon \tilde{f}(x) > 0\} \cup A^c$.

(iii) *Si A^c es numerable y $\tilde{f}$ es integrable, entonces, dado $B \in \mathcal{B}$,*

$$P(B) = \int_B \tilde{f}(t)\,dt$$

Demostración. Otro ejercicio. □

2.4. Funciones de distribución absolutamente continuas

Definición 16. Una función de distribución F asociada a una medida de probabilidad P se dice que es **absolutamente continua** cuando existe una función $f:\mathbb{R}\to\mathbb{R}$ integrable y no negativa tal que

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

Esta función f se denomina **función de densidad de F**.

Proposición 9. *Considérese una función de distribución absolutamente continua F asociada a una medida de probabilidad P y con función de densidad f. Se verifican las siguientes propiedades:*

(i) $D_F=\emptyset$.

(ii) $S_F=\{x\in\mathbb{R}: f(x)>0\}$.

(iii) *Si $B\in\mathcal{B}$, entonces*

$$P(B)=\int_B f(t)dt$$

Demostración. Más ejercicios. □

2.5. Funciones de distribución mixtas

Definición 17. Una función de distribución F asociada a una medida de probabilidad P se dice que es **mixta** cuando

$$0<\sum_{x\in\mathbb{R}} p(x)<1$$

Teorema 1. *Toda función de distribución F admite una descomposición de la forma*

$$F=\alpha F_d+(1-\alpha)F_c,$$

donde $\alpha\in[0,1]$, F_d es una función de distribución discreta y F_c una continua.

Demostración. Se define

$$\alpha=\sum_{x\in D_F} p(x)$$

Si fuese $\alpha=0$ o $\alpha=1$, F sería continua o discreta (respectivamente) y la descomposición sale gratis. Se supone entonces que $0<\alpha<1$. Sea $G:\mathbb{R}\to[0,1]$ la función definida por

$$G(x)=\frac{1}{\alpha}\sum_{\substack{t\in D_F\\ t\le x}} p(t)$$

Es fácil comprobar que G es una función de distribución, y además cumple

$$\sum_{x\in\mathbb{R}}(G(x)-G(x^-))=\sum_{x\in\mathbb{R}}\frac{1}{\alpha}\Bigg(\sum_{\substack{t\in D_F\\ t\le x}} p(t)-\sum_{\substack{t\in D_F\\ t<x}} p(t)\Bigg)=\frac{1}{\alpha}\sum_{x\in\mathbb{R}} p(x)=\frac{1}{\alpha}\alpha=1,$$

luego G es una función de distribución discreta. Se define también la función $H:\mathbb{R}\to[0,1]$ mediante

$$H(x)=\frac{F(x)-\alpha G(x)}{1-\alpha}$$

Es claro que H es función de distribución y que $F(x) = \alpha G(x) + (1-\alpha)H(x)$. Además,

$$F(x) - \alpha G(x) - (F(x^-) - \alpha G(x^-)) = F(x) - F(x^-) - \alpha(G(x) - G(x^-)) = p(x) - \alpha \frac{1}{\alpha} p(x) = 0,$$

de donde se deduce que $H(x) - H(x^-) = 0$, y por tanto H es continua. $\square$

Proposición 10. *Sea F una función de distribución mixta asociada a una medida de probabilidad P. Se verifican las siguientes propiedades:*

(i) $P(D_F) + P(S_F) = 1$.

(ii) *Si $B \in \mathcal{B}$, entonces*

$$P(B) = \sum_{x \in D_F \cap B} p(x) + \int_B \tilde{f}(t)dt,$$

siempre que F_c sea derivable salvo en un conjunto a lo sumo numerable y siempre que la función de pseudodensidad de F_c sea integrable.

Demostración. Tampoco se va a demostrar. $\square$

Ejemplo. Sea $F : \mathbb{R} \to [0,1]$ la función de distribución definida por

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^2}{16} & \text{si } 0 \leq x < 2 \\ \frac{1}{4} & \text{si } 2 \leq x < 4 \\ \frac{x}{4} - \frac{5}{8} & \text{si } 4 \leq x < 5 \\ 1 - \frac{5}{4x} & \text{si } x \geq 5 \end{cases}$$

Se tratará de realizar la descomposición del teorema anterior. En primer lugar, se tiene que

$$D_F = \{4, 5\}, \qquad p(4) = \left(\frac{4}{4} - \frac{5}{8}\right) - \frac{1}{4} = \frac{1}{8} \qquad \text{y} \qquad p(5) = \frac{1}{8},$$

así que $\alpha = \frac{1}{4}$. La parte discreta viene dada por

$$G(x) = \begin{cases} 0 & \text{si } x < 4 \\ \frac{1}{\alpha} p(4) & \text{si } 4 \leq x < 5 \\ \frac{1}{\alpha}(p(4) + p(5)) & \text{si } 5 \leq x \end{cases} = \begin{cases} 0 & \text{si } x < 4 \\ \frac{1}{2} & \text{si } 4 \leq x < 5 \\ 1 & \text{si } 5 \leq x \end{cases}$$

La parte continua sería

$$H(x) = \frac{4}{3}(F(x) - \frac{1}{4} G(x)) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^2}{12} & \text{si } 0 \leq x < 2 \\ \frac{1}{3} & \text{si } 2 \leq x < 4 \\ \frac{x}{3} - 1 & \text{si } 4 \leq x < 5 \\ 1 - \frac{5}{3x} & \text{si } 5 \leq x \end{cases}$$

TEMA 3

Variables aleatorias

3.1. Nociones básicas

Definición 18. Sean $(X, \mathcal{M}_X)$ e $(Y, \mathcal{M}_Y)$ dos espacios medibles. Una función $f : X \to Y$ se dice que es **medible con respecto a $\mathcal{M}_X$ y $\mathcal{M}_Y$** (o simplemente **medible**) si $f^{-1}(B) \in \mathcal{M}_X$ para todo $B \in \mathcal{M}_Y$. Si (X, τ_X) y (X, τ_Y) son espacios topológicos, se dirá que f es una **función de Borel** o que f es **medible-Borel** si es medible con respecto a $\mathcal{B}(X)$ y $\mathcal{B}(Y)$.

Definición 19. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Una **variable aleatoria con respecto a $\mathcal{A}$** (o simplemente **variable aleatoria**) es una función $X : \Omega \to \mathbb{R}$ medible, es decir, tal que $X^{-1}(B) \in \mathcal{A}$ para cada $B \in \mathcal{B}$.

Definición 20. Dada una variable aleatoria $X : \Omega \to \mathbb{R}$ en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, la **medida de probabilidad inducida por X** no es más que la función $P_X : \mathcal{B} \to [0,1]$ definida por $P_X(B) = P(X^{-1}(B))$ para cada $B \in \mathcal{B}$.

Definición 21. Dada una variable aleatoria $X : \Omega \to \mathbb{R}$ en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, la **función de distribución asociada a X** no es más que la función $F_X : \mathbb{R} \to [0,1]$ definida mediante $F_X(a) = P_X((-\infty, a]) = P(X^{-1}((-\infty, a]))$ para cada $a \in \mathbb{R}$.

Definición 22. Sea X una variable aleatoria y sea F la función de distribución asociada a X. El conjunto $D_{F_X} = \{x \in \mathbb{R} : p(x) > 0\}$, también denotado por D_X, se denomina **rango de X**.

En lugar de $P_X((-\infty, a])$, a veces se utiliza la notación $P(X \leq a)$. En el caso de $P_X(\{a\})$, también se escribe $P(X = a)$, y en general, será habitual escribir $P(X \in B)$ en lugar de $P_X(B)$.

Definición 23. Sea $X : \Omega \to \mathbb{R}$ una variable aleatoria en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.

(i) Se dice que X es **discreta** cuando F_X lo es.

(ii) Se dice que X es **continua** cuando F_X lo es.

(iii) Se dice que X es **mixta** cuando F_X lo es.

(iv) Se dice que X es **absolutamente continua** cuando F_X lo es, es decir, cuando existe una función $f_X : \mathbb{R} \to \mathbb{R}$ integrable y no negativa tal que

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

De f_X se dirá que es la **función de densidad de X**.

En cuanto a las variables aleatorias discretas, lo más natural habría sido definirlas como aquellas que toman un número finito o infinito numerable de valores. ¿Será esto equivalente a la definición que se ha dado en la página anterior?

Supóngase que X toma un conjunto infinito numerable de valores distintos, llámense $\{x_i\}_{i=1}^{\infty}$ (en el caso finito se razona igual). Entonces

$$\sum_{x\in\mathbb{R}} p(x) = \sum_{i=1}^{\infty} p(x_i) = \sum_{i=1}^{\infty} P_X(\{x_i\}) = \sum_{i=1}^{\infty} P(X^{-1}(\{x_i\})) = P\Big(\bigcup_{i=1}^{\infty} X^{-1}(\{x_i\})\Big) = P(\Omega) = 1,$$

luego la función de distribución asociada a X es discreta. Para tratar de probar el recíproco, supóngase que

$$\sum_{x\in\mathbb{R}} p(x) = 1$$

Es bien sabido que el conjunto $D(F_X) = \{x \in \mathbb{R} \colon p(x) > 0\}$ es finito o infinito numerable. Se va a suponer que es infinito numerable (el caso finito se razona igual) y se va a considerar una enumeración $\{x_i\}_{i=1}^{\infty}$ de los elementos de $D(F_X)$ (lo que quiere decir que todos los x_i son distintos). Entonces

$$\sum_{x\in\mathbb{R}} p(x) = \sum_{x\in D(F_X)} p(x) = \sum_{i=1}^{\infty} p(x_i) = P\Big(X^{-1}\Big(\bigcup_{i=1}^{\infty}\{x_i\}\Big)\Big) = 1 = P(\Omega),$$

pero esto no implica que se tenga la igualdad

$$\Omega \overset{(*)}{=} X^{-1}\Big(\bigcup_{i=1}^{\infty}\{x_i\}\Big)$$

Sin embargo, lo que sí puede afirmarse es que Ω y $X^{-1}(\bigcup_{i=1}^{\infty}\{x_i\})$ se diferencian en un conjunto de probabilidad cero. Desde el punto de vista probabilístico, estos conjuntos van a ignorarse por completo, y como el rigor matemático en esta asignatura se suele tomar por el pito del sereno, puede afirmarse sin temor alguno que *una variable aleatoria X toma un número finito o infinito numerable de valores si y solo si F_X es una función de distribución discreta*.

Ejemplo. Sea $\Omega = \{w_1, w_2, w_3, w_4\}$ un espacio muestral cualquiera y sea $A = \{w_1, w_2\}$. Se considera la σ-álgebra $\mathcal{A} = \{\Omega, \emptyset, A, A^c\}$ sobre Ω y se define la variable aleatoria $X \colon \Omega \to \mathbb{R}$ mediante

$$X(w_1) = X(w_2) = -1, \qquad X(w_3) = X(w_4) = 1$$

Veamos que, en efecto, X es una variable aleatoria. Se recuerda que $\mathcal{B}$ es la σ-álgebra generada por el conjunto

$$E = \{(-\infty, x] \colon x \in \mathbb{R}\}$$

Por tanto, será suficiente demostrar que para todo $a \in \mathbb{R}$ se verifica que $X^{-1}((-\infty, a]) \in \mathcal{A}$. Tomamos $a \in \mathbb{R}$.

(i) Si $a < -1$, entonces $X^{-1}((-\infty, a]) = \emptyset$, que está en $\mathcal{A}$ por ser σ-álgebra.

(ii) Si $1 \le a < 2$, entonces $X^{-1}((-\infty, a]) = \{w_1, w_2\} = A$, que está en $\mathcal{A}$ por cómo se ha definido.

(iii) Si $2 \le a$, entonces $X^{-1}((-\infty, a]) = \{w_1, w_2, w_3, w_4\} = \Omega$, que está en $\mathcal{A}$ por ser σ-álgebra.

Ya puede afirmarse con valentía que X es una variable aleatoria.

Ejemplo. Sean $\Omega = [0,4]$, $\mathcal{A} = [0,4] \cap \mathcal{B}$. Se define la función $X \colon \Omega \to \mathbb{R}$ mediante

$$X(\omega) = \begin{cases} \omega + 1 & \text{si } 0 \le \omega < 2 \\ 2 & \text{si } 2 \le \omega < 3 \\ 9 - 2\omega & \text{si } 3 \le \omega \le 4 \end{cases}$$

Se tratará de ver que X es variable aleatoria. Sea $a \in \mathbb{R}$.

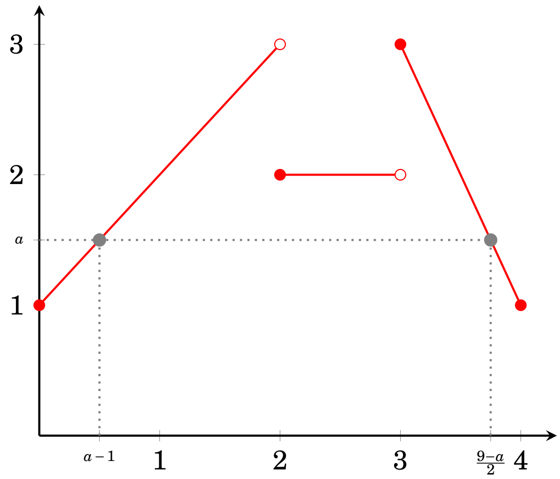
(i) Si $a < 1$, entonces $X^{-1}((-\infty, a]) = \emptyset \in \mathcal{A}$.

(ii) Si $1 \le a < 2$, entonces $X^{-1}((-\infty, a]) = [0, a-1] \cup [\frac{9-a}{2}, 4] \in \mathcal{A}$.

