PROBABILIDAD II

Grado en Matemáticas

Tema 2
Nociones básicas de probabilidad.

Jesús Munárriz

Departamento de Matemáticas
Universidad Autónoma de Madrid
jesus.munarriz@uam.es

1. Espacios de probabilidad.
2. Probabilidad condicional.
3. Variables aleatorias, densidades, funciones de distribución. Media o esperanza.
4. Independencia. Correlación.
5. Algunas distribuciones importantes.
6. Modos de convergencia.

La teoría de la Probabilidad fué axiomatizada por Kolmogorov en 1933.

Un **espacio de probabilidad** $(\Omega, \mathcal{F}, P)$ es un espacio de medida tal que $P(\Omega) = 1$.

El conjunto (no vacío) Ω se denomina espacio muestral. Desde el punto de vista de la modelización matemática, el espacio muestral debe ser lo bastante rico como para que $\mathcal{F}$ contenga todos los sucesos de interés. Los conjuntos en $\mathcal{F}$ se denominan eventos o sucesos. Cuanto mayor sea $\mathcal{F}$, "más información" puede contener.

Principio de inclusión-exclusión:

a) Dos eventos: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

b) Tres: $P(A \cup B \cup C) =$
$P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$.

c) n: $P(\cup_{i=1}^{n} A_i) = \sum_{k=1}^{n} \sum_{I\subset\{1,\ldots,n\},|I|=k} (-1)^{k-1} P(\cap_{i\in I} A_i)$.

Ejemplo 1: **Modelo clásico o de Laplace**

Dado un espacio muestral finito $\Omega = \{\omega_1, \ldots, \omega_n\}$, la aplicación:

$$P(A) = \frac{\text{card}(A)}{n}, \quad A \subseteq \Omega,$$

es una medida de probabilidad sobre $(\Omega, \mathcal{P}(\Omega))$.

En este ejemplo, cada elemento ω_i $(i = 1, \ldots, n)$ tiene la misma probabilidad $(P(\{\omega_i\}) = 1/n,\ i = 1, \ldots, n)$. Este ejemplo es una instancia de la **probabilidad uniforme**.

Ejemplo 2: **Medidas discretas** (podemos tomar todos los conjuntos como medibles.)

Sea $\{\omega_1, \omega_2, \ldots\} \subset \Omega$ numerable, sea $\mathcal{F} = \mathcal{P}(\Omega)$ y sea $P(\{w_i\}) \in [0, 1]$ tal que $\sum_i P(\{\omega_i\}) = 1$. Entonces $(\Omega, \mathcal{F}, P)$ es un espacio de probabilidad.

Ejemplo 3: **Medidas definidas mediante densidades con respecto a la medida de Lebesgue.**

Sea $\Omega = I$ (un intervalo de $\mathbb{R}$). Una función $f : \mathbb{R} \longrightarrow \mathbb{R}^+$ $(f \geq 0)$ se llama **función de densidad** sobre I si cumple:

(a) f es integrable.

(b) $\int_I f(t)\, dt = 1$.

Tomando como σ-álgebra $\mathcal{F} = \mathcal{B}(I)$, podemos definir

$$P(A) = \int_A f(x)\, dx, \quad A \in \mathcal{F}.$$

Entonces $(\Omega, \mathcal{F}, P)$ es un espacio de probabilidad.

Ejercicio: Buscar en la literatura el Teorema Fundamental del Cálculo para la integral de Lebesgue, y entender lo que dice.

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad y A, $B \in \mathcal{F}$ con $P(B) > 0$. Se llama **probabilidad de A condicionada a B** a:

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

Esta expresión cuantifica en que medida la nueva información disponible (se ha dado B) modifica la probabilidad de A, es decir, nos dice como A depende de B.

¿Qué debería ocurrir con $P(A|B)$ en relación a $P(A)$ si A es independiente de B?

Definición: Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad. Dos sucesos $A, B \in \mathcal{F}$ se dicen **independientes** si $P(A \cap B) = P(A)P(B)$.

Notación: a veces se usa $AB = A \cap B$.

Observaciones:

- Si $P(A) = 0$, entonces A y B son independientes para todo $B \in \mathcal{F}$.
- Si $P(A) = 1$, entonces A y B son independientes para todo $B \in \mathcal{F}$.
- Si A y B son independientes, entonces también lo son A y B^c, A^c y B, y A^c y B^c.

N. B.: la aplicación $P(\cdot|B) : \mathcal{F} \longrightarrow [0,1]$ dada por $A \longmapsto P(A|B)$ es una medida de probabilidad sobre $(\Omega, \mathcal{F})$. Por tanto, $P(\cdot|B)$ verifica los axiomas y propiedades de una probabilidad.

Sea X una v.a. y sea $E(X)$ su media. Sabiendo que se ha dado B, ¿cómo deberíamos definir $E(X|B)$? Consideremos el caso especial $X = \mathbf{1}_A$.

Fórmulas asociadas a la probabilidad condicional

Fórmula del producto

Si $A_1, \ldots, A_n \in \mathcal{F}$ con $P(A_1 \cap \cdots \cap A_n) > 0$, se tiene

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

Fórmulas asociadas a la probabilidad condicional

Por brevedad usamos el siguiente abuso de notación: para denotar que todo conjunto en la sucesión $\{A_i\}_{i=1}^{\infty}$ pertenece a $\mathcal{F}$, escribimos $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$.

Fórmula de la probabilidad total

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, y sea $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$ una partición de Ω tal que $P(A_i) > 0$ para $i \geq 1$. Entonces para cualquier $B \in \mathcal{F}$,

$$P(B) = \sum_{i=1}^{\infty} P(B \cap A_i) = \sum_{i=1}^{\infty} P(A_i) P(B|A_i).$$

Fórmula de Bayes

Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad, y sea $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$ una partición de Ω tal que $P(A_i) > 0$ cuando $i \geq 1$. Para todo $B \in \mathcal{F}$ con $P(B) > 0$, tenemos

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

Dem: usar la definición de la probabilidad condicional y la regla de la probabilidad total.

Dado el espacio de probabilidad $(\Omega, \mathcal{F}, P)$, una **variable aleatoria** es una función medible $X : \Omega \longrightarrow \mathbb{K}$, donde la σ-álgebra en el espacio de llegada son los conjuntos de Borel $\mathcal{B}$ de $\mathbb{K}$. Es decir, para todo $B \in \mathcal{B}$, se tiene que $X^{-1}(B) \in \mathcal{F}$.

Notación: usamos las siguientes abreviaciones:

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

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

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

Como acabamos de indicar, una variable aleatoria $X : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}$ ó $\mathbb{C}$ es una función $\mathcal{F} - \mathcal{B}$ medible, donde $\mathcal{B}$ denota los conjuntos de Borel. Se define la **media o esperanza** de X como

$$EX = \int_\Omega X \, dP.$$

Observación: La identificación de la esperanza con la integral de Lebesgue (respecto de la medida de probabilidad subyacente) unifica las definiciones de la esperanza que se muestran en los cursos elementales.

Dada la v.a. $X : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{K}, \mathcal{B})$, la **ley de X** es la probabilidad imagen $P_X : \mathcal{B} \longrightarrow [0,1]$ definida como $P_X(B) = P(X^{-1}(B)) = P(X \in B)$.

