

---

# PROBABILIDAD I

---

Segundo del Grado en Matemáticas

**Hugo Marquerie**

Profesor: Pablo Fernández Gallardo

Facultad de Ciencias - Universidad Autónoma de Madrid

Segundo cuatrimestre 2023 - 2024

1 de Febrero, 2024

# Índice

# 1 Tema 1: Sucesos y probabilidades

## 1.1 Formalizando

**Definición 1.1 (Espacio muestral).** En un experimento aleatorio, el espacio muestral es el conjunto (no vacío) de sus posibles resultados y se denota por  $\Omega$ . Puede ser:

1. Finito:  $\Omega = \{\omega_1, \dots, \omega_N\}$
2. Infinito numerable:  $\Omega = \{\omega_1, \omega_2, \dots\}$
3. Infinito no numerable, ej.:  $\Omega = [0, 1) \vee \Omega = \mathcal{P}([0, 1))$

**Definición 1.2 (Espacio de sucesos por Kolmogórov).** Dado el espacio muestral  $\Omega$  de un experimento aleatorio,  $\mathcal{F} \subset \mathcal{P}(\Omega)$  es su espacio de sucesos

$$\iff (\mathcal{F} \neq \emptyset) \wedge (A \in \mathcal{F} \implies A^C \in \mathcal{F}) \wedge \left( A_1, A_2, \dots \in \mathcal{F} \implies \bigcup_{j=1}^{\infty} A_j \in \mathcal{F} \right)$$

**Observación 1.1.** De la definición se deduce:

$$\bullet \phi \in \mathcal{F} \wedge \Omega \in \mathcal{F} \quad \bullet A_1, \dots, A_n \in \mathcal{F} \implies \bigcup_{j=1}^n A_j \in \mathcal{F} \quad \bullet \forall A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}$$

**Definición 1.3 (Función o medida de probabilidad).** Dado espacio muestral ( $\Omega$ ) y de sucesos ( $\mathcal{F}$ ) de un experimento aleatorio, la aplicación  $P: \mathcal{F} \rightarrow [0, 1]$  es una medida de probabilidad

$$\iff (P(\Omega) = 1) \wedge \left[ P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j) \iff A_i \cap A_j = \emptyset \text{ cuando } i \neq j \right]$$

**Proposición 1.1.** De la definición se deduce:

$$\begin{aligned} 1. P\left(\bigcup_{j=1}^n A_j\right) &= \sum_{j=1}^n P(A_j) & 2. P(A^C) &= 1 - P(A) & 3. P(\emptyset) &= 0 \\ 4. P(A \cup B) &= P(A) + P(B) - P(A \cap B) & 5. A \subseteq B &\implies P(A) \leq P(B) \end{aligned}$$

**Ejemplo 1.1.** En un experimento aleatorio con espacio muestral finito, tomamos

$\Omega = \{\omega_1, \dots, \omega_N\} \wedge \mathcal{F} = \mathcal{P} \rightarrow 2^N$ . Asignamos  $P(\{\omega_j\}) = p_j \wedge j = 1, \dots, N$  tales que  $p_j \geq 0 \wedge \sum_{j=1}^N p_j = 1$ . Entonces,  $\forall A \in \mathcal{F} : P(A) = \sum_{\omega \in A} P(\omega)$

**Caso particular:**  $\forall j \in \mathbb{N}_N : p_j = \frac{1}{N} \implies \forall A \in \mathcal{F} : P(A) = \frac{|A|}{N} = \frac{\text{“Casos favorables”}}{\text{“Total de casos”}}$

**Ejemplos varios:**

1. (Muy tonto)  $\Omega \neq \phi$ , tomas  $A \subset \Omega : A \neq \phi, \Omega$ .  
Dato  $p \in (0, 1)$ .  $\mathcal{F} = \{\phi, A, A^c, \Omega\}$  con  $P(A) = p$ .
2. (Bastante general)  $(\Omega = \{\omega_1, \dots, \omega_N\} \implies |\Omega| = N) \wedge (\mathcal{F} = P(\Omega) \rightarrow |\mathcal{F}| = 2^N)$ .  
Dato:  $p_1, \dots, p_N \geq 0 \implies \sum_{j=1}^N p_j = 1$ . Asignamos  $\forall j \in \mathbb{N}_N : p_j := P(\{\omega_j\})$ .  
Definimos  $\forall A \subset \Omega : P(A) = \sum_{\omega \in A} P(\omega)$ .
3. Lanzas  $n$  veces una moneda. Dato:  $p \in (0, 1)$ .  
 $\implies \Omega = \{111 \dots 1, \dots, 000 \dots 0\} \wedge |\Omega| = 2^N \wedge$  escogemos  $\mathcal{F} = P(\Omega)$   
 $\implies \forall \omega \in \Omega : P(\omega) = p^{\#\text{unos de } \omega} (1-p)^{\#\text{ceros de } \omega}$

Comprobamos:

$$\begin{aligned} \sum_{\omega \in \Omega} P(\omega) &= \sum_{k=0}^n \left( \sum_{\substack{\omega \in \Omega \\ \#\text{0s de } \omega = k}} P(\omega) \right) = \sum_{k=0}^n p^k (1-p)^{n-k} (|\{\omega \in \Omega : \#\text{1s de } \omega = k\}|) \\ &= \sum_{k=0}^n p^k (1-p)^{n-k} \binom{n}{k} = (p + 1 - p)^n = 1 \end{aligned}$$

4. Lanzamos moneda hasta que sale una cara. Dato  $p \in (0, 1)$ .  
 $\implies \Omega = \{C, XC, XXC, \dots\} \wedge$  escogemos  $\mathcal{F} = P(\Omega)$   
 $P(C) =: p \implies P(XC) = p(1-p) \wedge P(XXC) = p(1-p)^2 \wedge \dots$

Comprobamos:  $\sum_{j=1}^{\infty} p(1-p)^{j-1} = p \sum_{k=0}^{\infty} (1-p)^k = p \frac{1}{1-(1-p)} = 1$

**1.2 Probabilidad condicionada y total, independencia y Bayes**

Tienes  $(\Omega, \mathcal{F}, P)$  y un suceso  $A \in \mathcal{F} \rightarrow P(A)$ . Llega “nueva información”: ha ocurrido el suceso  $B \in \mathcal{F} \rightarrow$  ¿Debo reasignar la probabilidad de  $A$ ?

