PROBABILIDAD II

Grado en Matemáticas

Tema 3
Variables y vectores aleatorios

Javier Cárcamo

Departamento de Matemáticas
Universidad Autónoma de Madrid
javier.carcamo@uam.es

Tema 3: Variables y vectores aleatorios

1. Variables y vectores aleatorios
2. Ley imagen y distribución de probabilidad
3. Función de distribución
4. Distribuciones discretas
5. Distribuciones absolutamente continuas
6. Independencia de variables aleatorias
7. Esperanza matemática
8. El teorema de cambio de variable y sus consecuencias

Concepto de variable aleatoria

Una **variable** es cualquier aspecto relativo a un experimento. Si $(\Omega, \mathcal{F}, \mathrm{P})$ es un espacio de probabilidad, podemos pensar en una aplicación:

$$\begin{aligned} X : \Omega &\longrightarrow E \\ \omega &\longmapsto X(\omega). \end{aligned}$$

Las variables pueden ser cuantitativas o cualitativas. Nos centraremos en las **variables cuantitativas** ($E \subseteq \mathbb{R}$). Se denomina variable *aleatoria* ya que a priori no conocemos el resultado del experimento, luego no sabemos el valor que va a tomar X.

Como sabemos calcular probabilidades en $(\Omega, \mathcal{F})$ vía P, la aplicación $X : \Omega \longrightarrow E \subseteq \mathbb{R}$ inducirá una medida de probabilidad en $(E, \mathcal{B}(E))$.

Nos podemos preguntar por $\mathrm{P}(X \leq 2)$ o $\mathrm{P}(2 < X < 7)$. Para poder calcular estas probabilidades tendremos que pedir que $\{\omega \in \Omega : X(\omega) \leq 2\} \in \mathcal{F}$ y $\{\omega \in \Omega : 2 < X(\omega) < 7\} \in \mathcal{F}$.

Definición de variable aleatoria

Sea $(\Omega, \mathcal{F}, \mathrm{P})$ un espacio de probabilidad. Se llama **variable aleatoria** a toda aplicación

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \quad (\text{o } \overline{\mathbb{R}}) \\ \omega &\longmapsto X(\omega) \end{aligned}$$

verificando que para todo $B \in \mathcal{B}$, se tiene que $X^{-1}(B) \in \mathcal{F}$.

Recordamos que $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$, es la **anti-imagen** de B por X.

Notación: $X : \Omega \longrightarrow \mathbb{R}$ función.

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

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

$$\{a < X < b\} = X^{-1}((a, b)) = \{\omega \in \Omega : a < X(\omega) < b\}.$$

Función medible: Sean $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{F}')$ dos espacios medibles. Se dice que una función $f : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$ es $\mathcal{F}/\mathcal{F}'$-**medible** si para todo $A \in \mathcal{F}'$, se tiene que $f^{-1}(A) \in \mathcal{F}$.

Observación: Una variable aleatoria sobre $(\Omega, \mathcal{F})$ es una función $\mathcal{F}/\mathcal{B}$-medible (**medible Borel**).

Ejercicio: Si $f : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$ es $\mathcal{F}/\mathcal{F}'$-medible y $g : (\Omega', \mathcal{F}') \longrightarrow (\Omega'', \mathcal{F}'')$ es $\mathcal{F}'/\mathcal{F}''$-medible, entonces $g \circ f$ es $\mathcal{F}/\mathcal{F}''$-medible.

Ejercicio: $f : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$. Si $\mathcal{F}' = \sigma(\mathcal{C}')$, entonces f es $\mathcal{F}/\mathcal{F}'$-medible si y solo si para todo $C \in \mathcal{C}'$, $f^{-1}(C) \in \mathcal{F}$.

Aplicación: $X : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$ es v.a. si y sólo si $\{X \leq a\} \in \mathcal{F}$, para todo $a \in \mathbb{Q}$.

Proposición: Sea $f : (\Omega, \tau) \longrightarrow (\Omega', \tau')$, donde τ y τ' son dos topologías. Sean $\mathcal{B}_\tau$ y $\mathcal{B}_{\tau'}$ las σ-álgebras Borelianas asociadas. Si f es τ/τ' continua, entonces f es $\mathcal{B}_\tau/\mathcal{B}_{\tau'}$-medible.

Aplicación: Si $X : (\Omega, \mathcal{F}) \longrightarrow \mathbb{R}$ es v.a., entonces X^n, $|X|^\alpha$ $(\alpha > 0)$, e^X, $5X + 1$, $\frac{X-3}{X^2+1}, \ldots$, son v.a.

Proposición: Sean $f, g : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$ $\mathcal{F}/\mathcal{B}$-medibles y $\phi : \mathbb{R}^2 \longrightarrow \mathbb{R}$ continua. La función $\phi(f, g) : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$ es $\mathcal{F}/\mathcal{B}$-medible.

Aplicación: Si $X, Y : (\Omega, \mathcal{F}) \longrightarrow \mathbb{R}$ son v.a., entonces $X + Y$, XY, $X^2 + Y^2$, $\ldots$, son v.a.

Proposición: Sean $X_k : (\Omega, \mathcal{F}) \longrightarrow \overline{\mathbb{R}}$, $k \geq 1$ v.a. Las siguientes funciones son v.a.

(a) $\sup_k X_k$.
(b) $\inf_k X_k$.
(c) $\limsup_k X_k$
(d) $\liminf_k X_k$
(e) Si existe $\lim_k X_k$, entonces es v.a.

(1) **Constantes:** Toda $X : (\Omega, \mathcal{F}) \longrightarrow \mathbb{R}$ constante es v.a.

(2) **Indicadores:** Si $A \subset \Omega$, se llama **indicador de** A a la función $1_A : \Omega \longrightarrow \mathbb{R}$ tal que $1_A(\omega) = 1$ si $\omega \in A$ y $1_A(\omega) = 0$ si $\omega \notin A$. ¿Cuándo 1_A es una v.a.?

(3) **Funciones simples:** Una v.a. $X : (\Omega, \mathcal{F}) \longrightarrow \mathbb{R}$ se dice **simple** si $X(\Omega)$ es finito.

Propiedad: $X : (\Omega, \mathcal{F}) \longrightarrow \mathbb{R}$ con $X(\Omega) = \{a_1, \ldots, a_k\}$. X es v.a. simple si y solo si $\{X = a_i\} \in \mathcal{F}$, $i = 1, \ldots, k$.