Mientras que P está definida en subconjuntos de Ω, la ley P_X lo está en subconjuntos de $\mathbb{K}$: es una medida de probabilidad en $(\mathbb{K}, \mathcal{B})$.

Ello nos permite trabajar sobre $\mathbb{K}$ y olvidarnos de Ω.

Teorema: fórmula del cambio de variables

Dada la v.a. $X : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{K}$, para toda función Borel medible $g : \mathbb{K} \to \mathbb{K}$ se tiene que

$$\mathrm{E}(g(X)) := \int_{\Omega} g(X(\omega))\, dP(\omega) = \int_{\mathbb{K}} g(x)\, dP_X(x).$$

Dem: por reducciones sucesivas, basta probarlo para $g = \mathbf{1}_B$, donde B es Borel. Pero esto es obvio.

Ejemplo: Sea $X : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}$ una v.a. con distribución N(0,1). Hallar $E(X)$ y $E(X^2)$.

Por simplicidad, en lo sucesivo consideramos el caso $\mathbb{K} = \mathbb{R}$, salvo que explicitamente digamos otra cosa.

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

Notación: $X =_d Y$.

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

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

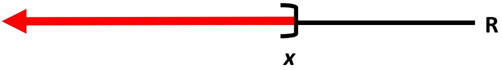
Proposición: Si $X =_{cs} Y$, entonces $X =_d Y$.

Observación: La igualdad en distribución $X =_d Y$ ni siquiera requiere que X e Y estén definidas en un mismo espacio de probabilidad. Incluso si lo están, es dedir, si $X, Y : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}$, y además $X =_d Y$, ello *no* implica que $X =_{cs} Y$.

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

Definición: Sea $X : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}$ una variable aleatoria. Se llama **función de distribución de** X a la función $F_X : \mathbb{R} \longrightarrow [0,1]$ dada por

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



Comentario: F_X contiene la misma información que P_X, pero expresada de manera más sucinta.

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

Teorema: Propiedades básicas de la función de distribución

Sea X variable aleatoria y F su función de distribución.

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

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

Recordatorio: Si $f : \mathbb{R} \longrightarrow \mathbb{R}$ es monótona, entonces el conjunto de puntos de discontinuidad de f es numerable. Por tanto, su complementario es denso.

Tres tipos puros de distribuciones

1. Discretas
2. Absolutamente continuas
3. Continuas singulares

Definición: Una variable aleatoria tiene **distribución discreta** F_d si existe un conjunto numerable (incluyendo el caso finito) no vacio $E \subset \mathbb{R}$ tal que para toda $x \in E$, $P(X = x) > 0$, y $P(E^c = 0)$. En tal caso decimos que la v.a. es disctreta.

Ojo: en principio no hay ninguna topología definida en Ω.

Ejemplo: Sea $X_n = 1/n$. Entonces $P_{X_n} = \delta_{1/n}$. Suponiendo que las v. a.'s X_n están definidas sobre el mismo Ω, ¿a qué convergen? ¿A qué convergen las funciones de distribución F_{X_n}?

Definición: Decimos que $X_n \to X$ en distribución si para todo punto de continuidad t de F_X tenemos que

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t).$$

Comentario: Esta noción de convergencia es muy débil, y tiene sentido incluso cuando las variables aleatorias están definidas en espacios distintos.

Ejemplo anterior: $X_n = 1/n$. Entonces $X_n \to 0$ en distribución. Notesé que

$$\lim_{n\to\infty} F_{X_n}(0) = 0 < F_X(0) = 1.$$

Puede demostrarse que una función continua y monótona $F : \mathbb{R} \longrightarrow [0, 1]$ es absolutamente continua si y solo si la imagen $F(A)$ de cualquier conjunto A de medida de Lebesgue 0, tiene medida de Lebesgue 0.

"Recordatorio": Las funciones F absolutamente continuas son precisamente aquellas para las que el TFC se cumple, es decir, F es derivable en casi todo punto, y puede recuperarse integrando su derivada.

Definición: Una variable aleatoria tiene **distribución absolutamente continua** F_{ac} si F_{ac} es una función absolutamente continua. En tal caso decimos que la v.a. es *continua*, con densidad $f = F'_{ac}$.

Ojo: en principio no hay ninguna topología definida en Ω. Decir que una v. a. es continua solo significa que tiene densidad.

Definición: Una variable aleatoria tiene **distribución continua y singular** F_{cs} si la función F_{cs} es continua y $F'_{cs} = 0$ en casi todo punto.

Ejemplo: la función "escalera" de Cantor (extendida del modo natural a toda la recta).

Puede demostrarse:

Teorema de descomposición

Toda función de distribución es combinación convexa de los tres tipos de distribuciones vistos anteriormente. En otras palabras, para toda función de distribución F, existen constantes

$$a, b, c \geq 0 \quad \text{con} \quad a + b + c = 1$$

y distribuciones puras

$$F_d, F_{ac}, \quad \text{y} \quad F_{cs},$$

tales que la función F se descompone de modo único como

$$F = aF_d + bF_{ac} + cF_{cs}.$$

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

Dado $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$, la **región suroeste generada por x** es $R_{\mathbf{x}} = (-\infty, x_1] \times \cdots \times (-\infty, x_k]$.

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

Proposición: La aplicación $\mathbf{X} : \Omega \longrightarrow \mathbb{R}^k$ dada por $\mathbf{X} = (X_1, \ldots, X_k)$ es un vector aleatorio si y solo si las funciones $X_1, \ldots, X_k$ son variables aleatorias.

Corolario: Sean $X_1, \ldots, X_k$ v.a. sobre $(\Omega, \mathcal{F})$ con valores en $\mathbb{R}$ y sea $g : \mathbb{R}^k \longrightarrow \mathbb{R}$ medible Borel. Entonces $X = g(X_1, \ldots, X_k)$ es una v.a.

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

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

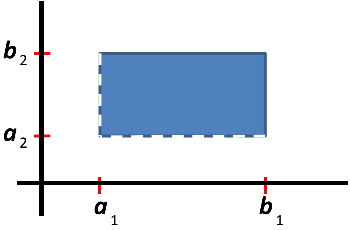
Teorema: Propiedades básicas de la función de distribución

Sea $\mathbf{X}$ un vector aleatorio y sea $F_{\mathbf{X}}$ su función de distribución.

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

Pregunta: ¿Qué significa que una función sea monótona (no decreciente) en $\mathbb{R}^k$? Unicamente ilustramos el caso $k = 2$.