**Ejemplo 1.2 (Dependencia).** Lanzas 10 veces una moneda (regular).

$$A = \{\text{salen 6 caras}\} \implies P(A) = \frac{\binom{10}{6}}{2^{10}} \approx 20.51\%$$

$$B = \{\text{sale C en 1}^{\text{o}}\} \implies P(A) \text{ sube a } \frac{\binom{9}{5}}{2^9}$$

**Definición 1.4.** Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y los sucesos  $A, B \in \mathcal{F} : P(B) > 0$ ,  $P(A|B)$  es la probabilidad de  $A$  condicionada a  $B$

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

**Observación 1.2.** En general,  $P(A|B) \neq P(B|A)$

**Proposición 1.2** (Cálculo de  $P(A|B)$  para cada  $A \in \mathcal{F}$ ). Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y  $B \in \mathcal{F}$  un suceso con  $P(B) > 0$

$\implies (\Omega, \mathcal{F}, Q_B)$ , con  $Q_B: \mathcal{F} \rightarrow [0, 1] : Q_B(A) = P(A|B)$ , es un espacio de probabilidad

**Demostración.** Basta ver que  $Q_B$  es una función de probabilidad.

$$\left( Q_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} \in [0, 1] \right) \wedge \left( Q_B(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1 \right)$$

Sean  $A_1, A_2, \dots \in \mathcal{F}$  disjuntos dos a dos.

$$\begin{aligned} \implies Q_B \left( \bigcup_{j=1}^{\infty} A_j \right) &= P \left( \bigcup_{j=1}^{\infty} A_j \mid B \right) = \frac{P \left( \bigcup_{j=1}^{\infty} A_j \cap B \right)}{P(B)} \\ &= \frac{1}{P(B)} P \left( \bigcup_{j=1}^{\infty} (A_j \cap B) \right) = \frac{1}{P(B)} \sum_{j=1}^{\infty} P(A_j \cap B) = \sum_{j=1}^{\infty} Q_B(A_j) \end{aligned}$$

■

**Definición 1.5 (Independencia).** Sean  $A, B \in \mathcal{F}$  dos sucesos con  $P(A), P(B) \geq 0$  son independientes

$$\Longleftrightarrow P(A|B) = P(A) \wedge P(B|A) = P(B) \text{ (para entender)}$$

$$\Longleftrightarrow P(A \cap B) = P(A) \cdot P(B) \text{ (la adecuada)}$$

•  $A, B$  disjuntos  $\implies$  no independientes.

•  $A_1, \dots, A_N \in \mathcal{F}$  independientes  $\Longleftrightarrow \forall J \subset \mathbb{N}_N : P \left( \bigcap_{j \in J} A_j \right) = \prod_{j \in J} P(A_j)$

$$\Longleftrightarrow P \left( \overline{A_1} \cap \dots \cap \overline{A_N} \right) = \prod_{i=1}^N P \left( \overline{A_i} \right) \text{ donde } \overline{A_i} = A_i^c, (A_i)^c$$

**Ejercicio 1.2.1.** Encontrar un espacio de probabilidad en el que haya un conjunto de sucesos independientes dos a dos pero no completamente independientes.

$$\text{SOL: } \Omega = \{1, 2, 3, 4\} \wedge A = \{1, 2\} \wedge B = \{2, 3\} \wedge C = \{1, 3\}$$

**Proposición 1.3** (Regla de Bayes). Sean  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y  $A, B \in \mathcal{F}$  sucesos con  $P(A), P(B) > 0$

$$\implies P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)}$$

**Demostración.**

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

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

■

07/02/2024

### 1.2.1 Probabilidad total

**Proposición 1.4.** Sean  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y  $\{B_1, B_2, \dots\}$  una partición de  $\Omega : (\forall i, j : i \neq j : B_i \cap B_j = \emptyset) \wedge \left( \bigcup_{j=1}^{\infty} B_j = \Omega \right)$

$$\implies \forall A \in \mathcal{F} : P(A) = P\left(\bigcup_{j=1}^{\infty} (A \cap B_j)\right) = \sum_{j=1}^{\infty} P(A \cap B_j) = \sum_{j=1}^{\infty} P(A|B_j) \cdot P(B_j)$$

$$\implies \forall A \in \mathcal{F} : \boxed{P(A) = \sum_{j=1}^{\infty} P(A|B_j) \cdot P(B_j)}$$

■

**Ejemplo 1.3.** Sean  $U_1 = \{10b, 3n\} \wedge U_2 = \{5b, 5n\} \wedge U_3 = \{2b, 6n\}$  tres urnas con bolas blancas ( $b$ ) y negras ( $n$ ). Procedimiento:

1. Sorteamos una urna  $P(U_1) = \frac{1}{4} \wedge P(U_2) = \frac{1}{4} \wedge P(U_3) = \frac{1}{2}$
2. Sacamos bola de la urna seleccionada.