(iii) Si $2 \le a < 3$, entonces $X^{-1}((-\infty, a]) = [0, a-1] \cup [2,3] \cup [\frac{9-a}{2}, 4] \in \mathcal{A}$.

(iv) Si $3 \le a$, entonces $X^{-1}((-\infty, a]) = [0,4] \in \mathcal{A}$.

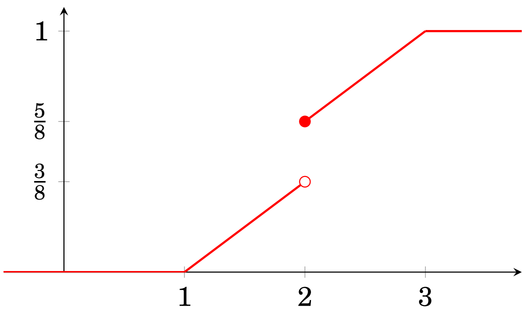
Por tanto, X es variable aleatoria. Su gráfica es



Si se considera ahora la medida de probabilidad dada por $P(A) = \frac{1}{4} l(A)$, con $l(A)$ la longitud de A, la función de distribución asociada a la variable aleatoria X vendría dada por

$$F_X(a) = \begin{cases} 0 & \text{si } a < 1 \\ \frac{3a-3}{8} & \text{si } 1 \le 2 \\ \frac{3a-1}{8} & \text{si } 2 \le a < 3 \\ 1 & \text{si } a \ge 3 \end{cases}$$

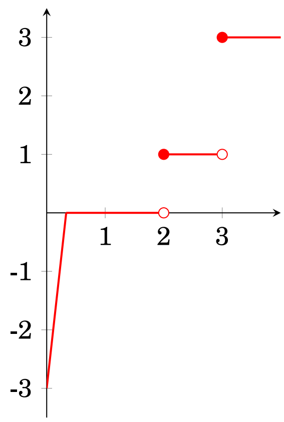
La gráfica de F_X sería



Ejemplo. Sea $\Omega = \mathbb{R}_+$, sea $\mathcal{A} = \mathcal{B} \cap \mathbb{R}_+$ y, para cierto $\lambda > 0$, sea $P : \mathcal{A} \to [0,1]$ la función definida mediante $P([0,\omega]) = P([0,\omega)) = 1 - e^{-\lambda\omega}$. Se considera la variable aleatoria $X : \Omega \to \mathbb{R}$ dada por

$$X(\omega) = \begin{cases} 3(3\omega - 1) & \text{si } 0 \le \omega < \frac{1}{3} \\ 0 & \text{si } \frac{1}{3} \le \omega < 2 \\ 1 & \text{si } 2 \le \omega < 3 \\ 3 & \text{si } 3 \le \omega \end{cases}$$

Se tratará de construir la función de distribución F_X. La gráfica de X es



Sea $a \in \mathbb{R}$.

(*i*) Si $a < -3$, entonces $X^{-1}((-\infty, a]) = \emptyset$, luego $F_X(a) = 0$.

(*ii*) Si $-3 \leq a < 0$, entonces $X^{-1}((-\infty, a]) = [0, \frac{a+3}{9}]$, luego $F_X(a) = 1 - e^{-\lambda(\frac{a+3}{9})}$.

(*iii*) Si $0 \leq a < 1$, entonces $X^{-1}((-\infty, a]) = [0, 2)$, luego $F_X(a) = 1 - e^{-2\lambda}$.

(*iv*) Si $1 \leq a < 3$, entonces $X^{-1}((-\infty, a]) = [0, 3)$, luego $F_X(a) = 1 - e^{-3\lambda}$.

(*v*) Si $3 \leq a$, entonces $F_X(a) = 1$.

En resumidas cuentas,

$$F_X(a) = \begin{cases} 0 & \text{si } a < -3 \\ 1 - e^{-\lambda(\frac{a+3}{9})} & \text{si } -3 \leq a < 0 \\ 1 - e^{-2\lambda} & \text{si } 0 \leq a < 1 \\ 1 - e^{-3\lambda} & \text{si } 1 \leq a < 3 \\ 1 & \text{si } 3 \leq a \end{cases}$$

Ejemplo. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sea $X : \Omega \to \mathbb{R}$ una variable aleatoria discreta cuyos valores son todos los números naturales. Supóngase que la medida de probabilidad inducida por la variable aleatoria X viene dada por $P(X^{-1}(\{n\})) \equiv P(X = n) = p(1-p)^{n-1}$ para todo $n \in N$, siendo $p \in (0,1)$ constante. Se define ahora la función $Y : \Omega \to \mathbb{R}$ que envía cada $\omega \in \Omega$ al resto de la división entera $\frac{X}{3}$. Es claro que Y es también una variable aleatoria discreta, y se tiene que

(*i*) $P(Y = 0) = P(X = 3) + P(X = 6) + P(X = 9) + \ldots = p[(1-p)^2 + (1-p)^5 + (1-p)^8 + \ldots]$.

(*ii*) $P(Y = 1) = P(X = 1) + P(X = 4) + P(X = 7) + \ldots = p[1 + (1-p)^3 + (1-p)^6 + \ldots]$.

(*iii*) $P(Y = 2) = P(X = 2) + P(X = 5) + P(X = 8) + \ldots = p[1 - p + (1-p)^4 + (1-p)^7 + \ldots]$.

Proposición 11. *Sea $\mathcal{S}$ un subconjunto de $\mathcal{B}$ tal que $\sigma(\mathcal{S}) = \mathcal{B}$. Una función $X : \Omega \to \mathbb{R}$ es una variable aleatoria si y solo si $X^{-1}(E) \in \mathcal{A}$ para todo $E \in \mathcal{S}$.*

Demostración. Una de las implicaciones es trivial. Supóngase que $X^{-1}(E) \in \mathcal{A}$ para cada $E \in \mathcal{S}$. Sea $\mathcal{D} = \{D \in \mathcal{B} : X^{-1}(D) \in \mathcal{A}\}$. Probar que X es variable aleatoria es lo mismo que probar que $\mathcal{D} = \mathcal{B}$. Se tiene que $\mathcal{S} \subset \mathcal{D} \subset \mathcal{B}$, luego $\sigma(\mathcal{S}) = \mathcal{B} \subset \sigma(\mathcal{D})$. Basta demostrar entonces que $\mathcal{D}$ es σ-álgebra sobre $\mathbb{R}$.

(*i*) $\mathbb{R} \in \mathcal{D}$ porque $X^{-1}(\mathbb{R}) = \Omega \in \mathcal{A}$ al ser $\mathcal{A}$ σ-álgebra.

(*ii*) Si $D \in \mathcal{D}$, entonces $X^{-1}(D^c) = X^{-1}(D)^c \in \mathcal{A}$ porque $X^{-1}(D) \in \mathcal{A}$ y $\mathcal{A}$ es σ-álgebra. Por tanto, $D^c \in \mathcal{D}$.

(*iii*) Si $\{D_i\}_{i \in I}$ es una familia numerable de elementos de $\mathcal{D}$, entonces

$$X^{-1}\Big(\bigcup_{i \in I} D_i\Big) = \bigcup_{i \in I} X^{-1}(D_i)$$

Por definición de $\mathcal{D}$, se tiene que $X^{-1}(D_i) \in \mathcal{A}$ para cada $i \in I$, y como $\mathcal{A}$ es σ-álgebra, la unión de los $X^{-1}(D_i)$ también es un elemento de $\mathcal{A}$. Esto demuestra que

$$\bigcup_{i \in I} D_i \in \mathcal{D}$$

Se concluye que $\sigma(\mathcal{D}) = \mathcal{D}$. Por tanto, $\mathcal{B} \subset \sigma(\mathcal{D}) = \mathcal{D}$, luego $\mathcal{D} = \mathcal{B}$ y queda demostrado que X es variable aleatoria. □

Proposición 12. *Sea $\varphi\colon \mathbb{R} \to \mathbb{R}$ una función de Borel y sea $X : \Omega \to \mathbb{R}$ una variable aleatoria. Entonces la composición $\varphi \circ X$ es también una variable aleatoria.*

Demostración. Sea $B \in \mathcal{B}$. Por ser φ medible, se tiene que $\varphi^{-1}(B) \in \mathcal{B}$, y como X es variable aleatoria, entonces $X^{-1}(\varphi^{-1}(B)) = (\varphi \circ X)^{-1}(B)$, luego $\varphi \circ X$ es variable aleatoria. □

Ejemplo. Sea $X : \Omega \to \mathbb{R}$ una variable aleatoria y sea $\varphi\colon \mathbb{R} \to \mathbb{R}$ la función definida por $\varphi(x) = ax + b$, con $a > 0$. Es claro que φ es una función de Borel, luego $Y = \varphi \circ X$ es variable aleatoria. Además, si $y \in \mathbb{R}$,

$$F_Y(y) = P(Y^{-1}(-\infty, y]) = P(X^{-1}(\varphi^{-1}((-\infty, y]))) = P(X^{-1}(-\infty, \tfrac{y-b}{a}]) = F_X(\tfrac{y-b}{a})$$

Ejemplo. Sea X una variable aleatoria absolutamente continua con función de densidad

$$f_X(x) = \begin{cases} e^{-x} & \text{si } x > 0 \\ 0 & \text{en caso contrario} \end{cases}$$

Sea $Y = [X]$ la variable aleatoria dada por la parte entera de X. Es fácil comprobar que la parte entera es una función de Borel, implicando que Y es, efectivamente, una variable aleatoria. Además,

(i) $P(Y = 0) = P(0 \leq X < 1) = \displaystyle\int_0^1 e^{-x}dx = 1 - \frac{1}{e}$;

(ii) $P(Y = 1) = P(1 \leq X < 2) = \displaystyle\int_1^2 e^{-x}dx = \frac{1}{e} - \frac{1}{e^2}$;

(iii) $P(Y = n) = P(n \leq X < n+1) = \displaystyle\int_n^{n+1} e^{-x}dx = \frac{1}{e^n} - \frac{1}{e^{n+1}}$ para cada $n \in \mathbb{N}$.

Proposición 13. *Sean X e Y variables aleatorias en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Entonces*

(i) *$X + Y$ es una variable aleatoria.*

(ii) *λX es una variable aleatoria para todo $\lambda \in \mathbb{R}$.*

(iii) *XY es una variable aleatoria.*

Demostración. Ya se vio en la asignatura *Introducción a la Probabilidad y a la Estadística*. □

Proposición 14. *Sean (Ω, τ) y (Ω', τ') espacios topológicos y sea $f : \Omega \to \Omega'$ una aplicación continua. Entonces f es medible con respecto a $\mathcal{B}(\Omega)$ y $\mathcal{B}(\Omega')$.*

Demostración. Como τ' genera $\mathcal{B}(\Omega')$, basta ver que $f^{-1}(B) \in \mathcal{B}(\Omega)$ para todo $B \in \tau'$. Por ser f continua y B abierto, se tiene que $f^{-1}(B)$ es abierto de τ, luego $f^{-1}(B) \in \sigma(\tau) = \mathcal{B}(\Omega)$. □

Proposición 15. *Sea $(\Omega, \mathcal{A})$ un espacio medible. Dadas n funciones $f_1, f_2, \ldots, f_n : \Omega \to \mathbb{R}$, se define $f = (f_1, f_2, \ldots, f_n)\colon \Omega \to \mathbb{R}^n$. Entonces f es medible (con respecto a $\mathcal{A}$ y $\mathcal{B}_{\mathbb{R}^n}$) si y solo si f_i es medible (con respecto a $\mathcal{A}$ y $\mathcal{B}_{\mathbb{R}}$) para todo $i = 1, 2, \ldots, n$.*

Demostración. Supóngase que f es medible. Para cada $i = 1, 2, \ldots, n$, se considera la proyección

$$\begin{aligned} \pi_i\colon \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x_1, x_2, \ldots, x_n) &\longmapsto x_i \end{aligned}$$

Es claro que π_i es una aplicación continua. Como π_i es medible con respecto a $\mathcal{B}_{\mathbb{R}^n}$ y $\mathcal{B}_{\mathbb{R}}$ (usando la proposición anterior), entonces $\pi_i^{-1}(B) \in \mathcal{B}_{\mathbb{R}^n}$, y como f es medible con respecto a $\mathcal{A}$ y $\mathcal{B}_{\mathbb{R}^n}$ (por hipótesis), entonces $f^{-1}(\pi_i^{-1}(B)) = (\pi_i \circ f)^{-1}(B) = f_i^{-1}(B) \in \mathcal{A}$. Por tanto, f_i^{-1} es medible.

Supóngase ahora que f_i es medible para cada $i = 1, 2, \ldots, n$. Basta ver que $f^{-1}((-\infty, b]) \in \mathcal{A}$ para cada $b = (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$, siendo

$$(-\infty, b] = (-\infty, b_1] \times (-\infty, b_2] \times \ldots \times (-\infty, b_n]$$

En efecto, es fácil probar que

$$f^{-1}((-\infty, b]) = \bigcap_{i=1}^{\infty} f_i^{-1}((-\infty, b_i]),$$

de donde se deduce que $f^{-1}((-\infty, b]) \in \mathcal{A}$ por ser intersección numerable de elementos de $\mathcal{A}$. □

3.2. Variables aleatorias simples

Definición 24. Sea X una variable aleatoria en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Se dice que X es **simple** si toma un número finito de valores, en cuyo caso se puede escribir

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

donde los α_i son los valores que toma X, $\mathbb{1}$ es la función indicadora y $A_i = X^{-1}(\{\alpha_i\})$ para cada $i = 1, \ldots, n$. Dicha representación es denominada **representación canónica de X**.