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



Sea $\mathbf{X} = (X_1, \ldots, X_k) : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}^k$ un vector aleatorio. La ley de $\mathbf{X}$ es

$$P_{\mathbf{X}} : \mathcal{B}(\mathbb{R}^k) \longrightarrow [0, 1],$$

$$P_{\mathbf{X}}(B) = P(\mathbf{X}^{-1}(B)) = P(\mathbf{X} \in B).$$

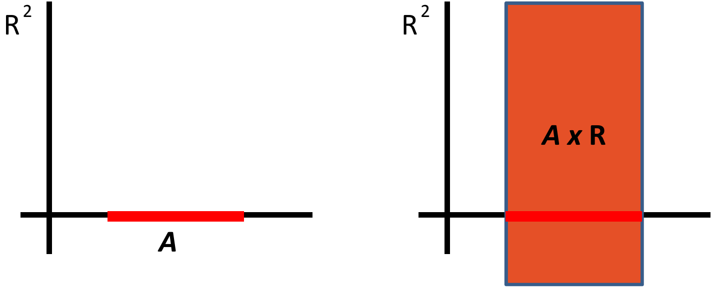
La ley $P_{\mathbf{X}}$ ó $P_{(X_1,\ldots,X_k)}$ también se denomina **distribución de probabilidad** de $\mathbf{X}$ o **distribución conjunta** de las variables $X_1, \ldots, X_k$, y es una medida de probabilidad en $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$.

A las leyes $P_{X_1}, \ldots, P_{X_k}$ de las variables aleatorias $X_1, \ldots, X_k$, se las llama **distribuciones marginales** de $\mathbf{X}$.

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

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

Ilustración en el caso $n = 2$, $P(X \in A) = P((X, Y) \in A \times \mathbb{R})$.



La "masa" o probabilidad 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 I un conjunto de indices (no necesariamente numerable) y sea $\{A_i\}_{i\in I}$ una sucesión de eventos en $\mathcal{F}$. Se dice que los sucesos $\{A_i\}_{i\in I}$ son **(mutuamente) independientes** si

$$\forall\{A_{i_1},\ldots,A_{i_n}\}\subset\{A_i\}_{i\in I},\ P(A_{i_1}\cap\cdots\cap A_{i_n})=P(A_{i_1})\cdots P(A_{i_n}).$$

Observación: $A, B, C\in\mathcal{F}$ son independientes si:

$(*)$

$$\left\{\begin{array}{c} P(A\cap B)=P(A)P(B)\\ P(A\cap C)=P(A)P(C)\\ P(B\cap C)=P(B)P(C)\end{array}\right\}\quad \{A,B,C\} \text{ son indep. dos a dos}$$

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

Proposición: $\{A_i\}_{i\in I}$ independientes $\Longrightarrow$ $\{A_i^c\}_{i\in I}$ independientes.

Definición: Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidad.
Un familia $\{\mathcal{C}_i\}_{i \in I}$ de clases (no vacías) de sucesos en $\mathcal{F}$ se dice que es una **familia de clases independientes** si para cada elección $\{C_i : C_i \in \mathcal{C}_i,\ i \in I\}$, los sucesos $\{C_i\}_{i \in I}$ son independientes. Es decir, si para cada subcolección finita $\{i_1, \ldots, i_k\} \subset I$, se verifica $P(C_{i_1} \cap \cdots \cap C_{i_k}) = P(C_{i_1}) \cdots P(C_{i_k})$.

Comentario: Comprobar la independencia de σ-álgebras es complicado. Recurrimos a colecciones que encajan mejor con la noción de independencia.

Definición: Una colección $\mathcal{C}$ de $\mathcal{P}(\Omega)$ es un π-**sistema** si

(a) $\Omega \in \mathcal{C}$.

(b) $A, B \in \mathcal{C}$, entonces $A \cap B \in \mathcal{C}$
(estable bajo las intersecciones finitas).

Teorema fundamental de la independencia

Sean $\mathcal{C}_1, \ldots, \mathcal{C}_n \subset \mathcal{F}$ π-sistemas independientes. Entonces las σ-álgebras $\sigma(\mathcal{C}_1), \ldots, \sigma(\mathcal{C}_n)$ son independientes.

Sin dem.

Teorema fundamental de la independencia

Sean $\mathcal{C}_1, \ldots, \mathcal{C}_n \subset \mathcal{F}$ π-sistemas independientes. Entonces las σ-álgebras $\sigma(\mathcal{C}_1), \ldots, \sigma(\mathcal{C}_n)$ son independientes.

Idea de por qué el resultado es creible: Los conjuntos en $\sigma(\mathcal{C}_i)$ pueden aproximarse con precisión arbitraria por conjuntos de $\mathcal{C}_i \cup \{A^c : A \in \mathcal{C}_i\}$ (demostración: los conjuntos aproximables con precisión arbitraria forman una clase que contiene a $\mathcal{C}_i$, está cerrada bajo complementación y bajo uniones numerables).

Ejemplo: Sean A, B, C, D, y E sucesos independientes. Entonces las σ-álgebras $\mathcal{F}_1 = \{\emptyset, A, A^c, \Omega\}, \ldots, \mathcal{F}_5 = \{\emptyset, E, E^c, \Omega\}$ son independientes.

Las σ-álgebras $\sigma(\mathcal{F}_1 \cup \mathcal{F}_2)$ y $\sigma(\mathcal{F}_3 \cup \mathcal{F}_4 \cup \mathcal{F}_5)$ ¿son independientes?
Por ejemplo, los sucesos $A \cap B^c$ y $C \cup (D \setminus E)$ ¿son independientes?

Corolario: Teorema de agrupamiento

Sea $\{\mathcal{F}_i\}_{i\in I}$ familia independiente de σ-álgebras de $\mathcal{F}$. Para toda partición $\{I_j\}_{j\in J}$ de I, definiendo $\mathcal{F}_{I_j} = \sigma\left(\bigcup_{i\in I_j} \mathcal{F}_i\right)$, se tiene que $\{\mathcal{F}_{I_j}\}_{j\in J}$ es una familia de σ-álgebras independientes.

Dem: fácil.

Límites inferior y superior de sucesiones de funciones.

Sea $\{f_n\}_{n\geq 1}$ con $f_n : \Omega \longrightarrow \overline{\mathbb{R}}$. Se define la función **límite superior** de $\{f_n\}$ como

$$\left(\limsup_{n\to\infty} f_n\right)(\omega) = \limsup_{n\to\infty} f_n(\omega), \quad \omega \in \Omega.$$

Análogamente, el **límite inferior** de $\{f_n\}$ es

$$\left(\liminf_{n\to\infty} f_n\right)(\omega) = \liminf_{n\to\infty} f_n(\omega), \quad \omega \in \Omega.$$

Recordatorio: los límites superior e inferior de funciones medibles son medibles ¿Cuál es la manera estándar de probar que $\lim_{n\to\infty} f_n$ existe c.s.?