¿Cuál es la probabilidad de que sea blanca?

$$P(b) = P(b|U_1)P(U_1) + P(b|U_2)P(U_2) + P(b|U_3)P(U_3)$$

**Ejemplo 1.4 (Peso de la evidencia).** Sean  $U_1 = \{80\% b, 20\% n\} \wedge U_2 = \{20\% b, 80\% n\}$  dos urnas con bolas blancas ( $b$ ) y negras ( $n$ ). Procedimiento:

1. Sorteamos la urna con  $1/2$  y  $1/2$  de probabilidad.
2. Sacamos 10 bolas (con reemplazamiento).

Observamos la evidencia:  $bb \dots nb$  ¿qué urna se usó?

$$P(U_1|5b5n) = P(5b5n|U_1) \frac{P(U_1)}{P(5b5n)} = \frac{P(5b5n|U_1)P(U_1)}{P(5b5n|U_1)P(U_1) + P(5b5n|U_2)P(U_2)}$$

$$\implies P(5b5n|U_1) = \binom{10}{5} 0.8^5 0.2^5 = P(5b5n|U_2) \implies P(U_1|5b5n) = \frac{1}{2}$$

Este es el resultado que esperábamos, prestemos atención a otro caso más contraintuitivo.

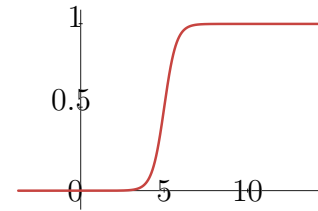
$$P(U_1|6b4n) = \frac{P(6b4n|U_1)P(U_1)}{P(6b4n|U_1)P(U_1) + P(6b4n|U_2)P(U_2)}$$

$$= \frac{\binom{10}{6} 0.8^6 0.2^4 \cdot 1/2}{\binom{10}{6} 0.8^6 0.2^4 \cdot 1/2 + \binom{10}{6} 0.8^4 0.2^6 \cdot 1/2} \approx 90\%$$

Si dibujamos la gráfica de la función

$$f(x) = P(U_1|xb(10-x)n)$$

podemos ver que el cambio es muy brusco. Es decir, una pequeña diferencia en la evidencia puede cambiar mucho la probabilidad de que se haya usado una urna u otra.

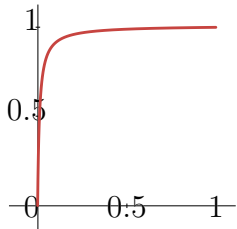


**Ejemplo 1.5 (Falsos positivos/negativos).** Hay una enfermedad ( $E \vee S$ ) y hay una prueba para detectar ( $+ \vee -$ ). Datos:  $P(+|E) = 95\% \wedge P(-|S) = 99\%$ .

Te haces la prueba y sale +:

$$P(E|+) = P(+|E) \frac{P(E)}{P(+)} = \frac{P(+|E)P(E)}{P(+|E)P(E) + P(+|S)P(S)}$$

Conozco todas estas probabilidades excepto  $p := P(E) \implies P(S) = 1 - p$ .



Si definimos  $f(p) := P(E|+)$

$$\implies \{f(0.5) = 98.95\% \wedge f(1/100) = 48.97\% \wedge f(1/1000) = 8.68\%\}$$

Es decir, si la incidencia es muy baja, no tiene sentido hacer pruebas masivamente porque la mayoría de positivos serán falsos.

08/02/2024

**Ejemplo 1.6 (Sobre independencia).**  $(\Omega, \mathcal{F}, P) \wedge A_1, \dots, A_n$  sucesos independientes tal

que  $\forall j \in \mathbb{N}_n : P(A_j) = \frac{1}{n}$ . ¿Qué sabemos sobre  $P\left(\bigcup_{j=1}^n A_j\right)$ ?

$$\text{En general, sabemos que } \frac{1}{n} \leq P\left(\bigcup_{j=1}^n A_j\right) \leq \sum_{j=1}^n P(A_j) \leq 1$$

$$n = 2: P(A \cup B) = P(A) + P(B) - P(A \cap B) = 3/4$$

$$n = 3: P(A \cup B \cup C) = \binom{3}{1} \frac{1}{3} - \binom{3}{2} \frac{1}{3^2} + \binom{3}{3} \frac{1}{3^3} = 19/27$$

$$\begin{aligned}
n \text{ general: } P\left(\bigcup_{j=1}^n A_j\right) &= \sum_{j=1}^n P(A_j) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \cdots \text{ (Inclusión exclusión)} \\
\Rightarrow P\left(\bigcup_{j=1}^n A_j\right) &= \binom{n}{1} \frac{1}{n} - \binom{n}{2} \frac{1}{n^2} + \binom{n}{3} \frac{1}{n^3} + \cdots = \sum_{j=1}^n \binom{n}{j} \frac{1}{n^j} (-1)^{j+1} \\
\Rightarrow P\left(\bigcup_{j=1}^n A_j\right) &= 1 - \sum_{j=0}^n \binom{n}{j} \left(\frac{-1}{n}\right)^j = 1 - \left(1 - \frac{1}{n}\right)^n \xrightarrow{n \rightarrow \infty} 1 - \frac{1}{e}
\end{aligned}$$

### 1.2.2 Continuidad de la probabilidad: detalle técnico

**Proposición 1.5.** Sean  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidades y  $A_1, \dots : A_1 \subset A_2 \subset \dots$  una sucesión creciente de conjuntos

$$\Rightarrow P\left(\bigcup_{j=1}^{\infty} A_j\right) = \lim_{n \rightarrow \infty} P(A_n)$$