Ejemplo. Considérese el experimento consistente en el lanzamiento de 3 monedas, proporcionando el espacio muestral siguiente:

$$\Omega = \{CCC, CCF, CFC, FCC, CFF, FCF, FFC, FFF\}$$

Sea X la variable aleatoria que toma como valores el número de caras en cada lanzamiento y defínanse los conjuntos

(i) $A_1 = X^{-1}(\{0\}) = \{CCC\}$.

(ii) $A_2 = X^{-1}(\{1\}) = \{CCF, CFC, FCC\}$.

(iii) $A_3 = X^{-1}(\{2\}) = \{CFF, FCF, FFC\}$.

(iv) $A_4 = X^{-1}(\{3\}) = \{FFF\}$.

La representación canónica de X es

$$X(\omega) = 0 \cdot \mathbb{1}_{A_1}(\omega) + 1 \cdot \mathbb{1}_{A_2}(\omega) + 2 \cdot \mathbb{1}_{A_3}(\omega) + 3 \cdot \mathbb{1}_{A_4}(\omega)$$

Definición 25. Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, una **partición de Ω** no es más que una familia numerable $\mathcal{D} = \{D_i\}_{i \in I}$ de elementos de $\mathcal{A}$ tales que $D_i \cap D_j = \emptyset$ para todo $i \neq j$ y

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

Proposición 16. *Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sea $\mathcal{D} = \{D_i\}_{i=1}^{\infty}$ una partición de Ω. Entonces $X : \Omega \to \mathbb{R}$ es una variable aleatoria con respecto a $\sigma(\mathcal{D})$ si y solo si es constante en cada D_i.*

Demostración. Supongamos que X es variable aleatoria. Sea $n \in \mathbb{N}$ y veamos que X es constante en D_n. Sea $\omega_n \in D_n$ y sea $x_n = X(\omega_n)$. Como $\{x_n\}$ es un cerrado de $\mathbb{R}$, entonces $\{x_n\} \in \mathcal{B}$, y como X es variable aleatoria, entonces $X^{-1}(\{x_n\}) \in \sigma(\mathcal{D})$. Se sabe que

$$\sigma(\mathcal{D}) = \left\{ \bigcup_{i \in I} D_i : I \subset \mathbb{N} \right\}$$

Por tanto, existe $I \subset \mathbb{N}$ tal que
$$X^{-1}(\{x_n\}) = \bigcup_{i\in I} D_i$$
Como se tiene que $\omega_n \in X^{-1}(\{x_n\})$ y D_n es el único elemento de la partición que contiene a ω_n, entonces
$$D_n \subset \bigcup_{i\in I} D_i = X^{-1}(\{x_n\}),$$
luego para todo $\omega \in D_n$ se verifica $X(\omega) = X(\omega_n) = x_n$.

Suponemos ahora que X es constante x_i en cada D_i y veamos que $X^{-1}(B) \in \sigma(\mathcal{D})$ para cada $B \in \mathcal{B}$. Sea $B \in \mathcal{B}$. Entonces
$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} = \bigcup_{i=1}^{\infty} \{w \in D_i : X(\omega) \in B\}$$
Ahora bien,
$$\{w \in D_i : X(\omega) \in B\} = \begin{cases} D_i & \text{si } x_i \in B \\ \emptyset & \text{si } x_i \notin B \end{cases}$$
Se concluye por tanto que $X^{-1}(B)$ es unión de elementos de $\sigma(\mathcal{D})$. $\square$

Obsérvese que X es constante en cada D_i si y solo si existe una sucesión de números reales $\{\alpha_i\}_{i=1}^{\infty}$ de modo que
$$X(\omega) = \sum_{i=1}^{\infty} \alpha_i \mathbb{1}_{D_i}(\omega)$$
para cada $\omega \in \Omega$. Aunque no lo parezca, este simple resultado será de cierta utilidad en el futuro.

3.3. Cambio de variable

Dada una función de Borel $\varphi : \mathbb{R} \to \mathbb{R}$ y una variable aleatoria $X : \Omega \to \mathbb{R}$, es bien sabido que $Y = \varphi \circ X$ es también una variable aleatoria. Los siguientes resultados, cuyas demostraciones no corresponden a esta asignatura, serán útiles en ciertos casos en los que X sea absolutamente continua y se quieran hallar las funciones P_Y o F_Y.

> **Teorema 2 (Teorema del cambio de variable).** *Considérese una variable aleatoria absolutamente continua X con función de densidad f_X y sea $\varphi : I \to \mathbb{R}$ una función de Borel, estrictamente monótona y con derivada continua. Entonces $Y = \varphi \circ X$ es una variable aleatoria absolutamente continua cuya función de densidad viene dada por*
> $$f_Y(y) = f_X(\varphi^{-1}(y))|(\varphi^{-1})'(y)|$$

> **Teorema 3 (Teorema del cambio de variable generalizado).** *Considérese una variable aleatoria absolutamente continua X con función de densidad f_X. Sea $\{D_i\}_{i=1}^{N}$ una partición de un intervalo I y sea $\varphi : I \to \mathbb{R}$ una función de Borel tal que $\varphi|_{D_i}$ es estrictamente monótona y con derivada continua. Entonces $Y = \varphi \circ X$ es una variable aleatoria absolutamente continua cuya función de densidad viene dada por*
> $$f_Y(y) = \sum_{i=1}^{N} f_X(\varphi_i^{-1}(y))|(\varphi_i^{-1})'(y)|$$

Distribuciones de probabilidad usuales

4.1. Distribuciones de probabilidad discretas

Definición 26. Sea X una variable aleatoria discreta en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.

(i) Dado $x_0 \in \mathbb{R}$, se dice que X sigue una **distribución degenerada en el punto x_0**, y se denota $X \sim D(x_0)$, cuando

$$P(X=x) = \begin{cases} 1 & \text{si } x = x_0 \\ 0 & \text{si } x \neq x_0 \end{cases}$$

para cada $x \in \mathbb{R}$.

(ii) Dados n puntos $\{x_1, \ldots, x_n\}$ distintos, se dice que X sigue una **distribución uniforme sobre los n puntos**, y se denota $X \sim U(\{x_1, \ldots, x_n\})$, cuando

$$P(X=x) = \begin{cases} \dfrac{1}{n} & \text{si } x \in \{x_1, \ldots, x_n\} \\ 0 & \text{si } x \in \{x_1, \ldots, x_n\}^c \end{cases}$$

para cada $x \in \mathbb{R}$.

(iii) Dado $p \in (0,1)$, se dice que X sigue una **distribución de Bernoulli de parámetro p**, y se denota $X \sim Ber(p)$, cuando

$$P(X=0) = 1-p, \quad P(X=1) = p \quad \text{y} \quad P(X=x) = 0 \ \text{ en el resto}$$

Usualmente se escribe $q = 1-p$.

(iv) Dado $n \in \mathbb{N}$ y dado $p \in (0,1)$, se dice que X sigue una **distribución binomial de parámetros n y p**, y se denota $X \sim Bin(n,p)$, cuando

$$P(X=k) = \binom{n}{k} p^k q^{n-k} \ \text{ para cada } k \in \mathbb{N} \text{ con } k \leq n, \quad \text{y} \quad P(X=x) = 0 \ \text{ en el resto}$$

(v) Dado $p \in (0,1)$, se dice que X sigue una **distribución geométrica de parámetro p**, y se denota $X \sim Geo(p)$, cuando

$$P(X=n) = q^n p \ \text{ para cada } n \in \mathbb{N} \cup \{0\}, \quad \text{y} \quad P(X=x) = 0 \ \text{ en el resto}$$

(vi) Dado $\lambda > 0$, se dice que X sigue una **distribución de Poisson de parámetro λ**, y se denota $X \sim P(\lambda)$, cuando

$$P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!} \ \text{ para cada } k \in \mathbb{N} \cup \{0\}, \quad \text{y} \quad P(X=x) = 0 \ \text{ en el resto}$$

Ejemplo. Dada una variable aleatoria discreta, a cada distribución de probabilidad se le asocia de manera más o menos natural un experimento aleatorio con diversas características:

(*i*) Una distribución degenerada está asociada a un experimento aleatorio con un solo resultado, como tirar un dado de una cara.

(*ii*) Una distribución uniforme está asociada a un experimento aleatorio con varios resultados de igual probabilidad, como tirar un dado de seis caras.

(*iii*) Una distribución de Bernoulli de parámetro $p \in (0,1)$ está asociada a un experimento aleatorio con dos posibles resultados. Los valores que toma la variable aleatoria son 0 (*fracaso*, de probabilidad $q = 1 - p$) o 1 (*éxito*, de probabilidad p).

(*iv*) Una distribución binomial de parámetros $n \in \mathbb{N}$ y $p \in (0,1)$ está asociada a un experimento que consiste en realizar n pruebas de Bernoulli, todas con probabilidad de éxito p. Los valores que toma la variable aleatoria son el número de éxitos (o de fracasos).

(*v*) Una distribución geométrica de parámetro $p \in (0,1)$ está asociada a un experimento aleatorio que consiste en realizar un número indefinido de pruebas de Bernoulli hasta que aparezca el primer éxito, de probabilidad p. Los valores que toma la variable aleatoria son el número de fracasos antes de conseguir el primer éxito.

(*vi*) Una distribución de Poisson de parámetro $\lambda > 0$ está asociada a un experimento aleatorio que consiste en contar el número de ocurrencias de un fenómeno aleatorio, de frecuencia de ocurrencia media λ, durante un periodo de tiempo fijo o en una región fija del espacio. Los valores que toma la variable aleatoria son el número de ocurrencias por unidad de tiempo o de espacio.

4.2. Distribuciones de probabilidad absolutamente continuas

Definición 27. Sea X una variable aleatoria absolutamente continua con función de densidad f en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.

(*i*) Dado un intervalo $[a,b] \subset \mathbb{R}$, se dice que X sigue una **distribución uniforme en $[a,b]$**, y se denota $X \sim U([a,b])$, cuando

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in [a,b] \\ 0 & \text{en otro caso} \end{cases}$$

(*ii*) Dados $\mu \in \mathbb{R}$ y $\sigma^2 > 0$, se dice que X sigue una **distribución normal de parámetros μ y σ^2**, y se denota $X \sim N(\mu, \sigma^2)$, cuando

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{para cada } x \in \mathbb{R}$$

Cuando $\mu = 0$ y $\sigma^2 = 1$, se habla de **distribución normal estándar**.

(*iii*) Dado $\lambda > 0$, se dice que X sigue una **distribución exponencial de parámetro λ**, y se denota $X \sim Exp(\lambda)$, cuando

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

Probabilidad condicionada

5.1. Probabilidad condicionada de sucesos

Definición 28. Considérese un espacio de probabilidad $(\Omega,\mathcal{A},P)$. Dados $A,B\in\mathcal{A}$ con $P(B)>0$, la **probabilidad condicionada del suceso A con respecto al suceso B** se define como

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

Proposición 17. *Siempre que las probabilidades condicionadas tengan sentido, se verifica:*

(i) $P(A\,|\,A)=1$.

(ii) $P(\emptyset\,|\,A)=0$.

(iii) *Si* $B\subset A$, *entonces* $P(A\,|\,B)=1$.

(iv) *Si* A_1 *y* A_2 *son disjuntos, entonces* $P(A_1\cup A_2\,|\,B)=P(A_1|B)+P(A_2\,|\,B)$.

(v) $P(A\,|\,B)+P(A^c\,|\,B)=1$.

Demostración. Bastante inmediata. □

Proposición 18. *Sea* $(\Omega,\mathcal{A},P)$ *un espacio de probabilidad y sea* $B\in\mathcal{A}$ *con* $P(B)>0$. *Entonces* $(\Omega,\mathcal{A},P(\cdot|B))$ *es un espacio de probabilidad, donde*

$$\begin{aligned} P(\cdot|B)\colon \mathcal{A} &\longrightarrow [0,1] \\ A &\longmapsto P(A\,|\,B) \end{aligned}$$

Demostración. Es una simple comprobación. □

Teorema 4 (Teorema de la probabilidad total). *Sea* $(\Omega,\mathcal{A},P)$ *un espacio de probabilidad y sea* $\mathcal{D}=\{D_i\}_{i\in I}$ *una partición de* Ω *tal que* $P(D_i)>0$ *con* $i\in I$. *Entonces, dado* $A\in\mathcal{A}$,

$$P(A)=\sum_{i\in I}P(A\,|\,D_i)P(D_i)$$

Demostración. Ya se vio en la asignatura *Introducción a la Probabilidad y a la Estadística*. □

Teorema 5 (Teorema de Bayes). *Considérese un espacio de probabilidad* $(\Omega,\mathcal{A},P)$ *y sea* $\mathcal{D}=\{D_i\}_{i\in I}$ *una partición de* Ω *tal que* $P(D_i)>0$ *para cada* $i\in I$. *Dado* $A\in\mathcal{A}$ *con* $P(A)>0$, *se tiene*

$$P(D_i\,|\,A)=\frac{P(A\,|\,D_i)P(D_i)}{\sum_{j\in I}P(A\,|\,D_j)P(D_j)}$$

para todo $i\in I$.

Demostración. Ya se vio en la asignatura *Introducción a la Probabilidad y a la Estadística*. □

5.2. Dependencia e independencia de sucesos

Definición 29. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad.