Límites inferior y superior de sucesiones de conjuntos

Dado Ω, su conjunto de partes $\mathcal{P}(\Omega)$ está parcialmente ordenado por inclusión, es decir, $A \leq B$ si y solo si $A \subset B$. El ínfimo y el supremo de cualquier collección $\{A_i\}_{i \in I} \subset \mathcal{P}(\Omega)$ están bien definidos:

$$\sup_{i \in I} A_i = \bigcup_{i \in I} A_i \quad \text{e} \quad \inf_{i \in I} A_i = \bigcap_{i \in I} A_i.$$

Límites inferior y superior de sucesiones de conjuntos

Recordatorio: Se define el **límite superior** de una sucesión de números reales $\{a_n\}_{n\geq 1}$ como

$$\limsup_{n\to\infty} a_n = \inf_{n\geq 1}\left(\sup_{k\geq n} a_k\right) = \lim_{n\to\infty}\left(\sup_{k\geq n} a_k\right).$$

Análogamente, para toda sucesión de conjuntos $\{A_n\}_{n\geq 1}$ se define

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1}\left(\bigcup_{k\geq n} A_k\right) = \inf_{n\geq 1}\left(\sup_{k\geq n} A_k\right) = \lim_{n\to\infty}\left(\sup_{k\geq n} A_k\right).$$

Del mismo modo, el límite inferior de una sucesión de números reales $\{a_n\}_{n\geq 1}$ es

$$\liminf_{n\to\infty} a_n = \sup_{n\geq 1}\left(\inf_{k\geq n} a_k\right) = \lim_{n\to\infty}\left(\inf_{k\geq n} a_k\right),$$

y el límite inferior de una sucesión de conjuntos $\{A_n\}_{n\geq 1}$,

$$\liminf_{n\to\infty} A_n = \bigcup_{n\geq 1}\left(\bigcap_{k\geq n} A_k\right) = \sup_{n\geq 1}\left(\inf_{k\geq n} A_k\right) = \lim_{n\geq 1}\left(\inf_{k\geq n} A_k\right).$$

Identificando conjuntos con sus funciones indicatrices, tenemos dos formas de definir el límite superior y el límite inferior de una sucesión de conjuntos $\{A_n\}_{n\geq 1}$, directamente y por medio de los correspondientes límites de funciones.

¿Cuál es la relación entre ambas definiciones?

Caracterización puntual:

• $\omega \in \operatorname{lím\,sup} A_n \iff \omega \in A_n$ para infinitos valores de n.
Notación: $\omega \in A_n$ i. o.

• $\omega \in \operatorname{lím\,inf} A_n \iff \exists n_0$ tal que $\omega \in A_n$ para todo $n \geq n_0$; es decir, $\omega \in A_n$ eventualmente, o para "casi todo" n (para todo n salvo quizá un número finito de excepciones).

Propiedades:

- $\limsup A_n = (\liminf A_n^c)^c$, $\liminf A_n = (\limsup A_n^c)^c$.
- $\liminf A_n \subset \limsup A_n$.

Definición: Dada $\{A_n\}_{n\geq 1} \subset \mathcal{P}(\Omega)$, decimos que $A_n \to A$ si $\limsup A_n = \liminf A_n = A$.

Ejercicio: probar que si $A_n \uparrow A$ ó $A_n \downarrow A$, entonces $A_n \to A$ (recordatorio: $A_n \uparrow A$ significa que $A_n \subset A_{n+1}$ y $A = \cup_n A_n$, con la notación análoga para $A_n \downarrow A$).

Teorema: Sea $\{A_n\}_{n\geq 1} \subset \mathcal{F}$. Tenemos

$$P(\liminf A_n) \leq \liminf P(A_n) \leq \limsup P(A_n) \leq P(\limsup A_n).$$

Teorema: 1er Lema de Borel-Cantelli

Sea $\{A_n\}_{n\geq 1} \subset \mathcal{F}$. Si $\displaystyle\sum_{n=1}^{\infty} P(A_n) < \infty$, entonces $P(\limsup A_n) = 0$.

La demostración es trivial, pero el resultado es muy útil para probar convergencia c.s.

El siguiente teorema (no trivial) requiere independencia.

Teorema: 2ndo Lema de Borel-Cantelli

Sea $\{A_n\}_{n\geq 1} \subset \mathcal{F}$ una sucesión de eventos independientes. Si

$$\sum_{n=1}^{\infty} P(A_n) = \infty, \text{ entonces } P(\limsup A_n) = 1.$$

Dem: usar $e^{-x} \geq 1 - x$. Comentario: a menudo es más fácil trabajar con intersecciones que con uniones, especialmente si tenemos independencia.

Sea $\{\mathcal{F}_n\}_{n\geq 1}$ una sucesión de sub-σ-álgebras de $\mathcal{F}$. Consideramos:

$$\left\{\begin{array}{ccc} \mathcal{A}^1 = \sigma\left(\bigcup_{i\geq 1}\mathcal{F}_i\right) \\ \mathcal{A}^2 = \sigma\left(\bigcup_{i\geq 2}\mathcal{F}_i\right) \\ \vdots & & \vdots \\ \mathcal{A}^n = \sigma\left(\bigcup_{i\geq n}\mathcal{F}_i\right) \\ \vdots & & \vdots \end{array}\right\} \quad \mathcal{A}^1 \supseteq \mathcal{A}^2 \supseteq \mathcal{A}^3 \supseteq \cdots \supseteq \mathcal{A}^n \supseteq \cdots .$$

$$\mathcal{A}^\infty := \bigcap_{k=1}^{\infty} \mathcal{A}^k \quad \text{es una } \sigma\text{-álgebra}$$

Terminología: $\mathcal{A}^\infty$ se llama **σ-álgebra asintótica o terminal (relativa a $\{\mathcal{F}_n\}$)** (en inglés, tail σ-algebra). Si $A \in \mathcal{A}^\infty$, A se denomina **suceso asintótico**.

Comentario: para todo $n \geq 1$, tenemos que $\mathcal{A}^\infty = \bigcap_{k=n}^{\infty} \mathcal{A}^k$. Es decir, $\mathcal{A}^\infty$ no depende de lo que ocurra al principio: si reemplazamos $\mathcal{F}_1, \ldots, \mathcal{F}_k$ con $\mathcal{F}'_1, \ldots, \mathcal{F}'_k$, la σ-álgebra asintótica de $\mathcal{F}'_1, \ldots, \mathcal{F}'_k, \mathcal{F}_{k+1}, \ldots$ es la misma $\mathcal{A}^\infty$.

Ejemplos: Si $A_n \in \mathcal{F}_n$ para todo $n \in \mathbb{N}$, entonces

$$\limsup A_n \in \mathcal{A}^\infty \quad \text{y} \quad \liminf A_n \in \mathcal{A}^\infty.$$

Teorema: Ley 0-1 de Kolmogorov

Sea $\{\mathcal{F}_n\}_{n\geq 1} \subset \mathcal{F}$ una sucesión de sub-σ-álgebras independientes. Para todo $A \in \mathcal{A}^\infty$, se tiene $P(A) = 0$ ó 1.

Dem: probamos que $\mathcal{A}^\infty$ es independiente de $\mathcal{A}^\infty$, luego si $A \in \mathcal{A}^\infty$, entonces $P(A) = P(A \cap A) = P(A)P(A)$, y por tanto $P(A) = 0$ ó 1.
Nótese que para todo $n \geq 1$, $\mathcal{A}^\infty$ es independiente de $\sigma\left(\bigcup_{i=1}^n \mathcal{F}_i\right)$ por agrupación. Deducimos que $\mathcal{A}^\infty$ es independiente de $\mathcal{B} := \cup_{n=1}^\infty \sigma\left(\bigcup_{i=1}^n \mathcal{F}_i\right)$. Como $\mathcal{B}$ es un π-sistema, $\mathcal{A}^\infty$ es independiente de $\sigma(\mathcal{B})$, y a la vez $\mathcal{A}^\infty \subset \sigma(\mathcal{B}) = \mathcal{A}^1$.