**Demostración.** Se trata de describir  $\bigcup_{j=1}^n A_j$  como la unión de conjuntos disjuntos.

$$\begin{aligned}
P\left(\bigcup_{j=1}^n A_j\right) &= P\left(A_1 \cup \bigcup_{j=1}^{n-1} (A_{j+1} \setminus A_j)\right) = P(A_1) + \sum_{j=1}^{n-1} (P(A_{j+1}) - P(A_j)) = P(A_n) \\
\Rightarrow P\left(\bigcup_{j=1}^{\infty} A_j\right) &= P(A_1) + \sum_{j=1}^{\infty} (P(A_{j+1}) - P(A_j)) = \lim_{n \rightarrow \infty} P(A_n)
\end{aligned}$$

■

**Proposición 1.6.** Si la sucesión  $A_1, \dots$  es decreciente  $\Rightarrow P\left(\bigcap_{j=1}^{\infty} A_j\right) = \lim_{n \rightarrow \infty} P(A_n)$

**Teorema 1.1** (Continuidad de la probabilidad). En el espacio de probabilidades  $(\Omega, \mathcal{F}, P)$ , sea  $A_1, A_2, \dots$  una sucesión :  $\forall j \in \mathbb{N} : A_j \in \mathcal{F}$

$$\Rightarrow P\left(\bigcup_{j=1}^{\infty} A_j\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{j=1}^n A_j\right)$$

**Demostración.** Definimos  $B_i := \bigcup_{j=1}^i A_j \Rightarrow \bigcup_{j=1}^{\infty} A_j = \bigcup_{i=1}^{\infty} B_i$  ( $B_i$ ) es creciente.

$$\Rightarrow P\left(\bigcup_{j=1}^{\infty} A_j\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_{n \rightarrow \infty} P(B_n) = P\left(\bigcup_{j=1}^n A_j\right)$$

■



## 2 Tema 2: Variables aleatorias discretas

**Definición 2.1.** Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidades,  $X: \Omega \rightarrow \mathbb{R}$  es una variable aleatoria discreta\* (v.a.d.)

$$\iff (1) X(\Omega) \text{ es numerable}^* \wedge (2) \forall x \in \mathbb{R} : \{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$$

En realidad, solo interesa (2) cuando  $x = x_j$

**Definición 2.2.** Sea  $X$  una v.a.d. en  $(\Omega, \mathcal{F}, P)$ ,  $p_X: \mathbb{R} \rightarrow [0, 1]$  es su función de masa

$$\iff x \mapsto p_X(x) = P(X = x) := P(\{\omega \in \Omega : X(\omega) = x\})$$

Vemos que

$$\sum_{j \geq 1} p_X(x_j) = \sum_{j \geq 1} P(X = x_j) = \sum_{j \geq 1} P(\{\omega \in \Omega : X(\omega) = x_j\}) = P\left(\bigcup_{\omega \in \Omega} \{\omega\}\right) = P(\Omega) = 1$$

Lo relevante es el conjunto de posibles valores de  $X$  ( $\{x_1, x_2, \dots\}$ ) numerable y el conjunto (también numerable) de probabilidades  $\{p_1, p_2, \dots\}$  donde

$$\left(\forall j \geq 1 : p_j = P(x = x_j) \wedge p_j \geq 0\right) \wedge \sum_{j \geq 1} p_j = 1$$

**Teorema 2.1.** Sea  $S = \{x_1, x_2, \dots, x_j\}$  un conjunto y  $(\Pi_1, \Pi_2, \dots, \Pi_j)$  una lista tal que

$$\forall i \leq j : \Pi_i \geq 0 \wedge \sum_{j \geq 1} \Pi_j = 1$$

$$\implies \exists (\Omega, \mathcal{F}, P) \wedge X \text{ v.a.d.} : (\forall x \notin S : p_X(x) = 0) \wedge p_X(x_i) = \Pi_i$$

**Demostración.** Fijamos  $\Omega = S$  y  $\mathcal{F} = \mathcal{P}(S)$ .

$$A \in \mathcal{F} \implies P(A) = \sum_{j: x_j \in A} \Pi_j \wedge X(x_j) = x_j$$

■

**Ejemplo 2.1 (Diferentes modelos de distribución de probabilidad).**

1.  $X$  sigue una distribución **uniforme** en  $\{1, \dots, N\}$ ,  $N \geq 2$  ( $X \sim \text{UNIF}(N)$ ).

$$\iff S = \{1, \dots, N\} \wedge \Pi_j = 1/N, \dots, 1/N$$

Se usa para modelizar un lanzamiento de un dado regular de  $N$  caras.

2.  $X$  sigue una distribución de **Bernoulli** con parámetro  $p$  ( $X \sim \text{BER}(p)$ )

$$\iff \begin{cases} p_X(x) = 0 \iff x \neq 0, 1 \\ p_X(1) = p \wedge p_X(0) = 1 - p \end{cases} \iff \begin{cases} P(X = 1) = p \\ P(X = 0) = 1 - p \end{cases}$$

donde 1 es éxito y 0 fracaso. Se usa para modelizar el resultado de un experimento con dos posibles resultados, i.e. una moneda no necesariamente regular.

13/02/2024

3.  $X$  sigue una distribución **binomial** de parámetros  $n \geq 1 \wedge p \in (0, 1)$  ( $X \sim \text{BIN}(n, p)$ )

$$\iff S = \{0, 1, \dots, n\} \wedge \forall j \in \{0, 1, \dots, n\} : P(X = j) = \binom{n}{j} p^j (1 - p)^{n-j}$$

Sirve para modelizar el número de caras que salen al lanzar  $n$  veces una moneda de probabilidad  $p$ .