(i) Dos sucesos $A, B \in \mathcal{A}$ son **independientes** cuando

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

(ii) Los sucesos de una familia $\{A_1, A_2, \ldots, A_n\} \subset \mathcal{A}$ son **mutuamente independientes** si para toda subcolección finita $\{A_{i_1}, A_{i_2}, \ldots, A_{i_k}\}$, con $k \leq n$, se tiene que

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

Por ejemplo, para el caso de tres sucesos A_1, A_2 y A_3, la independencia mutua requerirá que se verifique

$$P(A \cap B) = P(A)P(B), \quad P(A \cap C) = P(A)P(C), \quad P(B \cap C) = P(B)P(C)$$

$$\text{y} \quad P(A \cap B \cap C) = P(A)P(B)P(C)$$

Evidentemente, la independencia mutua de sucesos implica la independencia a pares, pero el recíproco no es cierto.

Ejemplo. Se considera el espacio de probabilidad $(\Omega, \mathcal{P}(\Omega), P)$, donde $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ y $P(\omega_i) = \frac{1}{4}$ para todo $i = 1, 2, 3, 4$. Considérense los sucesos $A = \{\omega_1, \omega_2\}$, $B = \{\omega_1, \omega_3\}$ y $C = \{\omega_1, \omega_4\}$. Se tiene que

(i) $P(A) = \dfrac{1}{2}$

(ii) $P(B) = \dfrac{1}{2}$

(iii) $P(C) = \dfrac{1}{2}$

(iv) $P(A \cap B) = \dfrac{1}{4} = P(A)P(B)$

(v) $P(A \cap C) = \dfrac{1}{4} = P(A)P(C)$

(vi) $P(B \cap C) = \dfrac{1}{4} = P(B)P(C)$

Por tanto, los sucesos son independientes dos a dos. Sin embargo,

$$P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C),$$

luego los sucesos no son mutuamente independientes.

Ejemplo. Se considera el espacio $(\Omega, \mathcal{P}(\Omega), P)$, con $\Omega = \{(i,j) \in \mathbb{N}^2 : i, j = 1, 2, 3, 4, 5, 6\}$ y $P(i,j) = \frac{1}{36}$ para todo $(i,j) \in \Omega$. Sean

$$A = \{(i,j) \in \mathbb{N}^2 : j = 1, 2, 5\}, \quad B = \{(i,j) \in \mathbb{N}^2 : j = 4, 5, 6\} \quad \text{y} \quad C = \{(i,j) \in \mathbb{N}^2 : i + j = 9\}$$

Se tiene que

(i) $P(A) = \dfrac{1}{2}$

(ii) $P(B) = \dfrac{1}{2}$

(iii) $P(C) = \dfrac{1}{9}$

(iv) $P(A \cap B) = \dfrac{1}{6} \neq \dfrac{1}{4} = P(A)P(B)$

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

(vi) $P(B \cap C) = \dfrac{1}{12} \neq \dfrac{1}{18} = P(B)P(C)$

Por tanto, los sucesos no son independientes dos a dos. Sin embargo,

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

Aun así, los sucesos no son mutuamente independientes.

5.3. Probabilidad condicionada con respecto a una partición

Definición 30. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad finito y sea $\mathcal{D} = \{D_i\}_{i=1}^n$ una partición de Ω con $P(D_i) > 0$ para todo $i = 1, \dots, n$. Dado $A \in \mathcal{A}$, la **probabilidad condicionada de A con respecto a la partición $\mathcal{D}$** es la función $P(A \mid \mathcal{D})\colon \Omega \to \mathbb{R}$ definida por

$$P(A \mid \mathcal{D})(\omega) = \sum_{i=1}^{n} P(A \mid D_i) \mathbb{1}_{D_i}(\omega)$$

Proposición 19. *Se verifican las siguientes propiedades:*

(i) *$P(A \mid \mathcal{D})$ es una variable aleatoria.*

(ii) *Si A y B son disjuntos, entonces $P(A \cup B \mid \mathcal{D}) = P(A \mid \mathcal{D}) + P(B \mid \mathcal{D})$.*

Demostración. Bastante inmediata. □

5.4. Probabilidad condicionada con respecto a una variable aleatoria

Definición 31. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad finito y sea $X\colon \Omega \to \mathbb{R}$ una variable aleatoria simple con representación canónica

$$X(\omega) = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{D_i}(\omega),$$

siendo $\mathcal{D}_X = \{D_1, \dots, D_n\}$ la partición de Ω dada por $D_i = X^{-1}(\{\alpha_i\})$ para cada $i = 1, \dots, n$. Dado $A \in \mathcal{A}$, la **probabilidad condicionada de A con respecto a la variable aleatoria X** no es más que la probabilidad condicionada de A con respecto a la partición $\mathcal{D}_X$, esto es,

$$P(A \mid X) := P(A \mid \mathcal{D}_X)$$

Ejemplo. Sean X e Y dos variables aleatorias tomando los valores $\{0, 1\}$ y sea $p \in (0, 1)$ con

$$P(X = 0) = 1 - p, \quad P(X = 1) = p, \quad P(Y = 0) = 1 - p \quad \text{y} \quad P(Y = 1) = p$$

Se considera la variable aleatoria $X + Y$, que toma los valores $\{0, 1, 2\}$. Para cada $k \in \mathbb{N}$, se tratará de calcular la variable aleatoria $P((X + Y)^{-1}(\{k\}) \mid Y) \equiv P(X + Y = k \mid Y)$. Se define la partición

$$\mathcal{D}_Y = \{Y^{-1}(\{0\}), Y^{-1}(\{1\})\} = \{D_0, D_1\}$$

Que Y tome los valores 0 y 1 ha provocado la tentación de indexar los elementos de $\mathcal{D}_Y$ según estos valores y no según el orden, como se hace habitualmente. Se tiene que

$$\begin{aligned}
P(X + Y = k \mid D_0) \equiv P(X + Y = k \mid Y = 0) &= \frac{P(X + Y = k \cap Y = 0)}{P(Y = 0)} \\
&= \frac{P(X + 0 = k \cap Y = 0)}{P(Y = 0)} = \frac{P(X = k) P(Y = 0)}{P(Y = 0)} = P(X = k)
\end{aligned}$$

Se ha utilizado de manera crucial que los sucesos $X + 0 = k$ e $Y = 0$ son independientes. De forma análoga se llega a $P(X + Y = k \mid D_1) = P(X = k - 1)$, luego

$$\begin{aligned}
P(X + Y = k \mid Y)(\omega) &= P(X + Y = k \mid D_0) \mathbb{1}_{D_0}(\omega) + P(X + Y = k \mid D_1) \mathbb{1}_{D_1}(\omega) \\
&= P(X = k) \mathbb{1}_{D_0}(\omega) + P(X = k - 1) \mathbb{1}_{D_1}(\omega) \\
&= \begin{cases} (1 - p) \mathbb{1}_{D_0}(\omega) & \text{si } k = 0 \\ p \mathbb{1}_{D_0}(\omega) + (1 - p) \mathbb{1}_{D_1}(\omega) & \text{si } k = 1 \\ p \mathbb{1}_{D_1}(\omega) & \text{si } k = 2 \end{cases}
\end{aligned}$$

¿Son conocidas las variables aleatorias $\mathbb{1}_{D_0}$ y $\mathbb{1}_{D_1}$? En primer lugar, en cuanto a $\mathbb{1}_{D_0}$, se tiene *(i)* $\mathbb{1}_{D_0}(\omega)=0 \iff Y(\omega)\notin D_0 \iff Y(\omega)\in D_1 \iff Y(\omega)=1.$

(ii) $\mathbb{1}_{D_0}(\omega)=1 \iff Y(\omega)\in D_0 \iff Y(\omega)=0.$

En cuanto a $\mathbb{1}_{D_1}$,

(i) $\mathbb{1}_{D_1}(\omega)=0 \iff Y(\omega)\notin D_1 \iff Y(\omega)\in D_0 \iff Y(\omega)=0.$

(ii) $\mathbb{1}_{D_1}(\omega)=1 \iff Y(\omega)\in D_1 \iff Y(\omega)=1.$

Se observa que $\mathbb{1}_{D_0}=1-Y$ y $\mathbb{1}_{D_1}=Y$, así que

$$P(X+Y=k\,|\,Y)=\begin{cases}(1-p)(1-Y)(\omega) & \text{si } k=0\\ p(1-Y)(\omega)+(1-p)Y(\omega) & \text{si } k=1\\ pY(\omega) & \text{si } k=2\end{cases}$$

5.5. Variables aleatorias truncadas

Definición 32. Sea X una variable aleatoria en un espacio de probabilidad $(\Omega,\mathcal{A},P)$ y sea $A\in\mathcal{B}$ tal que $P(X\in A)>0$. La **variable aleatoria truncada de X por A** no es más que la variable aleatoria $Y=X|_{X^{-1}(A)}$, es decir, la función $Y: X^{-1}(A)\to\mathbb{R}$ dada por $Y(\omega)=X(\omega)$. En este contexto, de $P_X(A)=P(X\in A)$ se dirá que es el **grado de truncamiento**.

Es importante remarcar que $Y=X|_{X^{-1}(A)}$ es una variable aleatoria en el espacio de probabilidad $(X^{-1}(A),\mathcal{A},\tilde{P})$, donde

$$\begin{aligned}\tilde{P}:\mathcal{A}&\longrightarrow[0,1]\\ T&\longmapsto\frac{P(T)}{P(X\in A)}\end{aligned}$$

El motivo por el que la probabilidad de este espacio no puede ser P es que, generalmente, $P(X^{-1}(A))$ no es igual a 1, de ahí que haya que «normalizar» la probabilidad para que se tenga $\tilde{P}(X^{-1}(A))=1$. Por tanto, si $B\in\mathcal{B}$, la medida de probabilidad inducida por Y (que se va a denotar por P_Y en lugar de $\tilde{P}_Y$) viene dada por

$$\begin{aligned}P_Y(B)&=\tilde{P}(Y^{-1}(B))=\tilde{P}(\{\omega\in X^{-1}(A): X(\omega)\in B\})=\tilde{P}(X^{-1}(B)\cap X^{-1}(A))\\ &=\frac{P(X^{-1}(B)\cap X^{-1}(A))}{P(X^{-1}(A))}=\frac{P_X(B\cap A)}{P_X(A)}=P_X(B\,|\,A)\end{aligned}$$

Ejemplo. Sea X una variable aleatoria discreta tomando valores en $\mathbb{N}\cup\{0\}$, y considérese el boreliano $A=[2,\infty)\subset\mathbb{R}$. Entonces la variable aleatoria truncada Y de X por A toma los valores $\{2,3,4,...\}$, y además

$$P_Y(\{0\})=P(Y=0)=\frac{P((X=0)\cap(X\geq 2))}{P(X\geq 2)}=0$$

Ejemplo. Sea X una variable aleatoria discreta tal que $X\sim Ge(p)$ (se recuerda que en este caso X toma valores en $\mathbb{N}\cup\{0\}$ y $P(X=n)=q^np$), y sea $A=[0,a]$ para cualquier $a\in\mathbb{N}$. Entonces la variable aleatoria truncada Y de X por A toma los valores $\{0,1,\dots,a\}$, y además, para cada $y\in\{0,1,\dots,a\}$,

$$P_Y(\{y\})=P(Y=y)=\frac{P((X=y)\cap(X\leq a))}{P(X\leq a)}=\frac{P(X=y)}{P(X\leq a)}=\frac{q^yp}{\sum_{i=0}^a q^ip}$$

Esperanza

6.1. Nociones básicas

Definición 33. Sea $X : \Omega \to \mathbb{R}$ una variable aleatoria. Se define la **esperanza de X** como

$$E[X] := \int_\Omega X(\omega) dP(\omega),$$

siempre que esta integral exista.

La integral anterior es una integral en el sentido de Lebesgue, objeto de estudio de la asignatura *Teoría de la Medida e Integración*. Es por ello que algunas de las demostraciones de esta sección van a omitirse por completo. Por otro lado, es habitual encontrarse las notaciones

$$E[X] = \int_\Omega X(\omega) dP(\omega) \equiv \int_\Omega X(\omega) P(d\omega) \equiv \int_\Omega X dP$$

Proposición 20. *Dada una variable aleatoria no negativa X, la esperanza de X existe si y solo si la integral*

$$\int_0^\infty (1 - F_X(x)) dx = \int_0^\infty P(X > x) dx$$

es finita. Más generalmente, si X es una variable aleatoria arbitraria, entonces $E[X]$ existe si y solo si las integrales

$$\int_0^\infty (1 - F_X(x)) dx \quad \text{y} \quad \int_{-\infty}^0 F_X(x) dx$$

son finitas, en cuyo caso se tiene

$$E[X] = \int_0^\infty (1 - F_X(x)) dx - \int_{-\infty}^0 F_X(x) dx$$

Por motivos de comodidad, si no se hace mención alguna sobre la existencia de cierta esperanza es que se ha supuesto de antemano que dicha esperanza existe.

Proposición 21. *Se verifican las siguientes propiedades:*

(i) *Si X e Y son variables aleatorias, entonces $E[X + Y] = E[X] + E[Y]$.*

(ii) *Si $a, b \in \mathbb{R}$ y X es una variable aleatoria, entonces $E[aX + b] = aE[X] + b$.*

(iii) *Si $X, Y : \Omega \to \mathbb{R}$ son variables aleatorias con $X(\omega) \leq Y(\omega)$ para todo $\omega \in \Omega$, entonces $E[X] \leq E[Y]$.*

(iv) *Si $g : \mathbb{R} \to \mathbb{R}$ es medible-Borel y X es una variable aleatoria, entonces*

$$E[g \circ X] = \int_\mathbb{R} g(x) dP_X(x)$$

En particular,

$$E[X] = \int_{\mathbb{R}} x dP_X(x)$$

(*v*) *Si X es una variable aleatoria discreta, entonces*

$$E[X] = \sum_{x \in D_X} x P_X(\{x\}) = \sum_{x \in D_X} x P(X = x)$$

(*vi*) *Si X es una variable aleatoria absolutamente continua con función de densidad f, entonces*

$$E[X] = \int_{\mathbb{R}} x f(x) dx$$

(*vii*) *Si X es una variable aleatoria mixta, entonces*

$$E[X] = \sum_{x \in D_X} x P(X = x) + \int_{\mathbb{R}} x \tilde{f}(x) dx,$$

donde $\tilde{f}$ es la pseudodensidad de la parte continua de la función de distribución de X.