Ejemplos: si $\{A_n\}_{n\geq 1}$ es una sucesión de eventos independientes, entonces $\limsup \mathbf{1}_{A_n}$ es constante c.s., y también lo es $\liminf \mathbf{1}_{A_n}$.

Sea $X : \Omega \longrightarrow \mathbb{R}$ una función. La **σ-álgebra generada** por X es

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

Observación: $\sigma(X)$ es la mínima σ-álgebra sobre Ω que hace medible a X.

Del mismo modo, dada la colección $\{X_i : i \in I\}$ 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.

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

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

generadas por ellas son independientes.

Observaciones:

1. $X_1, \ldots, X_k$ ind. $\Longleftrightarrow \forall A_1 \in \sigma(X_1), \ldots, \forall A_k \in \sigma(X_k)$, $P(A_1 \cap \cdots \cap A_k) = P(A_1) \cdots P(A_k)$.

2. $X_1, \ldots, X_k$ ind. $\Longleftrightarrow \forall B_1, \ldots, B_k \in \mathcal{B}$,

$$P_{(X_1,\ldots,X_k)}(B_1 \times \cdots \times B_k) = P_{X_1}(B_1) \cdots P_{X_k}(B_k).$$

Independencia = Medida producto.

Sea $P_{(X_1,\ldots,X_k)}$ la medida imagen o ley del vector aleatorio $(X_1,\ldots,X_k)$.
$P_{(X_1,\ldots,X_k)}$ es una probabilidad en $(\mathbb{R}^k,\mathcal{B}^k)$, mientras que $P_{X_1},\ldots,P_{X_k}$ son probabilidades en $(\mathbb{R},\mathcal{B})$.

Las v.a's $X_1,\ldots,X_k$ son independientes $\Leftrightarrow$ $P_{(X_1,\ldots,X_k)}$ es la medida producto de $P_{X_1},\ldots,P_{X_k}$.

$$P_{(X_1,\ldots,X_k)} = P_{X_1}\otimes\cdots\otimes P_{X_k}.$$

1. Sean $X_1, \ldots, X_k$ v.a. independientes y sean $f_i : \mathbb{R} \longrightarrow \mathbb{R}$ medibles Borel ($i = 1, \ldots, k$). Entonces $f_1(X_1), \ldots, f_k(X_k)$ son independientes.

   Dem: considerar las σ-álgebras correspondientes.

2. Sean $X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+m}$ v.a. independientes y sean $f_1 : \mathbb{R}^n \longrightarrow \mathbb{R}$, $f_2 : \mathbb{R}^m \longrightarrow \mathbb{R}$ medibles. Entonces $f_1(X_1, \ldots, X_n)$ y $f_2(X_{n+1}, \ldots, X_{n+m})$ son independientes.

   Dem: Por agrupación y el resultado anterior.

Independencia de variables aleatorias

Teorema: Caracterización mediante funciones de distribución

Sean $X_1, \ldots, X_k$ v.a.'s con función de distribución conjunta F y funciones de distribución marginales $F_1, \ldots, F_k$. Entonces

$$X_1, \ldots, X_k \text{ son indep.} \iff F(x_1, \ldots, x_k) = F_1(x_1) \cdots F_k(X_k).$$

Teorema: Independencia de variables aleatorias continuas

Sean $X_1, \ldots, X_k$ v.a.'s continuas, con densidades $f_i(x_i)$ para $i = 1, \ldots, k$. Entonces las v.a.'s $X_1, \ldots, X_k$ son independientes $\iff$ el vector aleatorio $(X_1, \ldots, X_k)$ tiene densidad conjunta $f(x_1, \ldots, x_k) = f_1(x_1) \cdots f_k(x_k)$.

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

Sean X e $Y : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}$ v.a. independientes e integrables. Entonces la v.a. XY es integrable y $E(XY) = EXEY$.

Dem: por reducción al caso de las funciones indicatrices, para el cual es obvio. **OJO** con la independencia en cada reducción.

Comentario: comparar con Hölder.

Corolario: Sean $X_1, \ldots, X_n : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}$ v.a. independientes e integrables. Entonces la v.a. $X_1 \cdots X_n$ es integrable y $E(X_1 \cdots X_n) = EX_1 \cdots EX_n$.

Corolario: Sean $X_1, \ldots, X_n : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}$ v.a. independientes, y sean $f_1, \ldots, f_n : \mathbb{R} \longrightarrow \mathbb{R}$ funciones de Borel tales que las variables $f_i(X_i)$ son integrables. Entonces la v.a. $f_1(X_1) \cdots f_n(X_n)$ es integrable y $E(f_1(X_1) \cdots f_n(X_n)) = Ef_1(X_1) \cdots Ef_n(X_n)$. En particular, ello sucede si las funciones de Borel son acotadas.

Observación: Puede ocurrir que $E(XY) = EXEY$, pero X e Y *no* sean independientes. Por ejemplo, si X e Y no son 0 c.s., tienen soportes disjuntos, y sus medias son cero.

Teorema: Esperanza e independencia

Son equivalentes:

1. X e $Y : (\Omega, \mathcal{F}, P) \longrightarrow \mathbb{R}$ son v.a. independientes.
2. Para todo par $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ de funciones de Borel acotadas, $E(f(X)g(Y)) = Ef(X)Eg(Y)$.
3. Para todo par $f, g : \mathbb{R} \longrightarrow [0, \infty)$ de funciones de Borel, $E(f(X)g(Y)) = Ef(X)Eg(Y)$.
4. Para todo par $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ de funciones de Borel tales que $f(X), g(Y) \in L^1(P)$, $E(f(X)g(Y)) = Ef(X)Eg(Y)$.

Dem: 1) $\Longrightarrow$ 2), visto. 2) $\Longrightarrow$ 3), por aproximación desde abajo. 3) $\Longrightarrow$ 4), por descomposición en partes positivas y negativas de f y g. 4) $\Longrightarrow$ 1), dados $A, B \in \mathcal{B}(\mathbb{R})$, tomamos $f = \mathbf{1}_A$ y $g = \mathbf{1}_B$.

Definición: Dada una variable $X \in \mathcal{L}_2$, se llama **varianza** de X a

$$\sigma_X^2 = Var(X) := E[(X - E(X))^2] = E(X^2) - [E(X)]^2.$$

Definición: La **desviación típica** de X es

$$\sigma_X = \sqrt{Var(X)} = \|X - E(X)\|_2.$$

Definición: Si $X, Y, XY \in \mathcal{L}_1$, se llama **covarianza** de las v.a.'s X e Y a

$$Cov(X, Y) = E\left[(X - EX)(Y - EY)\right] = E(XY) - E(X)E(Y).$$

Recordemos que
$Cov(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - E(X)E(Y)$.

Observaciones:

1. Si $X, Y \in \mathcal{L}_2$, entonces por la desigualdad de Cauchy-Schwarz tenemos
$$|Cov(X, Y)| \leq \sigma_X \sigma_Y < \infty.$$
2. $Cov(X, X) = Var(X) \quad \text{y} \quad Cov(Y, X) = Cov(X, Y)$.
3. $Cov(aX, bY) = abCov(X, Y)$.
4. Si $X, Y \in \mathcal{L}_1$ son independientes, entonces $Cov(X, Y) = 0$.