Podemos estimar cual es la probabilidad de que salgan  $n/2$  caras con  $p = 1/2$  mediante la fórmula de Stirling:

$$\begin{aligned} n! \sim n^n e^{-n} \sqrt{2\pi n} &\implies \binom{n}{n/2} \sim \frac{n^n e^{-n} \sqrt{2\pi n}}{(n/2)^{(n/2)} e^{-(n/2)} \sqrt{2\pi(n/2)} (n/2)^{(n/2)} e^{-(n/2)} \sqrt{2\pi(n/2)}} \\ &\implies \frac{n^n \sqrt{n}}{(n/2)^{(n/2)} \sqrt{2\pi(n/2)} (n/2)^{(n/2)} \sqrt{(n/2)}} = \frac{n^n \sqrt{n}}{(n/2)^n \sqrt{2\pi(n/2)}} = \frac{n^n \sqrt{2}}{(n/2)^n \sqrt{\pi n}} = 2^n \sqrt{\frac{2}{\pi n}} \\ &\implies P\left(X = \frac{n}{2}\right) = \binom{n}{n/2} \frac{1}{2^n} \approx 2^n \sqrt{\frac{2}{\pi n}} \frac{1}{2^n} = \sqrt{\frac{2}{\pi n}} \end{aligned}$$

4.  $X$  sigue una distribución **geométrica** de parámetro  $p \in (0, 1)$  ( $X \sim \text{GEOM}(p)$ ).

$$\iff S = \{1, 2, \dots\} \wedge \forall j \geq 1 : P(X = j) = p(1 - p)^{j-1}$$

Sirve para modelizar el número de lanzamientos hasta que sale un resultado  $C$  en cuestión con  $P(X = C) = p$ .

**Observación 2.1.** Cuidado porque existen variables aleatorias que también se dicen de distribución geométrica en las que  $S = \{0, 1, 2, \dots\}$ . Se habla de cuantas veces has obtenido el resultado complementario a  $C$  antes de que halla salido  $C$ .

5.  $X$  sigue una distribución de **Poisson** con parámetro  $\lambda > 0$  ( $X \sim \text{POISSON}(\lambda)$ )

$$\iff S = \{0, 1, \dots\} \wedge \forall j \geq 0 : P(X = j) = e^{-\lambda} \frac{\lambda^j}{j!}$$

Se usa para modelizar la frecuencia de eventos determinados durante un intervalo de tiempo fijado a partir de la frecuencia media de aparición de dichos eventos.

**Proposición 2.1.** Sea  $X \sim \text{BIN}(n, p)$  una v.a.d.

$\implies$  cuando  $n$  es grande,  $\text{BIN}(n, p) \sim \text{POISSON}(np)$

**Demostración.** Fijo  $\lambda > 0$   $\wedge$   $p = \frac{\lambda}{n}$ .

$$\lim_{n \rightarrow \infty} \binom{n}{k} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} \frac{n!}{(n-k)!} \frac{1}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} = e^{-k} \frac{\lambda^k}{k!}$$

■

**Ejemplo 2.2** (¿Hay más ejemplos?).

- Binomial negativa
- Hipergeométrica
- Sea cualquier serie convergente  $\sum_{n=1}^{\infty} a_n = s$

$\implies$  se puede definir la variable aleatoria  $X : S = \{1, 2, \dots\} \wedge P(x = k) = \frac{a_k}{s}$

14/02/2024

## 2.1 Funciones transformadoras de variables aleatorias (discretas)

Sea  $X$  una v.a.d. y  $g : \mathbb{R} \longrightarrow \mathbb{R}$  una función, definimos  $Y := g(X)$ .

$\implies \{\omega \in \Omega : Y(\omega) = g(X(\omega)) = y\} \in \mathcal{F} \implies Y$  es una v.a.d

Por otro lado,

$$\forall y \in \mathbb{R} : P(Y = y) = P(g(X) = y) = P(X \in g^{-1}(y)) = \sum_{x \in g^{-1}(y)} P(X = x) = \sum_{x \in g^{-1}(y)} p_X(x)$$

**Ejemplo 2.3** ( $Y = x^2$ ).

$$p_Y(y) = P(Y = y) = P(X^2 = y) = \begin{cases} 0, & y < 0 \\ P(X = 0) = p_X(0), & y = 0 \\ P(X = \pm\sqrt{y}) = p_X(\sqrt{y}) + p_X(-\sqrt{y}), & y > 0 \end{cases}$$

## 2.2 Resúmenes: esperanza, varianza, momentos

**Definición 2.3 (Esperanza).** Sea  $X$  una v.a.d. en el espacio de probabilidad  $(\Omega, \mathcal{F}, P)$  y con función de masa  $p_X$ ,  $E(X)$  es la esperanza de  $X$  (también llamada media o *expectatio*)

$$\iff E(X) = \sum_{x \in X(\Omega)} x \cdot P(X = x) = \sum_{j \geq 1} x_j \cdot p_X(x_j)$$

Pero ojo, solo si la serie es absolutamente convergente.

- Si  $x_1, \dots, x_N$  finito, la suma obviamente converge.
- Si los  $x_j$  son positivos, la serie converge si y solo si es acotada. Si no lo es diverge a  $\infty$ .

**Ejemplo 2.4 (Cálculo de la esperanza).**

- $x_1, \dots, x_N \wedge p_1, \dots, p_N \implies E(X) = \sum_{j=1}^N x_j \cdot p_j$
- $X \sim \text{BER}(p) \implies E(X) = 1 \cdot p + 0 \cdot (1 - p) \implies \boxed{E(X) = p}$
- $X \sim \text{UNIF}(1, \dots, N) \implies E(X) = \frac{1}{N} \sum_{n=1}^N n \implies \boxed{E(X) = \frac{N+1}{2}}$
- $X \sim \text{BIN}(n, p) \implies E(X) = \sum_{j=0}^n j \binom{n}{j} p^j (1-p)^{n-j} \implies \boxed{E(X) = np}$