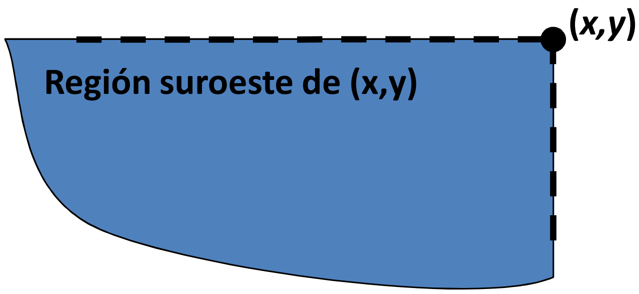
Nota: Si $X : (\Omega, \mathcal{F}) \longrightarrow \mathbb{R}$ es simple y $X(\Omega) = \{a_1, \ldots, a_k\}$, entonces $X = \sum_{i=1}^k a_i 1_{\{X=a_i\}}$.

(4) **Variables discretas:** Una v.a. X es **discreta** si $X(\Omega)$ es contable.
Ejercicio: $X : (\Omega, \mathcal{F}) \longrightarrow \mathbb{R}$ con $X(\Omega) = \{a_1, a_2 \ldots\}$. X v.a. discreta si y sólo si $\{X = a_i\} \in \mathcal{F}$, $i = 1, 2, \ldots$.

Se llama **vector aleatorio** k**-dimensional** a toda aplicación medible $\mathbf{X} : (\Omega, \mathcal{F}) \longrightarrow \mathbb{R}^k$. Es decir, $\{\mathbf{X} \in B\} \in \mathcal{F}$, $B \in \mathcal{B}(\mathbb{R}^k)$.

Dado $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$, la **región suroeste generada por x**:

$$S_{\mathbf{x}} = (-\infty, x_1] \times \cdots \times (-\infty, x_k].$$



Observación: $\mathcal{C} = \{S_{\mathbf{x}} : \mathbf{x} \in \mathbb{R}^k\}$ genera a $\mathcal{B}(\mathbb{R}^k)$. Por tanto, $\mathbf{X} : \Omega \longrightarrow \mathbb{R}$ es v.a. si y solo si $\{\mathbf{X} \in S_{\mathbf{x}}\} \in \mathcal{F}$ para todo $\mathbf{x} \in \mathbb{R}^k$.

Proposición: $\mathbf{X} = (X_1, \ldots, X_k) : \Omega \longrightarrow \mathbb{R}^k$ aplicación.
$\mathbf{X}$ vector aleatorio si y solo si $X_1, \ldots, X_k$ variables aleatorias.

Corolario: Sean $X_1, \ldots, X_k$ v.a. sobre $(\Omega, \mathcal{F})$ con valores en $\mathbb{R}$ y $g : \mathbb{R}^k \longrightarrow \mathbb{R}$ medible (en particular continua). Se tiene que $X = g(X_1, \ldots, X_k)$ es v.a.

Aplicación: $X_1, \ldots, X_k$ v.a., entonces $\sum X_i$, $\prod X_i$, $\sum X_i^2$, $e^{X_1+\cdots+X_k}, \ldots$ son v.a.

Sea $X : \Omega \longrightarrow \mathbb{R}$ (o $\overline{\mathbb{R}}$) aplicación. La colección

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}\} \quad \text{es una } \sigma\text{-álgebra}$$

denominada σ-**álgebra asociada o generada por** X.

Observación: $X : (\Omega, \mathcal{F}) \longrightarrow \mathbb{R}$ es una v.a. si y solo si $\sigma(X) \subset \mathcal{F}$.
$\sigma(X)$ es la mínima σ-álgebra sobre Ω que hace medible a X.

Los conjuntos de la forma $X^{-1}(B)$, $B \in \mathcal{B}$ se llaman **sucesos asociados a** X.

Sea $\{X_i : i \in I\}$ con $X_i : (\Omega, \mathcal{F}) \longrightarrow \mathbb{R}$.

$$\sigma(X_i : i \in I) = \sigma\left(\bigcup_{i \in I} \sigma(X_i)\right)$$

es la mínima σ-álgebra sobre Ω que hace medibles a todas las X_i simultáneamente.

Teorema fundamental de aproximación

Teorema fundamental de aproximación

Sea $X : (\Omega, \mathcal{F}) \longrightarrow [0, \infty]$.Existe una sucesión de funciones simples $\{X_n\}_{n=1}^{\infty}$ verificando:

(a) Las X_n son $\sigma(X)$-medibles.

(b) $0 \leq X_n \leq X_{n+1}$, para todo n.

(c) $\lim_{n\to\infty} X_n = X$.

Existe $\{X_n\}$ sucesión de v.a. simples $\sigma(X)$-medibles con $0 \leq X_n \uparrow X$.

Idea de la prueba: Para todo n y $\omega \in \Omega$, definimos

$$X_n(\omega) = \begin{cases} \frac{q-1}{2^n}, & \text{si} \quad \frac{q-1}{2^n} \leq X(\omega) < \frac{q}{2^n}, \quad q = 1, \ldots, n2^n, \\ n, & \text{si} \quad X(\omega) \geq n. \end{cases}$$

Es decir,

$$X_n = \sum_{q=1}^{n2^n} \frac{q-1}{2^n} 1_{\{\frac{q-1}{2^n} \leq X < \frac{q}{2^n}\}} + n 1_{\{X \geq n\}}, \quad \sigma(X)\text{-medible}.$$

Ley imagen

$(\Omega, \mathcal{F}, \mathrm{P})$ espacio de probabilidad.

$$f : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow (\Omega', \mathcal{F}') \text{ función } \mathcal{F}/\mathcal{F}'\text{-medible.}$$

Es decir, para todo $B \in \mathcal{F}'$, tenemos que $f^{-1}(B) \in \mathcal{F}$. Por tanto podemos calcular $\mathrm{P}(f^{-1}(B))$.

$$\begin{aligned} \mathrm{P}_f : \mathcal{F}' &\longrightarrow [0, 1] \\ B &\longmapsto \mathrm{P}_f(B) = \mathrm{P}(f^{-1}(B)). \end{aligned}$$