(*viii*) *Si X e Y son variables aleatorias independientes, entonces $E[XY] = E[X]E[Y]$. Si además $\varphi : \mathbb{R} \to \mathbb{R}$ es medible-Borel, entonces $E[\varphi(X)\varphi(Y)] = E[\varphi(X)]E[\varphi(Y)]$.*

6.2. Momentos de una variable aleatoria

Definición 34. Sea X una variable aleatoria y sea $n \in N$.

(*i*) Se define el **momento de orden *n* de *X*** como la esperanza de la variable aleatoria X^n.

(*ii*) Se define el **momento centrado de orden *n* de *X*** como el momento de orden n de $X - E[X]$.

(*iii*) Se define la **varianza de *X*** como su momento centrado de orden 2, es decir,

$$\mathrm{Var}[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2$$

(*iv*) Si Y es otra variable aleatoria, se define la **covarianza de *X* e *Y*** como

$$\mathrm{Cov}[X, Y] := E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

Cuando $\mathrm{Cov}[X, Y] = 0$ se dice que X e Y son **incorreladas**.

De la definición anterior se deduce inmediatamente que $\mathrm{Cov}[X, X] = 0$ y que si X e Y son variables aleatorias independientes entonces son incorreladas.

Ejemplo. Sea X una variable aleatoria discreta con $X \sim Deg(a)$ para cierto $a \in \mathbb{R}$. Entonces

$$E[X] = a \cdot P(X = a) = a, \qquad \mathrm{Var}[X] = E[X^2] - E[X]^2 = 0$$

Ejemplo. Sea X una variable aleatoria discreta con $X \sim Ber(p)$, donde $p \in (0, 1)$. Entonces

$$E[X] = 0(1 - p) + 1p = p, \qquad \mathrm{Var}[X] = p - p^2 = p(1 - p)$$

Ejemplo. Sea X una variable aleatoria discreta con $X \sim Bin(n, p)$, donde $n \in \mathbb{N}$, $p \in (0, 1)$. Como X representa el número de éxitos en n ensayos de Bernoulli con probabilidad p, se tiene $X = X_1 + \ldots + X_n$, donde para cada $i = 1, \ldots, n$, X_i toma los valores 0 o 1 dependiendo de si en el i-ésimo ensayo hay un fracaso o un éxito. Por tanto, $X_i \sim Ber(p)$, luego

$$E[X] = E[X_1 + \ldots + X_n] = E[X_1] + \ldots + E[X_n] = p + \ldots + p = np$$

Para la varianza, se observa que $E[X^2] = E[X(X-1)] + E[X]$, donde

$$
\begin{aligned}
E[X(X-1)] &= \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k q^{n-k} = \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k q^{n-k} \\
&= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k q^{n-k} = n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} q^{n-k} \\
&= n(n-1)p^2 \sum_{k=0}^{n-2} \binom{n-2}{k} p^k q^{n-k-2} \overset{(*)}{=} n(n-1)p^2 (p+q)^{n-2} = n(n-1)p^2
\end{aligned}
$$

Por tanto,

$$
\mathrm{Var}[X] = n(n-1)p^2 + np - n^2p^2 = npq
$$

Por si la igualdad $(*)$ causase incertidumbre, se recuerda que para todos $x, y \in \mathbb{R}$ y $n \in \mathbb{N}$ se tiene

$$
\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = (x+y)^n
$$

Ejemplo. Sea X una variable aleatoria absolutamente continua con $X \sim N(\mu, \sigma^2)$ para ciertos $\mu \in \mathbb{R}$, $\sigma^2 > 0$. Se tiene que

$$
\begin{aligned}
E[X] &= \int_{\mathbb{R}} x f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \overset{(*)}{=} \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} (\sqrt{2}\sigma t + \mu) e^{-t^2} dt \\
&= \frac{1}{\sqrt{\pi}} \left(\sqrt{2}\sigma \int_{\mathbb{R}} t e^{-t^2} dt + \mu \int_{\mathbb{R}} e^{-t^2} dt \right) = \frac{1}{\sqrt{\pi}} \left(\sqrt{2}\sigma \left[-\frac{1}{2} e^{-t^2} \right]_{-\infty}^{+\infty} + \mu\sqrt{\pi} \right) = \frac{\mu\sqrt{\pi}}{\sqrt{\pi}} = \mu,
\end{aligned}
$$

donde en $(*)$ se ha hecho $t = \frac{x-\mu}{\sqrt{2}\sigma}; dt = \frac{1}{\sqrt{2}\sigma} dx$. Cálculos similares demuestran que $\mathrm{Var}[X] = \sigma^2$.

Ejemplo. Sea X una variable aleatoria absolutamente continua con $X \sim Exp(\lambda)$ para cierto $\lambda > 0$. Se tiene que

$$
E[X] = \int_{\mathbb{R}} x f(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx
$$

Se halla primero la primitiva:

$$
\int \lambda x e^{-\lambda x} dx \overset{(*)}{=} -x e^{-\lambda x} - \int -e^{-\lambda x} dx = -x e^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x},
$$

donde en $(*)$ se ha integrado por partes:

$$
(*) \left\{ \begin{array}{ll} u(x) = \lambda x; & u'(x) = \lambda \\ v(x) = -\frac{1}{\lambda} e^{-\lambda x}; & v'(x) = e^{-\lambda x} \end{array} \right\}
$$

Por tanto,

$$
E[X] = \left[-x e^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} \right]_0^{\infty} = \frac{1}{\lambda}
$$

Cálculos similares demuestran que

$$
\mathrm{Var}[X] = \frac{1}{\lambda^2}
$$

Ejemplo. Sea X una variable aleatoria discreta con $X \sim P(\lambda)$ para cierto $\lambda > 0$. Se tiene que

$$
E[X] = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \lambda e^{\lambda} = \lambda
$$

Por otro lado,

$$E[X(X-1)] = \sum_{k=0}^{n} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^k}{(k-2)!} = \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}$$
$$= \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2$$

Por tanto,

$$\operatorname{Var}[X] = E[X(X-1)] + E[X] - E[X]^2 = \lambda$$

Proposición 22. *Se verifican las siguientes propiedades:*

(i) *Si $a, b \in \mathbb{R}$ y X es una variable aleatoria, entonces* $\operatorname{Var}[aX+b] = a^2\operatorname{Var}[X]$.

(ii) *Si $X_1, \ldots, X_n$ son variables aleatorias independientes dos a dos, entonces*

$$\operatorname{Var}[X_1 + \ldots + X_n] = \operatorname{Var}[X_1] + \ldots + \operatorname{Var}[X_n]$$

En consecuencia, si $a_1, \ldots, a_n \in \mathbb{R}$,

$$\operatorname{Var}[a_1X_1 + \ldots + a_nX_n] = a_1^2\operatorname{Var}[X_1] + \ldots + a_n^2\operatorname{Var}[X_n],$$

Demostración. Solo se va a probar el apartado segundo. Se tiene que

$$\begin{aligned}
\operatorname{Var}\Big[\sum_{i=1}^{n} X_i\Big] &= E\Big[\Big(\sum_{i=1}^{n} X_i - E\Big[\sum_{i=1}^{n} X_i\Big]\Big)^2\Big] \\
&= E\Big[\Big(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} E[X_i]\Big)^2\Big] \\
&= E\Big[\Big(\sum_{i=1}^{n}(X_i - E[X_i])\Big)^2\Big] \\
&= E\Big[\sum_{i=1}^{n}(X_i - E[X_i])^2 + 2\sum_{1\le i<j\le n} E[(X_i - E[X_i])(X_j - E[X_j])]\Big] \\
&\overset{(*)}{=} E\Big[\sum_{i=1}^{n}(X_i - E[X_i])^2\Big] \\
&= \sum_{i=1}^{n} E[(X_i - E[X_i])^2] \\
&= \sum_{i=1}^{n} \operatorname{Var}[X_i],
\end{aligned}$$

donde en la igualdad $(*)$ se ha usado que $E[(X_i - E[X_i])(X_j - E[X_j])] = 0$ gracias a la independencia de las variables. □

Proposición 23 (Desigualdad de Markov). *Dada una variable aleatoria no negativa X, para cada $t > 0$ se tiene*

$$P(X \ge t) \le \frac{E[X]}{t}$$

Demostración. Solo va a demostrarse para variables aleatorias discretas y absolutamente continuas. Si X toma los valores x_k con $k \in \mathbb{N}$ y $P(X = x_k) = p_k$, entonces para cada $t > 0$ se tiene

$$E[X] = \sum_{k=1}^{\infty} x_k p_k \ge \sum_{x_k \ge t} x_k p_k \ge \sum_{x_k \ge t} t p_k = t \cdot P(X \ge t),$$

de donde se deduce la desigualdad del enunciado. Si fuese X una variable aleatoria absolutamente continua disfrutando de una función de densidad f, entonces

$$E[X] = \int_{\mathbb{R}} xf(x)dx \geq \int_t^{\infty} xf(x)dx \geq \int_t^{\infty} tf(x)dx = t\int_0^{\infty} f(x)dx = t \cdot P(X \geq t),$$

de donde se deduce la desigualdad del enunciado. □

Corolario 2 (Desigualdad de Chebyshev). *Si X es una variable aleatoria con varianza finita, entonces para todo $\varepsilon > 0$ se verifica*

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{\text{Var}[X]}{\varepsilon^2}$$

Demostración. Basta aplicar la proposición anterior a la variable aleatoria no negativa $(X - E[X])^2$ tomando $t = \varepsilon^2$:

$$P((X - E[X])^2 \geq \varepsilon^2) \leq \frac{E[(X - E[X])^2]}{\varepsilon^2},$$

de donde se deduce inmediatamente la desigualdad del enunciado. □

Como caso particular, tomando $\varepsilon = h\sigma$ para cualquier $h > 0$, con $\sigma = \sqrt{\text{Var}[X]}$, la desigualdad de Chebyshev dice que

$$P(|X - E[X]| \geq h\sigma) \leq \frac{\text{Var}[X]}{h^2\sigma^2} = \frac{1}{h^2} \qquad \text{y} \qquad P(|X - E[X]| < h\sigma) \geq 1 - \frac{1}{h^2},$$

lo que permite acotar la probabilidad de que un valor de una variable aleatoria desconocida se encuentre en un intervalo de centro la esperanza y longitud $2h\sigma$.

6.3. Función generatriz de probabilidad

Definición 35. Sea X una variable aleatoria discreta que toma valores en $\mathbb{N} \cup \{0\}$, y sea $p_n = P(X = n)$ para cada $n \in \mathbb{N} \cup \{0\}$. La **función generatriz de probabilidad de X** es la función dada por

$$G_X(s) = \sum_{n=0}^{\infty} p_n s^n = E[s^X]$$

para cada $s \in \mathbb{R}$ que consiga que la serie anterior converja.

Obsérvese que la función generatriz siempre tiene sentido en el intervalo $(-1, 1)$, pues la serie de los valores absolutos puede acotarse por una serie geométrica de razón $|s| < 1$. Además, en $s = 1$ la serie converge y vale 1.

Ejemplo. Supóngase que $X \sim Bin(n, p)$ para cierto $n \in \mathbb{N}$ y $p \in (0, 1)$. Entonces, para todo $s \in \mathbb{R}$,

$$G_X(s) = \sum_{k=0}^{\infty} p_k s^k = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} s^k = (ps + q)^n$$

Ejemplo. Supóngase que $X \sim Ge(p)$ para cierto $p \in (0, 1)$. Entonces

$$G_X(s) = \sum_{n=0}^{\infty} p_n s^n = \sum_{n=0}^{\infty} q^n p s^n = p \sum_{n=0}^{\infty} (sq)^n,$$

luego $G_X(s)$ existe si y solo si $|sq| < 1$, es decir, si y solo si $|s| < \frac{1}{q}$, en cuyo caso se tiene

$$G_X(s) = \frac{p}{1 - sq}$$

Ejemplo. Supóngase que $X \sim P(\lambda)$ para cierto $\lambda > 0$. Entonces para todo $s \in \mathbb{R}$ se tiene

$$G_X(s) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} s^n = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda s)^n}{n!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}$$

Proposición 24. *Sean $X_1, \ldots, X_n$ variables aleatorias discretas independientes tomando valores en $\mathbb{N} \cup \{0\}$ y con funciones generatrices G_{X_i} para cada $i = 1, \ldots, n$. Entonces*

$$G_{X_1+\ldots+X_n}(s) = G_{X_1}(s) \ldots G_{X_n}(s)$$

para cada $s \in \mathbb{R}$ donde estén definidas todas las G_{X_i}.