Se obtiene derivando el binomio de Newton  $(1+x)^n = \sum_{j=0}^n \binom{n}{j} x^j$

$$\implies \frac{d}{dx}(1+x)^n = \sum_{j=0}^n \binom{n}{j} j \cdot x^{j-1} \implies xn(1+x)^{n-1} = \sum_{j=0}^n \binom{n}{j} j x^j$$

$$\begin{aligned} \text{Si } x = \frac{p}{1-p} \implies \frac{p}{1-p} n \left(1 + \frac{p}{1-p}\right)^{n-1} &= \sum_{j=0}^n \binom{n}{j} j \left(\frac{p}{1-p}\right)^j \\ \implies \frac{p}{1-p} n \left(\frac{1}{1-p}\right)^{n-1} &= \sum_{j=0}^n \binom{n}{j} j \left(\frac{p}{1-p}\right)^j \\ \implies np = (1-p)^n \sum_{j=0}^n \binom{n}{j} j \left(\frac{p}{1-p}\right)^j &= \sum_{j=0}^n j \binom{n}{j} p^j (1-p)^{n-j} = E(X) \end{aligned}$$

■

- $X \sim \text{GEOM}(p) \implies E(X) = \sum_{k=1}^{\infty} k \cdot p \cdot (1-p)^{k-1} \implies \boxed{E(X) = \frac{1}{p}}$

$$\forall x : |x| < 1 : \frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \implies \frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} n x^{n-1} \implies E(X) = p \frac{1}{p^2} = \frac{1}{p}$$

■

- $X \sim \text{POISSON}(\lambda) \implies E(X) = \sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} \implies \boxed{E(X) = \lambda}$

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

■

**Ejemplo 2.5.** Sea  $X$  una v.a.d. con  $\forall k \geq 0 : P(X = k) = \frac{6}{\pi^2} \frac{1}{k^2}$

$$\implies E(X) = \sum_{k=1}^{\infty} k \frac{6}{\pi^2} \frac{1}{k^2} = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k} \text{ que diverge a } \infty$$

**Ejemplo 2.6.** Sea  $X$  una v.a.d. que toma valores en  $\{(-1)^{k+1}k : k \geq 1\} = \{1, -2, 3, -4, \dots\}$

$$P(X = (-1)^{k+1}k) = \frac{6}{\pi^2} \frac{1}{k^2} \implies E(X) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \text{ que sabemos que tiende a } \ln 2$$

Sin embargo, la serie no converge absolutamente, por tanto, mediante argumentos de reordenación, se puede argumentar que  $E(X)$  toma cualquier valor real. Entonces  $E(X)$  no tiene sentido.

**Teorema 2.2.** Sea  $X$  una v.a.d. que toma los valores  $x_j$  con probabilidades  $p_j$  para  $j \geq 1$ . Sea  $g: \mathbb{R} \rightarrow \mathbb{R}$  una función.

$$\implies E(g(X)) = \sum_{j \geq 1} g(x_j) p_j$$

**Demostración.** Sabemos que  $g(X)$  es una v.a.d. que toma valores en  $\{g(x_j) : j \geq 1\}$ , donde  $|\{g(x_j)\}| \leq |\{x_j\}|$  porque  $g$  puede no ser inyectiva.

$$\text{Como } P(g(X) = y) = \sum_{x \in g^{-1}(y)} P(X = x)$$

$$\implies E(g(X)) = \sum_{y \in g(X(\Omega))} y \cdot P(g(X) = y) = \sum_{y \in g(X(\Omega))} y \left( \sum_{x \in g^{-1}(y)} P(X = x) \right)$$

Como  $\forall y \in g(X(\Omega)) : \exists |g^{-1}(y)|$  cantidad de  $i_s \geq 1 : g(x_i) = y$  se tiene

$$E(g(X)) = \sum_{j \geq 1} g(x_j) p_j$$

■

**Observación 2.2.**

1. Si  $X$  es tal que  $P(X = a) = 1 \implies E(X) = a$

2.  $X \sim \text{UNIF}(\{-1, 0, 1\}) \wedge Y = X^2 \implies E(X) = 0 \wedge E(Y) = 2/3$

3.  $E(aX + b) = aE(X) + b$  porque

$$\sum_{j \geq 1} (ax_j + b)p_j = a \sum_{j \geq 1} x_j p_j + b \sum_{j \geq 1} p_j = aE(X) + b$$

4. En general  $E(g(X)) \neq g(E(X))$

(Motivo de excomuni3n)

**Ejemplo 2.7** ( $X \sim \text{POISSON}(\lambda) \implies E(X) = \lambda$ ). Si  $Y = g(X) = e^X$

$$\implies E(Y) = E(e^Y) = \sum_{j=0}^{\infty} e^j \frac{\lambda^j}{j!} e^{-\lambda} = e^{-\lambda} \sum_{j=0}^{\infty} \frac{(e\lambda)^j}{j!} = e^{-\lambda} e^{e\lambda} = e^{\lambda(e-1)} \neq e^\lambda$$

21/02/2024

**Definición 2.4 (Varianza).** Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y  $X$  una v.a.d. con función de masa  $p_X$ ,  $V(X)$  es la varianza de  $X$

$$\iff V(X) = E[(X - E(X))^2]$$

Si  $X$  toma valores  $x_1, x_2, \dots$  con probabilidades  $p_1, p_2, \dots$  y denominamos  $\mu := E(X)$

$$\implies V(X) = \sum_{j \geq 1} (x_j - \mu)^2 \cdot p_j = \sum_{j \geq 1} (x_j - \mu)^2 \cdot p_j = \sum_{j \geq 1} (x_j^2 - 2x_j\mu + \mu^2) \cdot p_j$$

$$\implies V(X) = \sum_{j \geq 1} x_j^2 p_j - 2 \sum_{j \geq 1} x_j \mu p_j + \sum_{j \geq 1} \mu^2 p_j = E(X^2) - 2\mu^2 + \mu^2$$

$$\implies \boxed{V(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2}$$

**Observación 2.3.**

1.  $V(X)$  es medida de dispersión de  $X$  alrededor de  $E(X)$ .