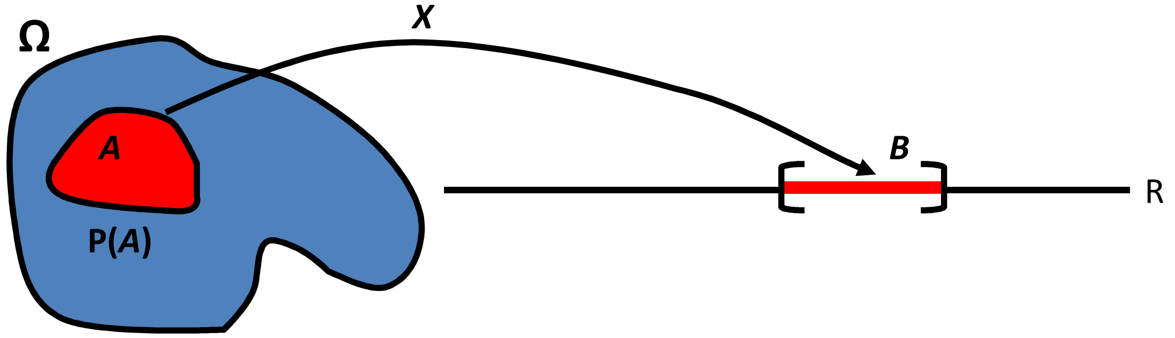
Teorema

P_f es una medida de probabilidad sobre $(\Omega', \mathcal{F}')$ que se llama **ley imagen** de P bajo f.

$X : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow (\mathbb{R}, \mathcal{B})$ variable aleatoria ($\Omega' = \mathbb{R}$ o $\overline{\mathbb{R}}$ y $\mathcal{F}' = \mathcal{B}$).

$$\begin{aligned} \mathrm{P}_X : \mathcal{B} &\longrightarrow [0,1] \\ B &\longmapsto \mathrm{P}_X(B) = \mathrm{P}(X \in B). \end{aligned}$$

P_X se llama **distribución de probabilidad de** X y es una medida de probabilidad en $(\mathbb{R}, \mathcal{B})$ o $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$.



Definición: $X : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ e $Y : (\Omega^*, \mathcal{F}^*, \mathrm{P}^*) \longrightarrow \mathbb{R}$ variables aleatorias. Se dice que X e Y están **idénticamente distribuidas** y lo denotamos $X =_d Y$ si $\mathrm{P}_X = \mathrm{P}^*_Y$.

Es decir, si para todo $B \in \mathcal{B}$, se tiene $\mathrm{P}_X(B) = \mathrm{P}^*_Y(B)$, o lo que es lo mismo, para todo $B \in \mathcal{B}$, $\mathrm{P}(X \in B) = \mathrm{P}^*(Y \in B)$.

Teorema: Sea $g : \mathbb{R} \longrightarrow \mathbb{R}$ medible. Si $X =_d Y$, entonces $g(X) =_d g(Y)$.

Aplicaciones: Si $X =_d Y$, entonces $X^2 =_d Y^2$, $e^X =_d e^Y$, $\sin X =_d \sin Y, \ldots$

Definición: $X, Y : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ variables aleatorias. Se dice que X e Y son **casi seguramente iguales** y lo denotamos $X =_{cs} Y$ si $\mathrm{P}(X = Y) = 1$ (o $\mathrm{P}(X \neq Y) = 0$).

Proposición: Si $X =_{cs} Y$, entonces $X =_d Y$.

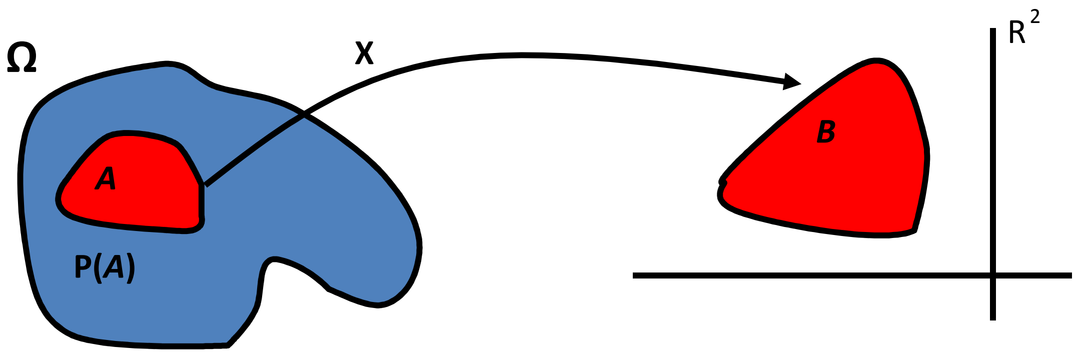
Observación: Si $X =_d Y$, X e Y ni siquiera tienen que estar definidas en un mismo espacio de probabilidad. Aun suponiendo que $X, Y : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ y $X =_d Y$, esto *no* implica que $X =_{cs} Y$.

Ejemplo: $(\Omega, \mathcal{F}, \mathrm{P})$ espacio de probabilidad donde existe $A \in \mathcal{F}$ con $\mathrm{P}(A) = 1/2$. Tomemos $X = 1_A$ e $Y = 1_{A^c}$.

$\mathbf{X} = (X_1, \ldots, X_k) : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^k$ vector aleatorio.

$$\begin{aligned} \mathrm{P}_{\mathbf{X}} : & \mathcal{B}(\mathbb{R}^k) \longrightarrow [0,1] \\ & B \longmapsto \mathrm{P}_{\mathbf{X}}(B) = \mathrm{P}(\mathbf{X} \in B). \end{aligned}$$

$\mathrm{P}_{\mathbf{X}}$ o $\mathrm{P}_{(X_1,\ldots,X_k)}$ se llama **distribución de probabilidad de X** o **distribución conjunta de las variables** $X_1, \ldots, X_k$ y es una medida de probabilidad en $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ o $(\overline{\mathbb{R}}^k, \mathcal{B}(\overline{\mathbb{R}}^k))$.