Demostración. Es una simple comprobación:

$$G_{X_1+\ldots+X_n}(s) = E[s^{X^1+\ldots+X_n}] = E[s^{X_1} \ldots s^{X_n}] \overset{(*)}{=} E[s^{X_1}] \ldots E[s^{X_n}] = G_{X_1}(s) \ldots G_{X_n}(s)$$

En la igualdad $(*)$ ha sido fundamental que las variables son independientes y que la función dada por $f(t) = s^t$ para cada $t \in \mathbb{R}$ es medible-Borel, pues es continua. $\square$

Teorema 6 (Fórmula de inversión). *Sea X una variable aleatoria discreta con valores en $\mathbb{N} \cup \{0\}$ y función generatriz de probabilidad G_X. Entonces, para cada $k \in \mathbb{N} \cup \{0\}$ se tiene*

$$p_k = \frac{G_X^{(k)}(0)}{k!}$$

Demostración. En primer lugar, por ser G_X una serie de potencias, entonces es de clase infinito en el interior del intervalo de convergencia I, y además, para todo $s \in \mathring{I}$ y $k \in \mathbb{N}$, se verifica

$$G_X^{(k)}(s) = \sum_{n=k}^{\infty} p_n n(n-1)(n-2)\ldots(n-k+1)s^{n-k} = p_k k! + \sum_{n=k+1}^{\infty} p_n n(n-1)(n-2)\ldots(n-k+1)s^{n-k}$$

Evaluando en 0 (obsérvese que $0 \in (-1,1) \subset \mathring{I}$) se obtiene $G_X^{(k)}(0) = p_k k!$, de donde se deduce la fórmula del enunciado. $\square$

La importancia de esta fórmula reside en el hecho de que la distribución de probabilidad de una variable aleatoria queda totalmente determinada por la función generatriz de probabilidad.

Ejemplo. Supóngase que $X \sim Bin(n,p)$ e $Y \sim Bin(m,p)$ son variables independientes, donde $m,n \in \mathbb{N}$ y $p \in (0,1)$. Entonces

$$G_{X+Y}(s) = G_X(s)G_Y(s) = (ps+q)^n(ps+q)^m = (ps+q)^{n+m}$$

Como se ha visto que la función generatriz determina completamente la distribución de $X+Y$, puede afirmarse que $X+Y \sim Bin(n+m,p)$.

Ejemplo. Supóngase que $X \sim P(\lambda)$ e $Y \sim P(\mu)$ son variables independientes, donde $\lambda, \mu > 0$. Entonces $X+Y \sim P(\lambda+\mu)$, ya que

$$G_{X+Y}(s) = G_X(s)G_Y(s) = e^{\lambda(s-1)}e^{\mu(s-1)} = e^{(\lambda+\mu)(s-1)}$$

Proposición 25. *Sea X una variable aleatoria discreta tomanndo valores en $\mathbb{N} \cup \{0\}$ y con función generatriz de probabilidad G_X. Se verifican las siguientes propiedades:*

(i) $E[X] = G'_X(1)$.

(ii) $E[X^2] = G''_X(1) + G'_X(1)$.

(iii) $\mathrm{Var}[X] = G''_X(1) + G'_X(1) - G'_X(1)^2$.

Demostración. Los dos primeros apartados son simples comprobaciones y el tercero es consecuencia inmediata de los anteriores. □

6.4. Función generatriz de momentos

Definición 36. Sea X una variable aleatoria. La **función generatriz de momentos asociada a X** se define como

$$M_X(t) := E[e^{tX}]$$

para cada $t \in \mathbb{R}$ donde la esperanza de e^{tX} exista.

En particular, si X es una variable aleatoria discreta tomando valores en $\mathbb{N} \cup 0$, entonces, recordando que $G_X(s) = E[s^X]$, se tiene

$$M_X(t) = G_X(e^t) = \sum_{n=0}^{\infty} e^{tn} P(X = n)$$

Por otro lado, si X es una variable aleatoria absolutamente continua con función de densidad f, se verifica

$$M_X(t) = \int_{\mathbb{R}} e^{tx} f(x) dx$$

Proposición 26. *Se verifican las siguientes propiedades:*

(i) *Si X e Y son variables aleatorias independientes, entonces*

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

para cada $t \in \mathbb{R}$ donde $M_X(t)$ y $M_Y(t)$ tengan sentido.

(ii) *Si $a, b \in \mathbb{R}$ y X es una variable aleatoria, entonces*

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

para cada $t \in \mathbb{R}$ donde $M_X(at)$ tenga sentido.

(iii) *Si X es una variable aleatoria y $k \in N$,*

$$M_X^{(k)}(0) = E[X^k]$$

En particular,

$$M_X''(0) - \mu^2 = \sigma^2,$$

donde $\mu = E[X]$ y $\sigma^2 = \mathrm{Var}[X]$.

Demostración. La prueba de los dos primeros apartados es inmediata. La del tercero es un poco menos inmediata pero los cálculos son lo suficientemente pesados como para omitir la prueba. □

6.5. Función característica

Definición 37. Dada una variable aleatoria X, la **función característica de X** es la función $\varphi_X : \mathbb{R} \to \mathbb{C}$ dada por

$$\varphi_X(t) = E[e^{itX}]$$

Se recuerda que para todo $\mu = \alpha + i\beta \in \mathbb{C}$ se tiene que $e^{\mu} = e^{\alpha}(\cos(\beta t) + i\,\mathrm{sen}(\beta t))$. Obsérvese además que la integral

$$\int_{\mathbb{R}} e^{itX(\omega)} dP(\omega) = \varphi_X(t)$$

siempre existe, de ahí que se haya tenido la valentía de definir φ_X en todo $\mathbb{R}$.

Proposición 27. *Dada una variable aleatoria X, se verifican las siguientes propiedades:*

(i) $\varphi_X(t) = M_X(it)$ *para todo* $t \in \mathbb{R}$.

(ii) $|\varphi_X(t)| \leq 1$ *para todo* $t \in \mathbb{R}$.

(iii) $\varphi_X(-t) = \overline{\varphi_X(t)}$ *para todo* $t \in \mathbb{R}$.

(iv) $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$ *para todo* $t \in \mathbb{R}$.

(v) *Si* $\varphi_X(t) \in \mathbb{R}$ *para todo* $t \in \mathbb{R}$, *entonces* $\varphi_{-X} = \varphi_X$.

(vi) *Si existe* $E[X^n]$ *para algún* $n \in \mathbb{N}$, *entonces*

$$E[X^r] = \frac{\varphi_X^{(r)}(0)}{i^r}$$

para todo $r \in \mathbb{N}$ *con* $r \leq n$.

Demostración. Comprobaciones fáciles e inmediatas. □

6.6. Otras características numéricas

Además de la esperanza y la varianza, algunas características numéricas de una variable aleatoria X que se han estudiado ya en múltiples ocasiones son las *medidas de posición* (la mediana, la moda o los percentiles), y las *medidas de dispersión* (la desviación típica o el coeficiente de variación).

Definición 38. Sea X una variable aleatoria en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Un número real Me[X] $\in \mathbb{R}$ se dice que es una **mediana de X** cuando

$$P(X \leq \text{Me}[X]) \geq \frac{1}{2} \qquad \text{y} \qquad P(X \geq \text{Me}[X]) \geq \frac{1}{2},$$

o, equivalentemente, cuando

$$\frac{1}{2} \leq F_X(\text{Me}[X]) \leq \frac{1}{2} + P(X = \text{Me}[X])$$

Ejemplo. Sea X una variable aleatoria cuya función de distribución viene dada por

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x - \frac{x^2}{2} & \text{si } 0 \leq x < 1 \\ \frac{1}{2} & \text{si } 1 \leq x < 2 \\ \frac{x}{4} & \text{si } 2 \leq x < 4 \\ 1 & \text{si} 4 \leq x \end{cases}$$

Cualquier punto de $[1,2]$ es una mediana de X.

Definición 39. Sea X una variable aleatoria en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Dado $\alpha \in \mathbb{R}$ y $P_\alpha \in \mathbb{R}$, se dice que P_α es un **percentil de orden α** cuando

$$P(X \leq P_\alpha) \geq \frac{\alpha}{100} \qquad \text{y} \qquad P(X \geq P_\alpha) \geq 1 - \frac{\alpha}{100}$$

Si $\alpha \in \{25, 50, 75\}$, se hablará de **cuartiles**, y si $\alpha = 10$, de **deciles**.

Definición 40. Sea X una variable aleatoria en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.

(i) Si X es discreta, se define la **moda de X** como su valor de mayor probabilidad.

(ii) Si X es absolutamente continua, se define la **moda de X** como el valor máximo de su función de densidad.

Definición 41. Sea X una variable aleatoria en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Se define la **desviación típica de X** como

$$\sigma = \sqrt{\operatorname{Var}[X]},$$

mientras que el **coeficiente de variación de X** no es más que

$$C_V = \frac{\sigma}{E[X]}$$

Vectores aleatorios

7.1. Nociones básicas

En este tema, la σ-álgebra de Borel de $\mathbb{R}^n$ junto con la topología usual será denotada por $\mathcal{B}^n$ en lugar de $\mathcal{B}(\mathbb{R}^n)$.

Definición 42. Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, un **vector aleatorio** es una función $X : \Omega \to \mathbb{R}^n$ medible con respecto a $\mathcal{A}$ y $\mathcal{B}^n$, es decir, tal que $X^{-1}(B) \in \mathcal{A}$ para todo $B \in \mathcal{B}^n$.

Proposición 28. *Dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, una función $X : \Omega \to \mathbb{R}^n$ es un vector aleatorio si y solo si cada compente $X_i : \Omega \to \mathbb{R}$ es una variable aleatoria.*

Demostración. Consecuencia inmediata de la Proposición 15. □

Definición 43. Sea X un vector aleatorio en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. La **medida de probabilidad inducida por X** es la función $P_X : \mathcal{B}^n \to [0,1]$ definida por

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

para cada $B \in \mathcal{B}^n$.

Definición 44. Sea X un vector aleatorio en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. La **función de distribución conjunta asociada a X** es la función $F : \mathbb{R}^n \to [0,1]$ definida por

$$F(a_1, \ldots, a_n) = P(X_1 \leq a_1, \ldots, X_n \leq a_n)$$

para cada $(a_1, \ldots, a_n) \in \mathbb{R}^n$

Respecto a la notación utilizada para referirse a $F(a_1, \ldots, a_n)$, es habitual encontrarse las expresiones

$$P(X \leq a), \quad P_X((-\infty, a_1] \times \ldots \times (-\infty, a_n]) \quad \text{o} \quad P(X_1 \leq a_1 \cap \ldots \cap X_n \leq a_n)$$

Proposición 29. *Sea X un vector aleatorio con función de distribución conjunta F. Entonces*

(i) *Se verifica que*

$$F(\infty, \ldots, \infty) \equiv \lim_{a_1 \to \infty, \ldots, a_n \to \infty} F(a_1, \ldots, a_n) = 1$$

(ii) *Para cada $i \in \{1, \ldots, n\}$ y cada $(a_1, \ldots, a_n) \in \mathbb{R}^n$,*

$$F(a_1, \ldots, a_{i-1}, -\infty, a_{i+1}, \ldots, a_n) \equiv \lim_{a_i \to -\infty} F(a_1, \ldots, a_n) = 0$$

(iii) *F es creciente en cada componente.*

(iv) *F es continua por la derecha en cada componente.*

Demostración. Como viene siendo habitual en esta asignatura, se omite la demostración. □

7.2. Vectores aleatorios bidimensionales

Proposición 30. *Sea (X,Y) un vector aleatorio bidimensional y sea F su función de distribución conjunta. Entonces*

(i) *Se verifica que*

$$F(\infty,\infty) \equiv \lim_{a_1\to\infty, a_2\to\infty} F(a_1,a_2) = 1$$

(ii) *Para cada $(x,y) \in \mathbb{R}^2$,*

$$F(-\infty,y) \equiv \lim_{x\to-\infty} F(x,y) = 0 \qquad y \qquad F(x,-\infty) \equiv \lim_{y\to-\infty} F(x,y) = 0$$

(iii) *Si $y \in \mathbb{R}$,*

$$F(x,y) \le F(x',y) \text{ para todos } x,x' \in \mathbb{R} \text{ con } x < x',$$

y si $x \in \mathbb{R}$,

$$F(x,y) \le F(x,y') \text{ para todos } y,y' \in \mathbb{R} \text{ con } y < y'$$

(iv) *Para cada $(x,y) \in \mathbb{R}$,*

$$\lim_{h\to0^+} F(x+h,y) = F(x,y) \qquad y \qquad \lim_{h\to0^+} F(x,y+h) = F(x,y)$$

(v) *Para cada $(x_1,y_1),(x_2,y_2) \in \mathbb{R}^2$,*

$$P(x_1 < X \le x_2, y_1 < Y \le y_2) = F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1)$$

(vi) *Si X e Y son variables aleatorias independientes, entonces $F(x,y) = F_X(x)F_Y(y)$ para todo $(x,y) \in \mathbb{R}^2$.*

(vii) *$D_{(X,Y)} = D_X \times D_Y$, donde $D_{(X,Y)} = \{(x,y) \in \mathbb{R}^2 : p((x,y)) = P(X=x, Y=y) > 0\}$.*

Demostración. Los cuatro primeros apartados vienen de reescribir la proposición anterior para el caso bidimensional, mientras que los tres últimos se comprueban fácilmente. □

Definición 45. Dado un vector aleatorio (X,Y), la **distribución marginal de X** no es más que la medida de probabilidad inducida por X, mientras que la **distribución marginal de Y** no es más que la medida de probabilidad inducida por Y.

El ejemplo que sigue demostrará que las cuatro propiedades de la Proposición 29 no caracterizan a las funciones de distribuciones conjuntas, esto es, hay funciones que verifican las propiedades mencionadas y que no son la función de distribución conjunta asociada a ningún vector aleatorio.