2.  $V(X) \geq 0$

3.  $V(X) = 0 \implies P(X = E(X)) = 1$

4.  $\boxed{V(aX + b)} = E[(aX + b - E(aX + b))^2] = E[(aX - aE(X))^2] = \boxed{a^2 V(X)}$

5. Las unidades de  $V(X)$  son las de  $X^2$

$\implies$  definimos la desviación típica de  $X$  como  $\boxed{\sigma(X) := \sqrt{V(X)}}$

6. ¿Por qué no  $E(|X - E(X)|)$ ?

Porque el valor absoluto no es diferenciable y no se puede trabajar con él.

**Ejemplo 2.8.**

1.  $X \sim \text{BER}(p) \implies E(X) = p \wedge \boxed{V(X)} = p - p^2 = \boxed{p(1 - p)}$

2.  $X \sim \text{UNIF}(\{1, \dots, N\}) \implies E(X) = \frac{N+1}{2} \wedge \boxed{V(X) = \frac{N^2 - 1}{12}}$

$$\implies V(X) = \frac{1}{N} \sum_{j=1}^N j^2 = \frac{1}{N} \cdot \frac{N(N+1)(2N+1)}{6} = \frac{(N+1)(N-1)}{12}$$

$$3. X \sim \text{BIN}(n, p) \implies E(X) = np \wedge \boxed{V(X) = np(1-p)}$$

**Demostración.** ■

$$4. X \sim \text{GEOM}(p) \implies E(X) = \frac{1}{p} \wedge \boxed{V(X) = \frac{1-p}{p^2}}$$

**Demostración.** ■

$$5. X \sim \text{POISSON}(\lambda) \implies E(X) = \lambda \wedge \boxed{V(X) = \lambda}$$

**Demostración.**

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

■

**Definición 2.5 (Momentos de  $X$ ).** Sea  $X$  una v.a.d. con función de masa  $p_X$ ,  $\mu_k$  es el  **$k$ -ésimo momento** de  $X \iff \mu_k = E[(X - E(X))^k]$

**Observación 2.4.** Algunos momentos tienen nombre propio:

1.  $\mu_1 = 0$       2.  $\mu_2 = V(X)$       3.  $\mu_3$  es la **asimetría** de  $X$       4.  $\mu_4$  es la **curtosis** de  $X$

**Teorema 2.3** (Desigualdad de Markov). Sea  $X$  una v.a.d. :  $P(X < 0) = 0 \wedge E(X) < \infty$   
 $\implies \forall t > 0 : P(X \geq t) \leq \frac{E(X)}{t}$

**Demostración. Notación:** En  $(\Omega, \mathcal{F}, P)$ ,  $A \subset \mathcal{F}$  y  $\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$

Fijamos  $t > 0$  y definimos  $Y_t(\omega) = t \cdot \mathbb{1}_{\{x \geq t\}}(\omega) = \begin{cases} t & \text{con probabilidad } P(x \geq t) \\ 0 & \text{con probabilidad } 1 - P(x \geq t) \end{cases}$

$$\implies \forall \omega : Y_t(\omega) \leq X(\omega) \implies E(Y_t) = t \cdot P(X \geq t) \leq E(X)$$

$$\text{Así que } \forall t > 0 : P(X \geq t) \leq \frac{E(X)}{t}$$

■

**Teorema 2.4** (Desigualdad de Chebyshev). Sea  $X$  una v.a.d. :  $E(X), V(X) < \infty$ .

$$\implies \forall \lambda > 0 : P(|X - E(X)| \geq \lambda \cdot \sigma(X)) \leq \frac{1}{\lambda^2}$$

$$\iff \forall \alpha > 0 : \boxed{P(|X - E(X)| \geq \alpha) \leq \frac{V(X)}{\alpha^2}}$$

**Demostración.** Definimos  $Y = |X - E(X)|^2$  y aplicamos la desigualdad de Markov.

$$\implies \forall t > 0 : P(Y \geq t) \leq \frac{E(Y)}{t} \implies \forall t > 0 : P(|X - E(X)|^2 \geq t) \leq \frac{E(Y)}{t}$$

Como  $E(Y) = E(|X - E(X)|^2) = V(X)$  por la def de varianza,

$$\implies \forall t > 0 : P(|X - E(X)| \geq \sqrt{t}) \leq \frac{V(X)}{t}$$

Definimos  $\alpha := \sqrt{t} \implies \forall \alpha > 0 : P(|X - E(X)| \geq \alpha) \leq \frac{V(X)}{\alpha^2}$

y para la desigualdad equivalente definimos  $\lambda := \frac{\alpha}{\sigma(X)}$

$$\implies \forall \lambda > 0 : P(|X - E(X)| \geq \lambda \cdot \sigma(X)) \leq \frac{1}{\lambda^2}$$

■

26/02/2024

### 2.2.1 Esperanza condicionada

**Definición 2.6 (Esperanza condicionada).** Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad  $B \in \mathcal{F}$  un suceso tal que  $P(B) > 0$  y  $X$  una v.a.d. con esperanza  $E(X)$ ,  $E(X|B)$  es la **esperanza de  $X$  condicionada a  $B$**

$$\iff E(X|B) := \sum_{x \in X(\Omega)} x \cdot P(X = x|B) = \sum_{x \in X(\Omega)} x \cdot \frac{P((X = x) \wedge B)}{P(B)}$$

(Siempre que la serie sea absolutamente convergente).