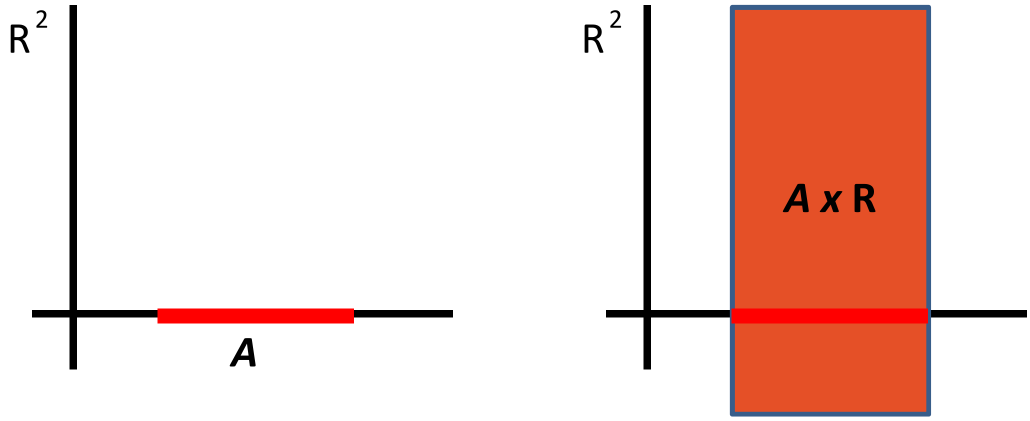


A las distribuciones de las variables aleatorias $X_1, \ldots, X_k$, $\mathrm{P}_{X_1}, \ldots, \mathrm{P}_{X_k}$, se las llama **distribuciones marginales de X**.

Observación: $A \in \mathcal{B}$, $\{X_1 \in A\} = \{\mathbf{X} \in A \times \mathbb{R} \times \cdots \times \mathbb{R}\}$.

Luego, $\mathrm{P}(X_1 \in A) = \mathrm{P}(\mathbf{X} \in A \times \mathbb{R} \times \cdots \times \mathbb{R})$.

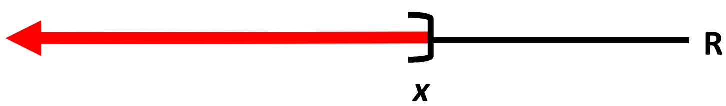
Idea física: $n = 2$, $\mathrm{P}(X \in A) = \mathrm{P}((X, Y) \in A \times \mathbb{R})$



La masa que la variable aleatoria X concentra en A es la masa que el vector aleatorio (X, Y) concentra en $A \times \mathbb{R}$.

Definición: Sea $X : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ variable aleatoria. Se llama **función de distribución de** X a la función

$$\begin{aligned} F_X : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto F_X(x) = \mathrm{P}(X \leq x) = \mathrm{P}_X((-\infty, x]). \end{aligned}$$



Proposición: X e Y variables aleatorias (no necesariamente definidas sobre el mismo espacio de probabilidad). Se tiene $X =_d Y$ si y solo si $F_X = F_Y$.

Demostración: P_X es la medida de Borel-Stieltjes asociada a la función F_X.

Teorema: Propiedades básicas de la función de distribución

Sea X variable aleatoria y F su función de distribución.

1. F es no decreciente ($x \leq y$, entonces $F(x) \leq F(y)$).
2. F es continua por la derecha ($x_n \downarrow x$, entonces $F(x_n) \downarrow F(x)$).
3. $\lim_{x \to -\infty} F(x) = 0$; $\lim_{x \to +\infty} F(x) = 1$.

Teorema

Sea $F : \mathbb{R} \longrightarrow \mathbb{R}$ verificando 1, 2 y 3 de arriba, entonces, existe $X : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ variable aleatoria tal que $F = F_X$.

Demostración: Se basa en el Teorema de extensión de medidas de Caratheodory.

Cálculo de probabilidades con F

1. $\mathrm{P}(X \leq x) = F(x)$.
2. $\mathrm{P}(X > x) = 1 - F(x)$.
3. $\mathrm{P}(X < x) = F(x^-)$.
4. $\mathrm{P}(X \geq x) = 1 - F(x^-)$.
5. $\mathrm{P}(x < X \leq y) = F(y) - F(x)$.
6. $\mathrm{P}(x \leq X \leq y) = F(y) - F(x^-)$.
7. $\mathrm{P}(x \leq X < y) = F(y^-) - F(x^-)$.
8. $\mathrm{P}(x < X < y) = F(y^-) - F(x)$.
9. $\mathrm{P}(X = x) = F(x) - F(x^-)$.

Corolario: F continua $\Longleftrightarrow$ F continua por la izquierda $\Longleftrightarrow$ $F(x^-) = F(x)$, $x \in \mathbb{R} \Longleftrightarrow \mathrm{P}(X = x) = 0$, $x \in \mathbb{R}$.

Observación: $f : \mathbb{R} \longrightarrow \mathbb{R}$ monótona, entonces $\{x : f \text{ discontinua en } x\}$ es contable. En particular $C_f = \{x : f \text{ continua en } x\}$ es denso.

Definición: Sea $\mathbf{X} = (X_1, \ldots, X_k) : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^k$ vector aleatorio. Se llama **función de distribución de X** o **función de distribución conjunta de las X_i-s** a la función $F_{\mathbf{X}} : \mathbb{R}^k \longrightarrow \mathbb{R}$ tal que

$$F_{\mathbf{X}}(\mathbf{x}) = \mathrm{P}(\mathbf{X} \in S_{\mathbf{x}}) = \mathrm{P}(X_1 \leq x_1, \ldots, X_k \leq x_k).$$

Teorema: Propiedades básicas de la función de distribución

Sea $\mathbf{X}$ vector aleatorio y F su función de distribución.

1. F es no decreciente (en $\mathbb{R}^k$).
2. F es continua por la derecha en cada variable.
3. $\lim_{x_i \to -\infty} F(x_1, \ldots, x_k) = 0$ $(i = 1, \ldots, k)$; $\lim_{x_1, \ldots, x_k \to +\infty} F(x_1, \ldots, x_k) = 1$.

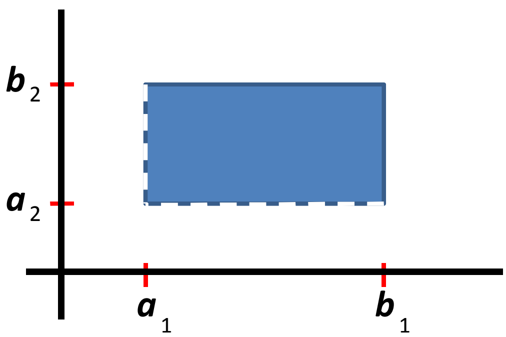
Pregunta: ¿Qué significa que una función sea monótona (no decreciente) en $\mathbb{R}^k$?