Definición: Sean $X, Y, XY \in \mathcal{L}_1$. Se dice que X e Y están **incorreladas** (o **incorrelacionadas**) si $Cov(X, Y) = 0$, es decir, si $E(XY) = EXEY$.

Sabemos que si X e Y son independientes, entonces X e Y son incorreladas. Sin embargo, X, Y incorreladas $\not\Rightarrow$ X, Y independientes (ejemplo con soportes disjuntos y medias 0).

Observaciones:

1. $Var(aX + b) = a^2 Var(X)$.
2. Varianza de una suma de v.a.: Si $X_1, \dots, X_n \in \mathcal{L}_2$, entonces
$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} Var(X_i) + 2 \sum_{1 \le i < j \le n} Cov(X_i, X_j).$$
3. Si $X_1, \dots, X_n$ son incorreladas (en particular si son independientes dos a dos) entonces
$$Var(X_1 + \cdots + X_n) = \sum_{i=1}^{n} Var(X_i).$$
4. Si para todo par $i \neq j$ tenemos $Cov(X_i, X_j) \le 0$, entonces
$$Var(X_1 + \cdots + X_n) \le \sum_{i=1}^{n} Var(X_i).$$

Definición: Dadas X, Y v.a. no degeneradas, el **coeficiente de correlación (lineal de Pearson)** se define mediante

$$\rho_{X,Y} = \frac{\text{Cov(X,Y)}}{\sigma_X \sigma_Y} \quad \text{(es adimensional).}$$

Nota: $\rho_{X,Y} \in [-1, 1]$ (Cauchy-Schwarz).

Observación: Análisis de los extremos

1. $\rho_{X,Y} = 0 \iff X, \ Y$ incorreladas.
2. $\rho_{X,Y} = 1 \iff Y = aX + b$ c.s. con $a > 0, \ b \in \mathbb{R}$.
3. $\rho_{X,Y} = -1 \iff Y = aX + b$ c.s. con $a < 0, \ b \in \mathbb{R}$.

1. **Distribución degenerada en un punto** a

$$P(X = a) = 1.$$

2. **Distribución de Bernoulli de parámetro** $p \in (0, 1)$

$$P(X = 0) = q = 1 - p, \quad P(X = 1) = p.$$

El valor 1 se identifica con el éxito y 0 con el fracaso. Es la base de la distribución binomial.
Notación: $X \sim \mathrm{B}(1; p)$.

3. **Distribución uniforme en** n **puntos** $a_1, \ldots, a_n$

$$P(X = a_i) = \frac{1}{n}, \quad i = 1, \ldots, n.$$

Notación: $X \sim \mathrm{U}(\{a_1, \ldots, a_n\})$.

4. **Distribución binomial de parámetros n, p**

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n.$$

La variable aleatoria

$$X \equiv \text{ número de éxitos en los } n \text{ intentos}$$

se dice que tiene distribución **binomial** (de parámetros n y p) y toma los valores $0, 1, \ldots, n$.

Notación: $X \sim \mathrm{B}(n; p)$.

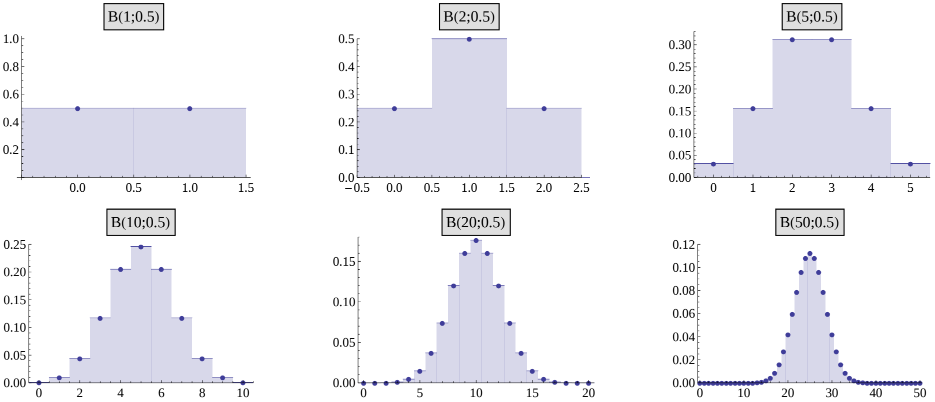
Suma de pruebas de Bernoulli independientes: Consideremos $X_1, \ldots, X_n$ v.a. $B(1, p)$ independientes, donde X_i toma el valor 1 si en la i-ésima prueba de Bernoulli ocurre el suceso A (éxito). Tenemos,

$$X = X_1 + \cdots + X_n, \quad X_i \sim \mathrm{B}(1; p) \quad (i = 1, \ldots, n).$$

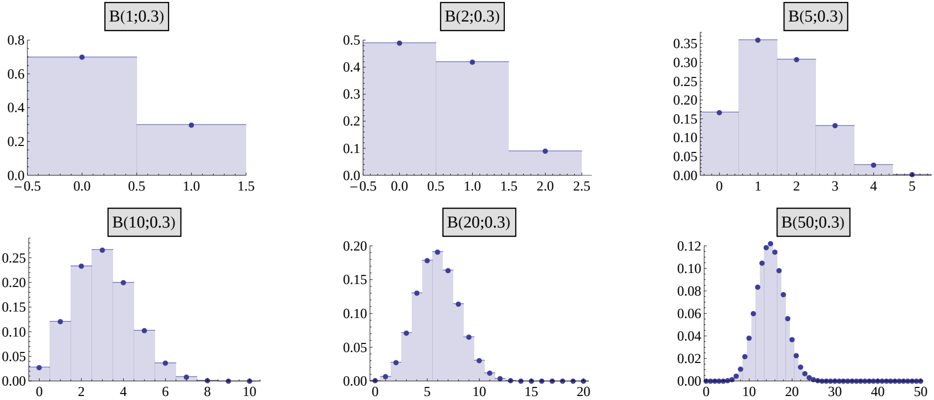
Luego, la v.a. con distribución binomial de parámetros n y p se puede expresar como suma de n v.a.s (independientes) de Bernoulli de parámetro p.

Suma de binomiales independientes:
Si $X_1, \ldots, X_k$ v.a. (independientes) con $X_i \sim \mathrm{B}(n_i; p)$ $(1 \leq i \leq n)$, la v.a. $X = X_1 + \cdots + X_k \sim \mathrm{B}(n; p)$, con $n = n_1 + \cdots + n_k$.

Función de probabilidad de B(n; 0,5)



Función de probabilidad de B$(n; 0{,}3)$

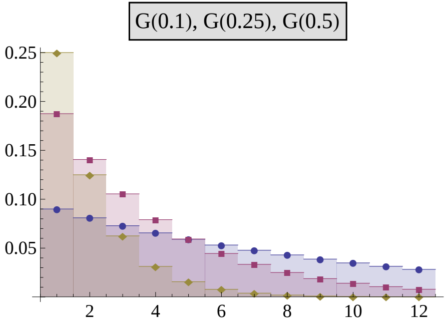


5. **Distribución geométrica de parámetro** $p \in (0,1)$

$$P(X=k) = pq^k, \quad k = 0, 1, \ldots$$

Esta distribución aparece al contar el número de fracasos antes de obtener el primer exito, con probabilidad p, cuando los intentos son independientes.