Ejemplo. Considérese la función $G : \mathbb{R}^2 \to [0,1]$ dada por

$$G(x,y) = \begin{cases} 0 & \text{si } x+y < 1 \\ 1 & \text{si } x+y \ge 1 \end{cases}$$

Si G fuera la función de distribución conjunta asociada a algún vector aleatorio X, se tendría

$$P_X((0,1] \times (0,1]) = G(1,1) - G(0,1) - G(1,0) + G(0,0) = -1,$$

lo cual es imposible.

Así, dada una función F, para asegurar que existe algún vector aleatorio con función de distribución conjunta F será necesario exigir alguna propiedad adicional.

7.3. Diferencia de una función con respecto a un rectángulo

Definición 46. Considérese una función $F: \mathbb{R}^n \to \mathbb{R}$, sea $A = (a,b]$ un rectángulo acotado de $\mathbb{R}^n$ y sea V el conjunto de los 2^n vértices de A. La **diferencia de F con respecto a A** se define como

$$\Delta_A F := \sum_{x \in V} \operatorname{sgn}(x) F(x),$$

donde $\operatorname{sgn}(x) = 1$ si el número de veces que se tiene $x_i = a_i$ es par, y $\operatorname{sgn}(x) = -1$ en caso contrario.

La definición puede parecer algo confusa, pero se trata en realidad de un concepto más simple que el mecanismo de un botijo. Por ejemplo, para $n = 1$, si $A = (a,b]$, la diferencia será

$$\Delta_A F = F(b) - F(a),$$

y para $n = 2$, si $A = (a_1,b_1] \times (a_2,b_2]$, se tendrá

$$\Delta_A F = F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2)$$

La diferencia de una función con respecto a un rectángulo acotado nos dará la condición que faltaba para asegurar que una función con ciertas propiedades es la función de distribución conjunta de algún vector aleatorio.

Proposición 31. *Si una función $F: \mathbb{R}^n \to [0,1]$ con las propiedades* ***(i)****,* ***(ii)****,* ***(iii)*** *y* ***(iv)*** *que aparecen en la Proposición 29 verifica*

(v) $\Delta_A F \geq 0$ *para todo* $A = (a,b] \subset \mathbb{R}^n$,

entonces F es la función de distribución conjunta asociada a algún vector aleatorio X.

Demostración. Requiere más trabajo del que le gustaría al probabilista promedio. □

De aquí en adelante, por motivos de comodidad en la notación, se trabajará exclusivamente con vectores aleatorios bidimensionales.

7.4. Vectores aleatorios discretos

Definición 47. Un vector aleatorio $(X,Y): \Omega \to \mathbb{R}^2$ se dice que es **discreto** si X e Y son variables aleatorias discretas.

Quizá habría sido más natural definir los vectores aleatorios discretos de la misma forma que se hizo en el caso unidimensional, es decir, diciendo que *un vector aleatorio es discreto cuando*

$$\sum_{(x,y) \in \mathbb{R}^2} p((x,y)) = \sum_{(x,y) \in D_{(X,Y)}} p((x,y)) = 1$$

Pues bien, si se tiene en cuenta que $D_{(X,Y)} = D_X \times D_Y$, es sencillo demostrar que ambas definiciones son equivalentes.

Por otra parte, en cuanto a las distribuciones marginales, si $D_X = \{x_i : i \in I\}$ es el rango de X y el de Y es $D_Y = \{y_j : j \in J\}$ (obviamente $I, J \subset \mathbb{N}$), entonces

$$P_X(\{x_i\}) = P(X = x_i) = \sum_{j \in J} P(X = x_i, Y = y_j) = \sum_{j \in J} P_{(X,Y)}(\{(x_i, y_j)\}),$$

lo que proporciona una forma de calcular la distribución marginal de X de un vector aleatorio (X,Y) a partir de la medida de probabilidad de dicho vector. Evidentemente, de forma análoga,

$$P_Y(\{y_j\}) = P(Y = y_j) = \sum_{i \in I} P(X = x_i, Y = y_j) = \sum_{i \in I} P_{(X,Y)}(\{(x_i, y_j)\})$$

Definición 48. Sea (X,Y) un vector aleatorio discreto y sea $y_j \in \mathbb{R}$ tal que $P(Y = y_j) > 0$. Se define la **variable aleatoria condicionada al suceso $Y = y_j$**, y se denota $X|Y = y_j$, como la variable aleatoria determinada por

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

para cada $x_i \in \mathbb{R}$. De forma totalmente análoga se define la **variable aleatoria condicionada al suceso $X = x_i$**, para cualquier $x_i \in \mathbb{R}$ que verifique $P(X = x_i) > 0$.

Ejemplo. Considérese el experimento aleatorio consistente en lanzar una moneda al aire cinco veces. El espacio de probabilidad sería $(\Omega, \mathcal{P}(\Omega), P)$, donde P es la medida uniforme y

$$\Omega = \{CCCCC, FFFFF, CCCCF, CCCFC, CCFCC, \dots\}$$

Sea X la variable aleatoria que cuenta el número de resultados distintos en el primer lanzamiento, y sea Y la variable aleatoria que cuenta el número de resultados distintos en el quinto lanzamiento. Por ejemplo, para ilustrar un poco el panorama, se tendría

$$\begin{aligned} X(CCCCC) &= 0 \qquad & Y(CCCCC) &= 0 \\ X(CFCFC) &= 2 \qquad & Y(CFCFC) &= 2 \\ X(CFFCF) &= 3 \qquad & Y(CFFCF) &= 2 \end{aligned}$$

Las probabilidades P_X, P_Y y $P_{(X,Y)}$ quedan totalmente determinadas por la tabla siguiente:

$X \backslash Y$	0	1	2	3	4	P_X
0	$\frac{1}{16}$	0	0	0	0	$\frac{1}{16}$
1	0	$\frac{3}{16}$	0	0	$\frac{1}{16}$	$\frac{4}{16}$
2	0	0	$\frac{3}{16}$	$\frac{3}{16}$	0	$\frac{6}{16}$
3	0	0	$\frac{3}{16}$	$\frac{1}{16}$	0	$\frac{4}{16}$
4	0	$\frac{1}{16}$	0	0	0	$\frac{1}{16}$
P_Y	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$	1

Por ejemplo, si se quisiera calcular la probabilidad de $(0,0)$, sería

$$\begin{aligned} P_{(X,Y)}(\{(0,0)\}) &= P(X = 0, Y = 0) = P(\{CCCCC, FFFFF\}) \\ &= P(\{CCCCC\}) + P(\{FFFFF\}) = \frac{1}{2^5} + \frac{1}{2^5} = \frac{1}{16} \end{aligned}$$

Por último, va a tratar de calcularse la distribución de probabilidad de la variable aleatoria $X|Y = 1$, cuyo rango es $D_{X|Y=1} = \{1,4\}$. Se tiene que

$$P_{X|Y=1}(\{1\}) = P(X = 1|Y = 1) = \frac{P(X = 1, Y = 1)}{P(Y = 1)} = \frac{3/16}{4/16} = \frac{3}{4}$$

$$P_{X|Y=1}(\{4\}) = P(X = 4|Y = 1) = \frac{P(X = 4, Y = 1)}{P(Y = 1)} = \frac{1/16}{4/16} = \frac{1}{4}$$

Ejemplo. Sea (X,Y) un vector aleatorio discreto y sean $D_X = \{0,1,2,3\}$ y $D_Y = \{1,2,3\}$ los rangos de X e Y, respectivamente. Supóngase además que $Y \sim U(\{1,2,3\})$ y que para todo $y \in D_Y$ se tiene que $X|Y = y \sim Bin(y, \frac{1}{2})$. Se va a tratar de obtener la distribución conjunta de (X,Y), es decir, de calcular

$$P(X = x_i, Y = y_j) = P(X = x_i | Y = y_j) P(Y = y_j),$$

siendo $D_X = \{x_i : i = 1,2,3,4\}$ y $D_Y = \{y_j : j = 1,2,3\}$. Fijando, por ejemplo $y_1 = 1$, se tiene que

$$
\begin{aligned}
P(X=0,Y=1) &= P(X=0\,|\,Y=1)P(Y=1) = \binom{1}{0}\frac{1}{2^0}\frac{1}{2^1}\frac{1}{3} = \frac{1}{6} \\
P(X=1,Y=1) &= P(X=1\,|\,Y=1)P(Y=1) = \binom{1}{1}\frac{1}{2^1}\frac{1}{2^0}\frac{1}{3} = \frac{1}{6} \\
< P(X=2,Y=1) &= P(X=2\,|\,Y=1)P(Y=1) = 0 \\
P(X=3,Y=1) &= P(X=3\,|\,Y=1)P(Y=1) = 0
\end{aligned}
$$

Realizando estas mismas cuentas para $y = 2$ e $y = 3$ queda resuelto el ejemplo.

Ejemplo. Sea (X,Y) un vector discreto y supongamos que $Y \sim U(\{1,2,3\})$, que $X\,|\,Y = y \sim Bin(y, \frac{1}{2})$ para cada $y \in D_Y = \{1,2,3\}$ y que $D_{X|Y=y} = \{0,1,\dots,y\}$. La distribución conjunta de (X,Y) ahora es

X \ Y	1	2	3	P_X
0	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{7}{24}$
1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{8}$	$\frac{11}{24}$
2	0	$\frac{1}{12}$	$\frac{1}{8}$	$\frac{5}{24}$
3	0	0	$\frac{1}{24}$	$\frac{1}{24}$
P_Y	$\frac{8}{24}$	$\frac{8}{24}$	$\frac{8}{24}$	1

Se va a exponer el cálculo de, por ejemplo, $P(X = 0, Y = 2)$. Se tiene que

$$P(X=0,Y=2) = P(X=0\,|\,Y=2)P(Y=2) = \binom{2}{0}\frac{1}{2^0}\frac{1}{2^2}\frac{1}{3} = \frac{1}{12}$$

Ejemplo. Supóngase que el número de huevos de insectos encima de una flor sigue una distribución de Poisson de parámetro $\lambda > 0$ (es decir, λ es la frecuencia media con la que los insectos ponen huevos). Supóngase además que, fijado un intervalo de tiempo, la probabilidad de que un huevo eclosione es $p \in (0,1)$. Los objetivos del ejemplo son

(*i*) *Calcular la probabilidad de que no nazca ningún insecto en dicho intervalo de tiempo.* Al traducir el panorama al universo de las matemáticas, se observa que hay calcular $P(X = 0)$, donde

$$X \equiv \textit{número de huevos en la flor que han eclosionado}$$

Asimismo, considérese la variable aleatoria

$$Y \equiv \textit{número de huevos en la flor}$$

Nótese que $D_X = D_Y = \mathbb{N}$, que $Y \sim P(\lambda)$ y que $X\,|\,Y = n \sim Bin(n,p)$ para cada $n \in \mathbb{N}$. Además,

$$
\begin{aligned}
P(X=0) &= \sum_{n=0}^{\infty} P(X=0,Y=n) = \sum_{n=0}^{\infty} P(X=0\,|\,Y=n)P(Y=n) = \sum_{n=0}^{\infty}(1-p)^n e^{-\lambda}\frac{\lambda^n}{n!} \\
&= e^{-\lambda}\sum_{n=0}^{\infty}\frac{(\lambda(1-p))^n}{n!} = e^{-\lambda}e^{\lambda(1-p)} = e^{-\lambda p}
\end{aligned}
$$

(*ii*) *Calcular la probabilidad de que nazcan k insectos en dicho intervalo de tiempo.* Razonando como antes, ahora es menester calcular $P(X=k)$. Se tiene que

$$\begin{aligned}
P(X=k) &= \sum_{n=0}^{\infty} P(X=k, Y=n) = \sum_{n=0}^{\infty} P(X=k \mid Y=n) P(Y=n) \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} = \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} \\
&= \frac{\lambda^k p^k e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \frac{(\lambda(1-p))^{n-k}}{(n-k)!} = \frac{\lambda^k p^k e^{-\lambda} e^{\lambda(1-p)}}{k!} = e^{-\lambda p} \frac{(\lambda p)^k}{k!}
\end{aligned}$$

Esto demuestra que $X \sim P(\lambda p)$.

(iii) *Calcular de que hubiesen menos de dos huevos en la flor tras haber observado que no ha nacido ningún insecto.* Se trata de calcular $P(Y<2 \mid X=0)$, para lo que habrá que echar mano del teorema de Bayes.

$$\begin{aligned}
P(Y<2 \mid X=0) &= P(Y=0 \mid X=0) + P(Y=1 \mid X=0) \\
&= \frac{P(X=0 \mid Y=0)P(Y=0)}{P(X=0)} + \frac{P(X=0 \mid Y=1)P(Y=1)}{P(X=0)} \\
&= \frac{e^{-\lambda}}{e^{-\lambda p}} + \frac{(1-p)e^{-\lambda}\lambda}{e^{-\lambda p}} \\
&= \frac{1+\lambda(1-p)}{e^{p}}
\end{aligned}$$

7.5. Vectores aleatorios absolutamente continuos

Definición 49. Un vector aleatorio (X,Y) con función de distribución conjunta $F:\mathbb{R}^2 \to [0,1]$ se dice que es **absolutamente continuo** (o que posee una distribución de probabilidad **absolutamente continua**) si existe una función $f:\mathbb{R}^2 \to \mathbb{R}$ integrable y no negativa tal que

$$F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u,v)\,du\,dv$$

para cada $(x,y) \in \mathbb{R}^2$. La función f se denominará **función de densidad conjunta**.