Notación: Si $G : \mathbb{R}^k \longrightarrow \mathbb{R}$ y $u \leq v$,

$\Delta_i(u, v) = G(x_1, \ldots, x_{i-1}, v, x_{i+1}, \ldots, x_k) - G(x_1, \ldots, x_{i-1}, u, x_{i+1}, \ldots, x_k)$.

En $k = 1$, para $a \leq b$, $\Delta(a, b)F(x) = F(b) - F(a) \geq 0$.

En $k = 2$, para $a_1 \leq b_1$ y $a_2 \leq b_2$, se pide que $\Delta_1(a_1, b_1)\Delta_2(a_2, b_2)F(x, y) = \Delta_1(a_1, b_1)(F(x, b_2) - F(x, a_2)) = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \geq 0$.



En general, se pide que, para $a_i \leq b_i$ $(i = 1, \ldots, k)$

$$\Delta_1(a_1, b_1)\Delta_2(a_2, b_2) \cdots \Delta_k(a_k, b_k)F(x_1, \cdots, x_k) \geq 0.$$

Teorema

Sea $F : \mathbb{R}^k \longrightarrow \mathbb{R}$ verificando ❶, ❷ y ❸ de antes, entonces, existe $\mathbf{X} : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^k$ vector aleatorio tal que $F = F_{\mathbf{X}}$.

Proposición: $\mathbf{X}$ e $\mathbf{Y}$ vectores aleatorios (no necesariamente definidas sobre el mismo espacio de probabilidad). Se tiene, $\mathbf{X} =_d \mathbf{Y}$ si y solo si $F_{\mathbf{X}} = F_{\mathbf{Y}}$.

Nota: (Relación entre $F_{\mathbf{X}}$ y $F_{X_1}, \dots, F_{X_k}$)
$\mathbf{X} = (X_1, \dots, X_k) : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^k$ vector aleatorio.
$F_{\mathbf{X}}(x_1, \dots, x_k) = \mathrm{P}(X_1 \le x_1, \dots, X_k \le x_k)$; $F_{X_i}(x_i) = \mathrm{P}(X_i \le x_i)$

$$F_{X_1}(x_1) = \lim_{x_2, \dots, x_k \to \infty} F_{\mathbf{X}}(x_1, \dots, x_k).$$

Los tres tipos puros de distribuciones

1. Discretas
2. Absolutamente continuas
3. Continuas singulares

Distribuciones discretas

Definición: Un vector aleatorio $\mathbf{X} : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^k$ se dice que tiene **distribución discreta** si existe un conjunto $S \subset \mathbb{R}^k$ contable tal que $\mathrm{P}(\mathbf{X} \in S) = 1$ ($\mathrm{P}(\mathbf{X} \in S^c) = 0$). Es decir, $\mathbf{X}$ concentra su masa en S. S se llama **soporte de la distribución** (de $\mathbf{X}$).

Nota: $\mathbf{X}$ vector discreto con soporte $S \subset \mathbb{R}^k$, entonces

$$\mathrm{P}(\mathbf{X} \in B) = \sum_{s \in S \cap B} \mathrm{P}(\mathbf{X} = s), \quad B \in \mathcal{B}^k.$$

Teorema: $\mathbf{X} = (X_1, \dots, X_k) : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^k$ vector aleatorio.
$\mathbf{X}$ vector discreto $\Longleftrightarrow X_1, \dots, X_k$ variables discretas.

Definición: Un vector aleatorio $\mathbf{X} = (X_1, \ldots, X_k) : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^k$ se dice que es **absolutamente continuo** si existe una función medible borel $f : \mathbb{R}^k \longrightarrow [0, \infty]$ tal que

$$\mathrm{P}(\mathbf{X} \in B) = \int_B f = \int_{\mathbb{R}^k} f\, 1_B, \quad B \in \mathcal{B}^k.$$

La función f se llama **densidad de probabilidad del vector X** o **densidad conjunta de las variables** $X_1, \ldots, X_k$.

En particular, la función de distribución de **X**

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{S_{\mathbf{x}}} f = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(t_1, \ldots, t_k)\, dt_1 \cdots dt_k.$$

Se puede demostrar que la densidad

$$f(x_1, \ldots, x_k) = \frac{\partial^k F_{\mathbf{X}}(x_1 \ldots, x_k)}{\partial x_1 \cdots \partial x_k} \quad \text{c.s. (medida de Lebesgue).}$$

Teorema: Si $\mathbf{X} = (X_1, \ldots, X_k)$ vector aleatorio con densidad f, entonces X_i $(i = 1, \ldots, k)$ tiene densidad

$$f_i(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_k)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_k.$$

Ejemplo: Si $\mathbf{X} = (X_1, X_2)$ vector aleatorio con densidad $f(x, y)$. Las densidades de X_1 y X_2 son:

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy, \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

Observación: Puede ocurrir que $X_1, \ldots, X_k$ variables aleatorias con densidad (es decir, cada una de ellas es absolutamente continua) y que el vector $\mathbf{X} = (X_1, \ldots, X_k)$ *no* tenga densidad.

Ejemplo:

Distribuciones continuas singulares

Definición: Un vector aleatorio $\mathbf{X} : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^k$ se dice que tiene distribución **continua singular** si $F_{\mathbf{X}}(x_1, \dots, x_k)$ es continua y existe $B \in \mathcal{B}^k$ con $m_k(B) = 0$ (medida de Lebesgue en $\mathbb{R}^k$) y $\mathrm{P}(\mathbf{X} \in B) = 1$. **No las manejaremos en la práctica.**

Teorema de descomposición

Sea $\mathbf{X} : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}^k$ un vector aleatorio. La medida de probabilidad $\mathrm{P}_{\mathbf{X}}$ se descompone **de manera única** como:

$$\mathrm{P}_{\mathbf{X}} = a\mathrm{P}_{\mathrm{d}} + b\mathrm{P}_{\mathrm{ac}} + c\mathrm{P}_{\mathrm{cs}},$$

donde $a, b, c \geq 0$ y $a + b + c = 1$ con

- P_{d} es una medida de probabilidad discreta sobre $(\mathbb{R}^k, \mathcal{B}^k)$.
- P_{ac} es una medida de probabilidad absolutamente continua sobre $(\mathbb{R}^k, \mathcal{B}^k)$.
- P_{cs} es una medida de probabilidad continua singular sobre $(\mathbb{R}^k, \mathcal{B}^k)$.