Notación: $X \sim \mathrm{G}(p)$.



5. **Distribución geométrica de parámetro** $p \in (0,1)$

Nota: En algunas ocasiones interesan conocer la distribución de la variable Y que cuenta el número de intentos hasta el primer éxito. Los intentos se suponen independientes unos de otros y la probabilidad de éxito es p. En este caso, la variable de interés es $Y = X + 1$, donde $X \sim \mathrm{G}(p)$, es decir, Y verifica

$$P(Y = k) = pq^{k-1}, \quad k = 1, 2, \ldots$$

Una variable aleatoria discreta $X \geq 0$ se dice que **no tiene memoria** si

$$P(X \geq n + m | X \geq m) = P(X \geq n) \quad \text{para todo } m, n \geq 0.$$

Observación: la distribución geométrica no tiene memoria.

7. **Distribución de Poisson de parámetro** $\lambda > 0$

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots$$

Notación: $X \sim \mathrm{P}(\lambda)$.

Recordatorio: si $X \sim \mathrm{P}(\lambda)$, entonces $E(X) = Var(X) = \lambda$.

Aproximación de Poisson a la binomial:

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!}, \quad n \to \infty,\ np \to \lambda.$$

1. **Distribución uniforme en el intervalo** (a, b)

$$f(x) = \frac{1}{b-a}\mathbf{1}_{(a,b)}(x).$$

Terminología: decimos que la variable U es uniforme estándar si $a = 0$ y $b = 1$.

Propiedad: generación de números aleatorios

Sea X variable aleatoria con función de distribución F continua. La variable $U = F(X)$ tiene distribución uniforme estándar.

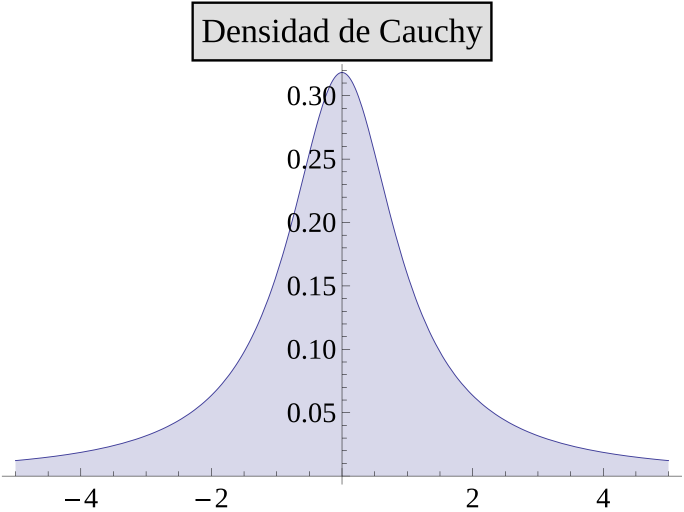
Notación: $X \sim \mathrm{U}(a, b)$.

3. **Distribución de Cauchy**

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad x \in \mathbb{R}.$$

Aplicaciones: En el estudio de emisiones de partículas. Si Z es un ángulo aleatorio distribuido uniformemente entre $-\pi/2$ y $\pi/2$, $\mathrm{tg}(Z)$ tiene distribución de Cauchy. El cociente de dos v.a. normales estándar independientes tiene también distribución de Cauchy.

Observación: La distribución de Cauchy no tiene media finita.


Densidad de Cauchy
0.30
0.25
0.20
0.15
0.10
0.05
−4
−2
2
4

4. **Distribución exponencial de parámetro** $\lambda > 0$

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

Aplicaciones: La distribución exponencial se utiliza para modelar el tiempo transcurrido hasta la ocurrencia por primera vez de un suceso. Es la versión continua de la distribución geométrica.

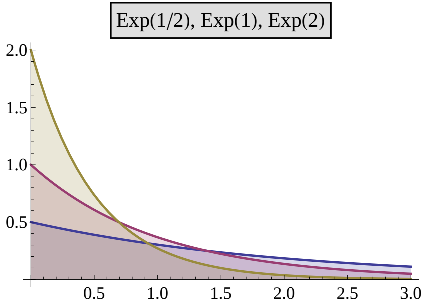
Notación: $X \sim \text{Exp}(\lambda)$.

Una v. a. continua $X \geq 0$ **no tiene memoria** si para todo par de números reales $s, t > 0$, $P(X > s + t | X > s) = P(X > t)$.

Propiedad: falta de memoria de la exponencial

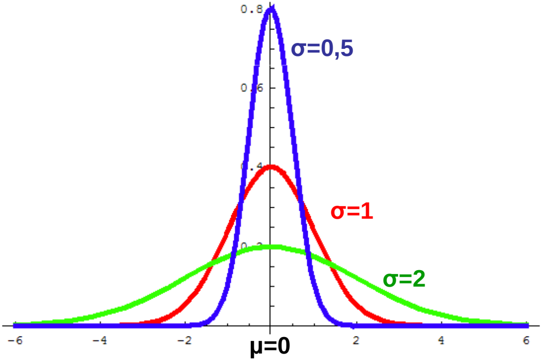
Sea $X \geq 0$ una v.a. continua. Son equivalentes:

1. X tiene distribución exponencial.
2. X no tiene memoria.



2. **Distribución normal de parámetros** $\mu \in \mathbb{R}, \sigma > 0$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x \in \mathbb{R}.$$



Notación: $X \sim \mathrm{N}(\mu; \sigma)$.

El vector aleatorio $\mathbf{X}$ es **normal** (d-dimensional) con parámetros $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ (notación: $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$), donde $\boldsymbol{\mu}$ es el vector de las medias de $X_1, \ldots, X_d$ y $\boldsymbol{\Sigma}$ la matriz de covarianzas) si tiene densidad dada por:

$$f(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1/2}(2\pi)^{-d/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}.$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{p2} & \cdots & \sigma_{dd} \end{bmatrix}.$$

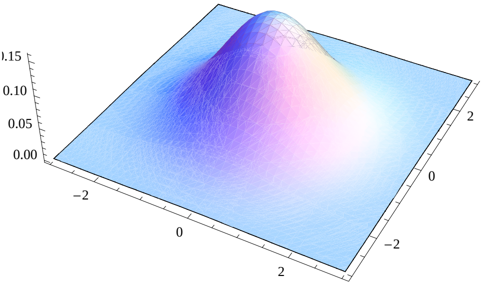
Nota: Se puede comprobar que las variables marginales también son normales. De hecho, $X_i \sim \mathrm{N}(\mu_i, \sigma_{ii})$.

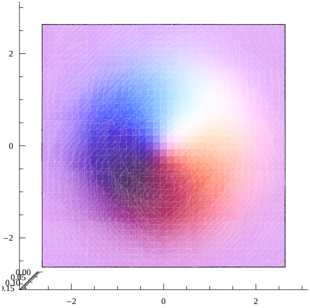
Teorema: v.a.'s incorreladas conjuntamente normales son independientes

Sea el vector aleatorio $\mathbf{X} = (X_1, \ldots, X_d)$ normal con distribución $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Entonces las v.a.'s $X_1, \ldots, X_d$ son independientes si y solo si son incorreladas.