Proposición 32. *Sea (X,Y) un vector aleatorio absolutamente continuo con probabilidad inducida $P_{(X,Y)}$, función de distribución F y función de densidad conjunta f. Entonces*

(i) *Se verifica que*

$$\int_{\mathbb{R}^2} f(u,v)\,du\,dv = 1$$

(ii) *Para todo $B \in \mathcal{B}^2$,*

$$P_{(X,Y)}(B) = \int_B f(u,v)\,du\,dv$$

(iii) *Para todo $(x_0,y_0) \in \mathbb{R}^2$,*

$$\frac{\partial^2 F(x_0,y_0)}{\partial x \partial y} = \frac{\partial^2 F(x_0,y_0)}{\partial y \partial x} = f(x_0,y_0)$$

(iv) *X e Y son variables aleatorias absolutamente continuas con funciones de densidad*

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy \qquad \text{y} \qquad f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx$$

(v) *Si $y_0 \in \mathbb{R}$ es tal que $f_Y(y_0) > 0$, entonces la variable aleatoria $X|_{Y=y_0}$ es absolutamente continua* *y su función de densidad es*

$$f_{X|_{Y=y_0}}(x) = \frac{f(x,y_0)}{f_Y(y_0)}$$

(*vi*) *Si $x_0 \in \mathbb{R}$ es tal que $f_X(x_0) > 0$, entonces la variable aleatoria $Y|_{X=x_0}$ es absolutamente continua y su función de densidad es*

$$f_{Y|_{X=x_0}}(y) = \frac{f(x_0,y)}{f_X(x_0)}$$

(*vii*) *Si X e Y son variables independientes, entonces $f(x,y) = f_X(x)f_Y(y)$ para todo $(x,y) \in \mathbb{R}^2$.*

Demostración. Tampoco. □

Obsérvese que puede que X e Y sean variables aleatorias absolutamente continuas y el vector aleatorio (X,Y) no sea absolutamente continuo (tómese $Y = X$, por ejemplo).

Proposición 33 (Cambio de variable). *Sea (X,Y) un vector aleatorio absolutamente continuo con densidad conjunta $f_{(X,Y)}$, y sea $T \subset \mathbb{R}$ el conjunto de puntos donde $f_{(X,Y)}$ toma valores positivos. Sea $g\colon T \to \mathbb{R}^2$ una función inyectiva y diferenciable tal que, para todo $(u,v) \in T$,*

$$\det Jg(u,v) = \begin{vmatrix} \frac{\partial g_1}{\partial u}(u,v) & \frac{\partial g_1}{\partial v}(u,v) \\ \frac{\partial g_2}{\partial u}(u,v) & \frac{\partial g_2}{\partial v}(u,v) \end{vmatrix} \neq 0$$

Entonces $(U,V) = g(X,Y)$ es un vector aleatorio absolutamente continuo, con densidad

$$f_{(U,V)}(u,v) = f_{(X,Y)}(g^{-1}(u,v))|\det Jg(g^{-1}(u,v))|^{-1}$$

Demostración. Utilícese el teorema de la función inversa. □

Ejemplo. Sea (X,Y) un vector aleatorio absolutamente continuo con densidad conjunta

$$f(x,y) = \begin{cases} x+y & \text{si } 0<x<1 \\ 0 & \text{en caso contrario} \end{cases}$$

Se tratará de calcular la función de distribución conjunta F. Sea $(x,y) \in \mathbb{R}^2$.

(*i*) Si $x<0$ o $y<0$,

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} 0\,du\,dv = 0$$

(*ii*) Si $0<x<1$ y $0<y<1$,

$$F(x,y) = \int_0^x \int_0^y (u+v)\,dv\,du = \int_0^x \Big[uv + \frac{v^2}{2}\Big]_0^y du = \Big[\frac{u^2y}{2} + \frac{uy^2}{2}\Big]_0^x = \frac{x^2y + xy^2}{2}$$

(*iii*) Si $0<x<1$ y $1 \le y$,

$$F(x,y) = \int_0^x \int_0^1 (u+v)\,du\,dv = \frac{x^2+x}{2}$$

(*iv*) Si $1 \le x$ y $0<y<1$,

$$F(x,y) = \int_0^1 \int_0^y (u+v)\,du\,dv = \frac{y^2+y}{2}$$

(*v*) Si $1 \le x$ y $1 \le 1$,

$$F(x,y) = \int_0^1 \int_0^1 (u+v)\,du\,dv = 1$$

Ejemplo. Sea (X,Y) un vector aleatorio absolutamente continuo con función de densidad dada por

$$f(x,y) = \begin{cases} ayx^2 & \text{si } 0 < y < x < 1 \\ 0 & \text{en caso contrario} \end{cases}$$

Los propósitos del ejemplo son

(*i*) *Calcular el valor de a*. Debe cumplirse lo siguiente:

$$\int_{\mathbb{R}^2} f(x,y)\,dx\,dy = 1$$

Se tiene que

$$\int_0^1 \int_0^x ayx^2\,dy\,dx = \int_0^1 \left[\frac{ay^2x^2}{2}\right]_0^x dx = \int_0^1 \frac{ax^4}{2}\,dx = \left[\frac{ax^5}{10}\right]_0^1 = \frac{a}{10},$$

de donde se deduce que $a = 10$.

(*ii*) *Hallar las distribuciones marginales*. Si $0 < x < 1$,

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy = \int_0^x 10yx^2\,dy = \left[5y^2x^2\right]_0^x = 5x^4$$

Por tanto,

$$f_X(x) = \begin{cases} 0 & \text{si } x \le 0 \\ 5x^4 & \text{si } 0 < x < 1 \\ 5x^2 & \text{si } 1 \le x \end{cases}$$

Por otra parte, si $0 < y < 1$,

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx = \int_y^1 10yx^2\,dx = \left[\frac{10yx^3}{3}\right]_y^1 = \frac{10y}{3} - \frac{10y^4}{3}$$

En consecuencia,

$$f_Y(y) = \begin{cases} \dfrac{10y}{3} & \text{si } y \le 0 \\ \dfrac{10y}{3}(1-y^3) & \text{si } 0 < y < 1 \\ 0 & \text{si } 1 \le y, \end{cases}$$

con lo que quedan determinadas las distribuciones marginales de X e Y.

(*iii*) *Hallar las densidades de las variables aleatorias $X|_{Y=y_0}$ e $Y|_{X=x_0}$, donde $x_0, y_0 \in (0,1)$*. En primer lugar, si $y_0 < x < 1$,

$$f_{X|_{Y=y_0}}(x) = \frac{f(x,y_0)}{f_Y(y_0)} = \frac{10y_0x^2}{\frac{10y_0}{3}(1-y_0^3)} = \frac{3x^2}{1-y_0^3}$$

Por tanto,

$$f_{X|_{Y=y_0}}(x) = \begin{cases} \dfrac{3x^2}{1-y_0^3} & \text{si } y_0 < x < 1 \\ 0 & \text{en caso contrario} \end{cases}$$

Por otra parte, si $0 < y < x_0$,

$$f_{Y|_{X=x_0}}(y) = \frac{f(x_0,y)}{f_X(x_0)} = \frac{10yx_0^2}{5x_0^4} = \frac{2y}{x_0^2}$$

En consecuencia,

$$f_{Y|_{X=x_0}}(y) = \begin{cases} \dfrac{2y}{x_0^2} & \text{si } 0 < y < x_0 \\ 0 & \text{en caso contrario} \end{cases}$$

(*iv*) *Calcular la esperanza de la variable aleatoria* $Y|_{X=\frac{1}{2}}$. Se tiene que

$$E[Y|_{X=\frac{1}{2}}] = \int_{\mathbb{R}} y f_{Y|_{X=\frac{1}{2}}}(y)\,dy = \int_0^{\frac{1}{2}} 8y^2\,dy = \Big[\frac{8y^3}{3}\Big]_0^{\frac{1}{2}} = \frac{1}{3}$$

Ejemplo. Sea (X,Y) un vector aleatorio absolutamente continuo con densidad conjunta

$$f(x,y) = \begin{cases} e^{-x-y} & \text{si } x > 0, y > 0 \\ 0 & \text{en caso contrario} \end{cases}$$

Se tratará de hallar la distribución de la variable aleatoria $U = \frac{X}{X+Y}$. Para ello, considérense la variable $V = X+Y$ y el vector $(U,V) = g(X,Y)$, donde

$$\begin{aligned} g\colon T &\longrightarrow \mathbb{R}^2 \\ (x,y) &\longmapsto g(x,y) = (\tfrac{x}{x+y}, x+y), \end{aligned}$$

siendo $T = \{(x,y) \in \mathbb{R}^2 : x > 0, y > 0\}$. Se tiene que g es inyectiva y diferenciable, y si $(x,y) \in T$,

$$\det Jg(x,y) = \begin{vmatrix} \dfrac{\partial g_1}{\partial x}(x,y) & \dfrac{\partial g_1}{\partial y}(x,y) \\ \dfrac{\partial g_2}{\partial x}(x,y) & \dfrac{\partial g_2}{\partial y}(x,y) \end{vmatrix} = \begin{vmatrix} \dfrac{y}{(x+y)^2} & -\dfrac{x}{(x+y)^2} \\ 1 & 1 \end{vmatrix} = \frac{1}{x+y} \neq 0$$

Además, la inversa de g viene dada por $g^{-1}(u,v) = (uv, v(1-u))$, luego

$$\det Jg(g^{-1}(u,v)) = \frac{1}{v}$$

Se concluye que (U,V) es una variable aleatoria absolutamente continua con función de densidad

$$f_{(U,V)}(u,v) = ve^{-uv-v(1-u)} = ve^{-v},$$

siempre que $(u,v) \in g(T)$. Ya puede hallarse fácilmente la función de densidad de U.

Teorema central del límite

8.1. Convergencia en distribución

Definición 50. Sea $\{X_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias, y sean $\{F_n\}_{n=1}^{\infty}$ sus funciones de distribución. Se dice que **la sucesión $\{X_n\}_{n=1}^{\infty}$ converge en distribución a la variable aleatoria X**, si

$$\lim_{n\to\infty} F_n(x) = F(x)$$

para todo $x \in C_F = \{x \in \mathbb{R} \colon F \text{ es continua en } x\}$, siendo F la función de distribución de X. Se suele denotar

$$X_n \xrightarrow[n\to\infty]{d} X$$

8.2. Teorema central del límite

Teorema 7. *Si X es una variable aleatoria absolutamente continua con $X \sim N(\mu,\sigma^2)$, entonces*

$$\frac{X-\mu}{\sigma} \sim N(0,1)$$

Demostración. Sea $\varphi \colon \mathbb{R} \to \mathbb{R}$ la función definida por

$$\varphi(x) = \frac{x-\mu}{\sigma}$$

La función φ es una función de Borel (pues es continua), es derivable, tiene derivada continua y es estrictamente monótona. El teorema del cambio de variable permite afirmar que

$$Z = \varphi \circ X = \frac{X-\mu}{\sigma}$$

es una variable aleatoria absolutamente continua, y su densidad es

$$f_Y(y) = f_X(\varphi^{-1}(y))|(\varphi^{-1})'(y)|,$$

donde $\varphi^{-1} \colon \mathbb{R} \to \mathbb{R}$ está definida por $\varphi^{-1}(y) = \sigma y + \mu$. Por tanto,

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\sigma y+\mu-\mu)^2}{2\sigma^2}} |\sigma| = \frac{|\sigma|}{\sqrt{2\pi}|\sigma|} e^{-\frac{\sigma^2 y^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

concluyéndose que $Z \sim N(0,1)$. □

Este teorema permite reducir el cálculo de probabilidades de distribuciones normales cualesquiera al cálculo de probabilidades de la distribución normal estándar, pues si $X \sim N(\mu,\sigma^2)$, entonces, para cada $a \in \mathbb{R}$,

$$P(X < a) = P\Big(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\Big) = P(Z < \frac{a-\mu}{\sigma}),$$

donde $Z \sim N(0,1)$.

Teorema 8 (Teorema central del límite de Lindeberg-Lévy). *Consideremos una sucesión de variables aleatorias* $\{X_n\}_{n=1}^{\infty}$ *verificando:*

(i) *Las variables aleatorias son mutuamente independientes.*

(ii) *Todas las variables aleatorias tienen la misma distribución de probabilidad.*

(iii) *Existen* $\mu = E[X_n]$ *y* $\sigma^2 = \text{Var}[X_n]$.

Entonces

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow[n\to\infty]{d} N(0,1),$$

donde $S_n = X_1 + \ldots + X_n$ *para cada* $n \in \mathbb{N}$.

Demostración. No hace falta. □

Corolario 3 (Teorema de De Moivre-Laplace). *Sea* $\{X_n\}_{n=1}^{\infty}$ *una sucesión de variables aleatorias mutuamente independientes tales que* $X_n \sim Ber(p)$ *para todo* $n \in N$. *Entonces*

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow[n\to\infty]{d} N(0,1),$$

donde $S_n = X_1 + \ldots + X_n$ *para cada* $n \in \mathbb{N}$.

Demostración. Consecuencia inmediata del teorema anterior. □

La importancia de este corolario radica en que, para un natural n lo suficientemente grande, se verifica

$$Bin(n,p) \approx N(np, np(1-p))$$

Esto se debe a que la variable aleatoria S_n del corolario anterior sigue una distribución binomial de parámetros n y p, siendo $\mu = np$ y $\sigma^2 = np(1-p)$ su esperanza y su varianza. Por tanto,

$$\frac{S_n - \mu}{\sigma^2} \approx N(0,1),$$

así que se podría decir que

$$Bin(n,p) \approx N(\mu, \sigma^2) = N(np, np(1-p))$$