Demostración: Teorema de descomposición de Lebesgue.

Independencia de variables aleatorias

Definición: Sea $\{X_i : i \in I\}$ una familia de v.a. Se dice que las v.a. X_i son **(mutuamente) independientes** si las σ-álgebras generadas por ellas $\sigma(X_i)$ son independientes.

$$\sigma(X_i) = \{X_i^{-1}(B) : B \in \mathcal{B}\} = \{\{X_i \in B\} : B \in \mathcal{B}\}.$$

Observaciones:

1. $\{X_i : i \in I\}$ ind. $\Longleftrightarrow \forall \{i_1, \dots, i_k\} \subset I$, $X_{i_1}, \dots, X_{i_k}$ ind.
2. $X_1, \dots, X_k$ ind. $\Longleftrightarrow \forall A_1 \in \sigma(X_1), \dots, \forall A_k \in \sigma(X_k)$, $\mathrm{P}(A_1 \cdots A_k) = \mathrm{P}(A_1) \cdots \mathrm{P}(A_k)$.
3. $X_1, \dots, X_k$ ind. $\Longleftrightarrow \forall B_1, \dots, B_k \in \mathcal{B}$,
$$\mathrm{P}(X_1 \in B_1, \cdots, X_k \in B_k) = \mathrm{P}(X_1 \in B_1) \cdots \mathrm{P}(X_k \in B_k).$$
4. $X_1, \dots, X_k$ ind. $\Longleftrightarrow \forall B_1, \dots, B_k \in \mathcal{B}$,
$$\mathrm{P}_{(X_1, \dots, X_k)}(B_1 \times \cdots \times B_k) = \mathrm{P}_{X_1}(B_1) \cdots \mathrm{P}_{X_k}(B_k).$$

$X_1, \dots, X_k$ ind. $\iff \forall B_1, \dots, B_k \in \mathcal{B}$,
$$\mathrm{P}_{(X_1,\dots,X_k)}(B_1 \times \cdots \times B_k) = \mathrm{P}_{X_1}(B_1) \cdots \mathrm{P}_{X_k}(B_k).$$

Nota (Teoría de la medida): Medida producto

$\mathrm{P}_{(X_1,\dots,X_k)}$ medida de probabilidad en $(\mathbb{R}^k, \mathcal{B}^k)$.
$\mathrm{P}_{X_1}, \dots, \mathrm{P}_{X_k}$ medidas de probabilidad en $(\mathbb{R}, \mathcal{B})$.

$X_1, \dots, X_k$ ind. $\Leftrightarrow$ $\mathrm{P}_{(X_1,\dots,X_k)}$ medida producto de $\mathrm{P}_{X_1}, \dots, \mathrm{P}_{X_k}$.
$$\mathrm{P}_{(X_1,\dots,X_k)} = \mathrm{P}_{X_1} \otimes \cdots \otimes \mathrm{P}_{X_k}.$$

5. $X_1, \dots, X_k$ v.a. y $f_i : \mathbb{R} \longrightarrow \mathbb{R}$ medible Borel ($i = 1, \dots, k$).
$$X_1, \dots, X_k \text{ ind. } \Longrightarrow f_1(X_1), \dots, f_k(X_k) \text{ ind.}$$

6. $X_1, \dots, X_n, X_{n+1}, \dots, X_{n+m}$ v.a. $f_1 : \mathbb{R}^n \longrightarrow \mathbb{R}$, $f_2 : \mathbb{R}^m \longrightarrow \mathbb{R}$ medibles.
$$X_1, \dots, X_{n+m} \text{ ind. } \Longrightarrow f_1(X_1, \dots, X_n), f_2(X_{n+1}, \dots, X_{n+m}) \text{ ind.}$$

Teorema: Caracterización mediante funciones de distribución

$X_1, \dots, X_k$ v.a. con f.d. conjunta F y f.d. marginales $F_1, \dots, F_k$.
$$X_1, \dots, X_k \text{ ind.} \iff F(x_1, \dots, x_k) = F_1(x_1) \cdots F_k(X_k).$$

Teorema: Independencia variables aleatorias discretas

X_i v.a. con soporte S_i contable $i = 1, \dots, k$. Son equivalentes:

- $X_1, \dots, X_k$ ind.
- Para todo $s_i \in S_i$ ($i = 1, \dots, k$), se tiene
$$\mathrm{P}(X_1 = s_1, \dots, X_k = s_k) = \mathrm{P}(X_1 = s_1) \cdots \mathrm{P}(X_k = s_k).$$

Teorema: Independencia variables aleatorias continuas

X_i v.a. absolutamente continua con densidad $f_i(x_i)$, $i = 1, \dots, k$.

(a) Si $(X_1, \dots, X_k)$ tiene densidad $f(x_1, \dots, x_k) = f_1(x_1) \cdots f_k(x_k)$ (c.s. Lebesgue), entonces $X_1, \dots, X_k$ independientes.

(b) Si $X_1, \dots, X_k$ independientes, entonces $(X_1, \dots, X_k)$ tiene densidad $f(x_1, \dots, x_k) = f_1(x_1) \cdots f_k(x_k)$.

Idea intuitiva: Sea Ω una población con:

n_1 personas de edad e_1.

n_2 personas de edad e_2.

$\vdots \qquad \vdots \qquad \vdots$

n_k personas de edad e_k.

$n_1 + \cdots + n_k = n \equiv$ número total de individuos.

$$\textbf{Edad media} = \frac{n_1 e_1 + \cdots + n_k e_k}{n} = e_1 \frac{n_1}{n} + \cdots + e_k \frac{n_k}{n}$$

ϵ : se elige un individuo al azar. Variable: $X \equiv$ edad del individuo.
X toma los valores $e_1, \ldots, e_k$ con probabilidades $\mathrm{P}(X = e_i) = \frac{n_i}{n}$.

$$\sum_{i=1}^{k} e_i \, \mathrm{P}(X = e_i) \equiv$$ **Valor medio o esperanza matemática de X.**

Nos da una idea de entorno a qué punto se distribuye la v.a.