Ejemplos de densidades normales en dimensión 2

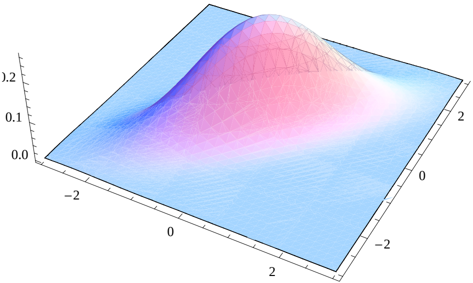
Densidad de $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $\boldsymbol{\mu} = (0,0)T$ y $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

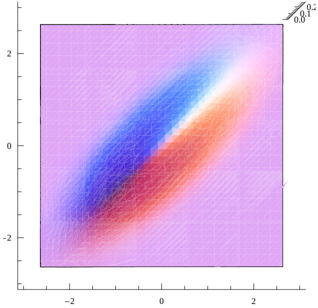




Ejemplos de densidades normales en dimensión 2

$$\text{Densidad de } \mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ con } \boldsymbol{\mu} = (0,0)^T \text{ y } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0{,}8 \\ 0{,}8 & 1 \end{pmatrix}.$$





A continuación definimos convergencia en casi todo punto (o casi por todo), en medida, y en L^p.

Definición: Sean $f, f_1, f_2, \dots$ funciones medibles definidas sobre el mismo espacio de medida $(\Omega, \mathcal{F}, \mu)$. Entonces:

$$f_n \xrightarrow{\text{c.p.t.}} f \quad \text{si} \quad \mu(\{\omega \in \Omega : f_n(\omega) \to f(\omega)\}^c) = 0.$$

$$f_n \xrightarrow{\mu} X \quad \text{si} \quad \forall \epsilon > 0,\ \mu(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| \geq \epsilon\}) \to 0.$$

$$f_n \xrightarrow{L^p} f \quad \text{si} \quad \|f_n - f\|_p \to 0 \qquad (p > 0).$$

Definición: Sean $X, X_1, X_2, \ldots$ variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Entonces:

$$X_n \xrightarrow{\text{c.s.}} X \quad \text{si} \quad P(X_n \to X) = 1.$$

$$X_n \xrightarrow{P} X \quad \text{si} \quad \forall \epsilon > 0,\ P(|X_n - X| \geq \epsilon) \to 0.$$

$$X_n \xrightarrow{\text{m-}p} X \quad \text{si} \quad E|X_n - X|^p \to 0 \qquad (p > 0).$$

Definición: Sean $X, X_1, X_2, \ldots$ variables aleatorias *no necesariamente definidas sobre el mismo espacio de probabilidad* y sean F y F_n las funciones de distribución de X y X_n, respectivamente. Entonces:

$$X_n \xrightarrow{\mathcal{D}} X \quad \text{si} \quad \forall x \in C_F,\ F_n(x) \to F(x),$$

donde C_F es el conjunto de puntos de continuidad de F.

Observación 1: Si indicamos por $\rightarrow$ uno cualquiera de los tres primeros modos de convergencia, es claro que $X_n \rightarrow X$ es equivalente a $X_n - X \rightarrow 0$.

Observación 2: Si $X \equiv a$ (constante), las definiciones de $X_n \xrightarrow{P} a$ y $X_n \xrightarrow{\text{m-}p} a$ tienen sentido también cuando las variables X_i *no* están definidas sobre el mismo espacio.

Observación 3: Cuando todas las variables involucradas son degeneradas, es decir constantes c.s., cada una de las cuatro convergencias coincide con la convergencia usual de sucesiones de números reales.

Lema: Son equivalentes:

(a) $X_n \xrightarrow{\text{c.s.}} X$.

(b) Para todo $\epsilon > 0$, $P\left(\limsup_{n\to\infty}\{|X_n - X| \geq \epsilon\}\right) = 0$.

(c) Para todo $\epsilon > 0$, $\lim_{n\to\infty} P\left(\bigcup_{k\geq n}\{|X_k - X| \geq \epsilon\}\right) = 0$.

Corolario: Se tiene:

$$\sum_{n\geq 1} P(|X_n - X| \geq \epsilon) < \infty, \quad \epsilon > 0 \quad \Longrightarrow \quad X_n \xrightarrow{\text{c.s.}} X.$$

Teorema: Si $X_n \xrightarrow{\text{c.s.}} X$, entonces $X_n \xrightarrow{P} X$.

Observación:

$$X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{\text{c.s.}} X.$$

$$X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{\text{c.s.}} X.$$

Teorema: Si $X_n \xrightarrow{P} X$, entonces $X_{n_k} \xrightarrow{\text{c.s.}} X$, para alguna subsucesión $\{X_{n_k}\}_{k\geq 1}$.

Convergencia en media-p versus convergencia en media-q
Convergencia en media-p vs convergencia en probabilidad

Teorema: Si $X_n \xrightarrow{\text{m-}q} X$, entonces $X_n \xrightarrow{\text{m-}p} X$, para $0 < p < q$.

Observación: $X_n \xrightarrow{\text{m-}p} X \not\Rightarrow X_n \xrightarrow{\text{m-}q} X$, para $0 < p < q$.

Teorema: Si $X_n \xrightarrow{\text{m-}p} X$, entonces $X_n \xrightarrow{P} X$.

Observación:

$$X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{\text{m-}p} X.$$

$$X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{\text{m-}p} X.$$

Algunos recíprocos parciales

Teorema: Si $|X_n - X| \leq K$ (constante) $n \geq 1$ y $X_n \xrightarrow{P} X$, entonces $X_n \xrightarrow{\text{m-}p} X$.

Teorema: Si $X_n \xrightarrow{\text{c.s.}} X$ y $|X_n| \leq U$ v.a. integrable $(n \geq 1)$, entonces $X_n \xrightarrow{\text{m-1}} X$.

Convergencia en probabilidad versus convergencia en distribución

Teorema: Si $X_n \xrightarrow{P} X$, entonces $X_n \xrightarrow{\mathcal{D}} X$.

Demostración: Para $x \in C_F$, bastará demostrar que:

(1) $\limsup F_n(x) \leq F(x)$.

(2) $F(x) \leq \liminf F_n(x)$.

Para ver (1), comprobar que para $\epsilon > 0$, se tiene que

$$\{X_n \leq x\} \subset \{X \leq x + \epsilon\} \cup \{|X_n - X| \geq \epsilon\}.$$

Para mostrar (2), comprobar que

$$\{X \leq x - \epsilon\} \subset \{X_n \leq x\} \cup \{|X_n - X| \geq \epsilon\}.$$

Convergencia en probabilidad versus convergencia en distribución

Observación: $X_n \xrightarrow{\mathcal{D}} X \nRightarrow X_n \xrightarrow{\mathrm{P}} X$.

Teorema: Si $X_n \xrightarrow{\mathcal{D}} a$ (constante), entonces $X_n \xrightarrow{\mathrm{P}} a$.