Caso discreto: Sea X una v.a. discreta, **la esperanza (matemática)** o **media de** X se define por:

$$\mathrm{E}X := \sum_x x \, \mathrm{P}(X = x),$$

siempre que la suma o serie sea convergente, es decir,

$$\sum_x |x| \, \mathrm{P}(X = x) < \infty \qquad \text{(Absolutamente convergente).}$$

Caso continuo: Sea X una v.a. continua con función de densidad f, **la esperanza (matemática)** o **media de** X se define por:

$$\mathrm{E}X := \int_{-\infty}^{+\infty} x \, f(x) \, dx,$$

siempre que la integral sea convergente, es decir,

$$\int_{-\infty}^{+\infty} |x| \, f(x) \, dx < \infty \qquad \text{(Absolutamente convergente).}$$

Esperanza matemática (definición formal)

Sea $X : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ variable aleatoria ($\mathcal{F}$-medible), se define la **esperanza (matemática)** de X como

$$\mathrm{E}X = \int_\Omega X \, d\mathrm{P}.$$

Observación: La $\mathrm{E}X$ es la integral (sobre Ω) de la función (medible) X respecto de la medida (de probabilidad) positiva P tal y como se estudia en el curso de Teoría de la medida y la integración.

Observación: La definición de la esperanza como integral (respecto de la medida de probabilidad subyacente) es consistente con la definición que se muestra en los cursos elementales.

Construcción de la esperanza matemática:

1. Caso de v.a. simples.
2. Extensión a v.a. no negativas.
3. Caso general.

Esperanza matemática (construcción para v.a. simples)

1. **Caso de v.a. simples:** Sea X una v.a. simple sobre $(\Omega, \mathcal{F}, \mathrm{P})$ que toma los valores $x_1, \ldots, x_m$. Es decir,

$$X = \sum_{i=1}^{m} x_i \, 1_{\{X = x_i\}}.$$

Se define la **media** o **esperanza matemática** de X

$$\mathrm{E}X \equiv \int_\Omega X \, d\mathrm{P} := \sum_{i=1}^{m} x_i \, \mathrm{P}(X = x_i).$$

Esperanza matemática (construcción para v.a. no negativas)

2. **Caso de v.a. no negativas:** $X : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow [0, \infty]$ v.a.
Existe una sucesión de v.a. simples $\{X_n\}$ tal que $0 \leq X_n \uparrow X$.
La **esperanza matemática** de X se define mediante
$$\mathrm{E}X \equiv \int_\Omega X \, d\mathrm{P} := \lim_{n\to\infty} \mathrm{E}X_n.$$
Alternativamente, se puede comprobar que
$$\mathrm{E}X \equiv \int_\Omega X \, d\mathrm{P} = \sup_{0\leq s\leq X,\ s \text{ simple}} \int_\Omega s \, d\mathrm{P}.$$

Ejercicio: Dado $(\Omega, \mathcal{F}, \mathrm{P})$ e.p. y $A \in \mathcal{F}$. Definimos la función
$$Y(\omega) = \begin{cases} +\infty & \text{si} \quad \omega \in A, \\ 0 & \text{si} \quad \omega \notin A. \end{cases}$$
Mostrar que Y es una v.a. y calcular $\mathrm{E}Y$.

Esperanza matemática (propiedades para v.a. positivas)

Sean $X, Y \geq 0$ v.a. Se tienen las siguientes propiedades:

1. $\mathrm{E}X$ siempre existe y $\mathrm{E}X \in [0, \infty]$.
2. **Linealidad:** $\mathrm{E}(X+Y) = \mathrm{E}X + \mathrm{E}Y$, $\mathrm{E}(cX) = c\mathrm{E}X$, para todo $c \geq 0$.
3. **Monotonicidad:** Si $X \leq Y$, entonces $\mathrm{E}X \leq \mathrm{E}Y$.
4. **Teorema de convergencia monótona:** Si X_n, X v.a. no negativas tales que $X_n \uparrow X$, entonces $\mathrm{E}X = \lim_{n\to\infty} \mathrm{E}X_n$.
5. Si $\mathrm{E}X < \infty$, entonces $\mathrm{P}(X = \infty) = 0$.
6. $\mathrm{E}X = 0$ si y sólo si $\mathrm{P}(X \neq 0) = 0$, es decir $X = 0$ c.s.

Esperanza matemática (construcción en el caso general)

3. **Caso general:** $X : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \overline{\mathbb{R}}$ v.a. Podemos escribir $X = X^+ - X^-$, donde $X^+ = \text{máx}\{X, 0\}$ y $X^- = \text{máx}\{-X, 0\}$ son v.a. no negativas. La v.a. X se dice que **tiene esperanza** si $\mathrm{E}X^+ < \infty$ ó $\mathrm{E}X^- < \infty$.

   Si X tiene esperanza, la **esperanza matemática** de X se define mediante
   $$\mathrm{E}X = \mathrm{E}X^+ - \mathrm{E}X^-.$$

   En particular, si $\mathrm{E}X^+, \mathrm{E}X^- < \infty$, se dice que X es **integrable** y se denota $X \in \mathcal{L}_1(\mathrm{P})$ (o $\mathcal{L}_1$).

Esperanza matemática (propiedades)

1. Si X tiene esperanza, entonces $\mathrm{E}X \in \overline{\mathbb{R}}$.
2. X integrable si y solo si $\mathrm{E}X \in \mathbb{R}$ ($X \in \mathcal{L}_1(\mathrm{P})$).
3. Si $X + Y$ está bien definida ($\mathrm{E}X^+ < \infty$ y $\mathrm{E}Y^+ < \infty$ ó $\mathrm{E}X^- < \infty$ y $\mathrm{E}Y^- < \infty$), entonces $X + Y$ tiene esperanza y $\mathrm{E}(X + Y) = \mathrm{E}X + \mathrm{E}Y$.
4. Si X tiene esperanza, entonces para todo $c \in \mathbb{R}$ cX tiene esperanza y $\mathrm{E}(cX) = c\mathrm{E}X$.
5. **Linealidad:** Si X, Y son integrables ($X, Y \in \mathcal{L}_1(\mathrm{P})$), entonces $aX + bY \in \mathcal{L}_1(\mathrm{P})$ y $\mathrm{E}(aX + bY) = a\mathrm{E}X + b\mathrm{E}Y$.
6. **Monotonicidad:** X, Y tienen esperanza y $X \leq Y$, entonces $\mathrm{E}X \leq \mathrm{E}Y$. En particular, $|\mathrm{E}X| \leq \mathrm{E}|X|$.
7. **Teorema de convergencia monótona**
8. **Lema de Fatou-Lebesgue**
9. **Teorema de la convergencia dominada**

Teorema de la convergencia monótona

Hacia arriba: Si $X_n \uparrow X$ y existe k tal que $\mathrm{E}X_k^- < \infty$, entonces $X_k, X_{k+1}, \ldots, X$ tienen esperanza y $\mathrm{E}X = \lim_{n\to\infty} \mathrm{E}X_n$.

Hacia abajo: Si $X_n \downarrow X$ y existe k tal que $\mathrm{E}X_k^+ < \infty$, entonces $X_k, X_{k+1}, \ldots, X$ tienen esperanza y $\mathrm{E}X = \lim_{n\to\infty} \mathrm{E}X_n$.

Lema de Fatou-Lebesgue

(a) Si $X_n \leq Y$ para todo n y $\mathrm{E}Y^+ < \infty$, entonces X_n, $\limsup_{n\to\infty} X_n$ tienen esperanza para todo n y

$$\mathrm{E}\left(\limsup_{n\to\infty} X_n\right) \geq \limsup_{n\to\infty} \mathrm{E}X_n.$$

(b) Si $X_n \geq Z$ para todo n y $\mathrm{E}Z^- < \infty$, entonces X_n, $\liminf_{n\to\infty} X_n$ tienen esperanza para todo n y

$$\mathrm{E}\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} \mathrm{E}X_n.$$

Teorema de la convergencia dominada

Si $X_n \to X$ y existe U variable aleatoria integrable tal que $|X_n| \leq U$ para todo n, entonces $X_1, \ldots, X_n, \ldots, X$ son integrables y

$$\mathrm{E}X = \lim_{n\to\infty} \mathrm{E}X_n.$$

Integral indefinida: Sea X una v.a. positiva o integrable sobre $(\Omega, \mathcal{F}, \mathrm{P})$ y $A \in \mathcal{F}$, se define

$$\int_A X \, d\mathrm{P} = \mathrm{E}(X 1_A) = \int_\Omega X \cdot 1_A \, d\mathrm{P}.$$

Una desigualdad de mucha aplicación

Sea X una variable aleatoria y $a, b \in \mathbb{R}$. Se tiene:

$$a\,\mathrm{P}(a \leq X \leq b) \leq \int_{\{a\leq X\leq b\}} X \, d\mathrm{P} \leq b\,\mathrm{P}(a \leq X \leq b).$$

El método de la cadena ascendente

El Teorema fundamental de aproximación y la construcción de la esperanza es la base de un método de demostración que llamaremos el **método de la cadena ascendente** (MCA):
Supongamos que queremos mostrar que la v.a. X verifica cierta propiedad $\mathcal{P}$, relacionada con esperanzas.

1. **Indicadores:** Se prueba que si $X = 1_A$, con $A \in \mathcal{F}$, entonces X verifica $\mathcal{P}$.
2. **V.a. simples:** Se prueba que si $X = \sum_{i=1}^{n} a_i 1_{A_i}$ v.a. simple, entonces X verifica $\mathcal{P}$.
3. **V.a. positivas:** Se usa el teorema fundamental de aproximación para mostrar que toda v.a. positiva X verifica $\mathcal{P}$. En este punto en ocasiones también se apela al Teorema de convergencia monótona.
4. **Caso general:** Se escribe $X = X^+ - X^-$ y se usa el punto anterior para mostrar que X verifica $\mathcal{P}$.

Esperanza matemática (independencia)

Teorema: Esperanza de un producto de v.a. ind.

$X\ Y : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ v.a. independientes e integrables. Se tiene:
La v.a. XY es integrable y $\mathrm{E}(XY) = \mathrm{E}X\mathrm{E}Y$.

Corolario: $X_1, \ldots, X_n : (\Omega, \mathcal{F}, \mathrm{P}) \longrightarrow \mathbb{R}$ v.a. independientes e integrables. Se tiene:
La v.a. $X_1 \cdots X_n$ es integrable y $\mathrm{E}(X_1 \cdots X_n) = \mathrm{E}X_1 \cdots \mathrm{E}X_n$.

Observación: Puede ocurrir que $\mathrm{E}(XY) = \mathrm{E}X\mathrm{E}Y$, pero X e Y *no* independientes.

Ejemplos:

Teorema: Cálculo de esperanzas cuando hay densidad

$$\begin{tikzcd}
(\Omega, \mathcal{F}, \mathrm{P}) \arrow[r, "\mathbf{X}=(X_1,\dots,X_k)"] \arrow[dr, dotted, "g(\mathbf{X})"'] & \mathbb{R}^k \arrow[d, "g"] \\
 & \overline{\mathbb{R}}
\end{tikzcd}$$

Supongamos que $\mathbf{X}$ y g son medibles y $\mathbf{X}$ tiene densidad f. Se tiene:

$$\mathrm{E}g(\mathbf{X}) = \int_\Omega g(\mathbf{X})\, d\mathrm{P} \quad \text{existe si y sólo si} \quad \int_{\mathbb{R}^k} g\, f\, dm_k \text{ existe,}$$

donde m_k es la medida de Lebesgue en $\mathbb{R}^k$, $dm_k = dx_1 \cdots dx_k$.

En tal caso coinciden. Es decir,

$$\int_\Omega g(\mathbf{X})\, d\mathrm{P} = \int_{\mathbb{R}^k} g\, f\, dm_k.$$

Ejemplos: Expresa en función de las densidades las siguientes esperanzas.

Sea X v.a. con densidad $f(x)$.

$\mathrm{E}X^3 =$

$\mathrm{E}e^{tX} =$

Sea (X, Y) v.a. con densidad $f(x, y)$.

$\mathrm{E}\left(\frac{X^2}{X^4+Y^4}\right) =$

Sea X, Y v.a. independientes con densidades $f_1(x)$ y $f_2(y)$.

$\mathrm{E}e^{X+Y} =$

$\mathrm{E}\sin(X+Y) =$