**Teorema 2.5 (Esperanza total).** En  $(\Omega, \mathcal{F}, P)$ , sea  $X$  una v.a.d. y  $\{B_1, B_2, \dots\}$  una partición de  $\Omega$

$$\implies E(X) = \sum_{i \geq 1} E(X|B_i) \cdot P(B_i)$$

(Siempre que la serie sea absolutamente convergente).

**Demostración.**

$$\begin{aligned} \sum_{i \geq 1} E(X|B_i) \cdot P(B_i) &= \sum_{i \geq 1} \sum_{x \in X(\Omega)} x \cdot P(X = x|B_i) \cdot P(B_i) \\ &= \sum_{x \in X(\Omega)} x \cdot \sum_{i \geq 1} \frac{P((X = x) \wedge B_i)}{P(B_i)} \cdot P(B_i) = \sum_{x \in X(\Omega)} x \cdot P(X = x) \end{aligned}$$

■

**Ejemplo 2.9.** Lanzamos una moneda con probabilidad  $p$  de cara y  $1 - p$  de cruz y definimos  $X$  como la longitud de la racha inicial, i.e. el número de caras/cruces consecutivas.

$$\begin{aligned} \implies E(X) &= E(X|C) \cdot p + E(X|\times) \cdot (1 - p) \\ \implies E(X) &= \left( \sum_{j \geq 1} j \cdot P(X = j|C) \right) p + \left( \sum_{j \geq 1} j \cdot P(X = j|\times) \right) (1 - p) \end{aligned}$$



$$\begin{aligned}\implies E(X) &= p \sum_{k=1}^{\infty} k p^{k-1} (1-p) + (1-p) \sum_{k=1}^{\infty} k (1-p)^{k-1} p \\ \implies E(X) &= \frac{1}{1-p} \cdot p + \frac{1}{p} \cdot (1-p) = \frac{1}{p(1-p)} - 2\end{aligned}$$

También se puede abordar el problema pensando en las variables geométricas:

$$\begin{aligned}E(X) &= E(X|C) \cdot p + E(X|\times) \cdot (1-p) = \frac{p}{1-p} + \frac{1-p}{p} = \frac{p^2 + (1-p)^2}{p(1-p)} \\ \implies E(X) &= \frac{p^2 + 1 - 2p + p^2}{p(1-p)} = \frac{2p(p-1) + 1}{p(1-p)} = \frac{1}{p(1-p)} - 2\end{aligned}$$

27/02/2024

## 2.3 Varias variables aleatorias

En  $(\Omega, \mathcal{F}, P)$ , sea  $\mathcal{X} = X_1, X_2, \dots, X_n$  una colección de variables aleatorias discretas.

En el caso  $n = 2$ , tenemos  $X$  e  $Y$  variables aleatorias discretas, se genera una tabla con las probabilidades conjuntas:

$$p_{X,Y}(x, y) := P(\omega \in \Omega : X(\omega) = x \wedge Y(\omega) = y) = P(X = x \wedge Y = y)$$

**Definición 2.7 (Función de masa conjunta).** En  $(\Omega, \mathcal{F}, P)$ , sean  $X$  e  $Y$  v.a.d.,  $p_{X,Y}$  es su función de masa conjunta

$$\iff p_{X,Y} : \mathbb{R}^2 \longrightarrow [0, 1] \wedge \forall (x, y) \in X(\Omega) \times Y(\Omega) : p_{X,Y}(x, y) = P(X = x \wedge Y = y)$$

tal que  $\sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} p_{X,Y}(x, y) = 1$  y  $\forall (x, y) \notin X(\Omega) \times Y(\Omega) : p_{X,Y}(x, y) = 0$

Ya tenemos  $X$  e  $Y$  en  $(\Omega, \mathcal{F}, P)$  con  $P_{X,Y}$

1. ¿Qué sabemos de  $X$  e  $Y$  por separado?
2. Esperanzas: nos interesa calcular  $E(X)$ ,  $E(Y)$ ,  $E(X + Y)$ ,  $E(X \cdot Y)$
3. Independencia

**Definición 2.8 (Funciones marginales).** En  $(\Omega, \mathcal{F}, P)$ , sean  $X$  e  $Y$  v.a.d.,  $p_X, p_Y$  son sus funciones de masa marginales

$$\iff p_X(x) = \sum_{y \in Y(\Omega)} p_{X,Y}(x, y) \quad \wedge \quad p_Y(y) = \sum_{x \in X(\Omega)} p_{X,Y}(x, y)$$

**Teorema 2.6.** En  $(\Omega, \mathcal{F}, P)$ , sean  $X$  e  $Y$  v.a.d. y sea  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$  una función

$$\implies E(g(X, Y)) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} g(x, y) \cdot p_{X,Y}(x, y)$$

Si converge absolutamente.

**Demostración.** Si consideramos la variable aleatoria  $Z$  que toma valores en  $X(\Omega) \times Y(\Omega)$  con función de masa  $p_Z = p_{X,Y}$ , entonces  $E(g(X, Y)) = E(g(Z))$ . Como  $X(\Omega) \times Y(\Omega)$  es numerable, podemos renombrar sus elementos como  $\{z_1, z_2, \dots\}$  y entonces del teorema ?? obtenemos:

$$E(g(X, Y)) = E(g(Z)) = \sum_{j \geq 1} g(z_j) \cdot p_Z(z_j) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} g(x, y) \cdot p_{X,Y}(x, y)$$

■

**Observación 2.5.** Si  $g: \mathbb{R} \rightarrow \mathbb{R}$

$$\begin{aligned} \implies E(g(X)) &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} g(x) \cdot p_{X,Y}(x, y) = \sum_{x \in X(\Omega)} \left( g(x) \sum_{y \in Y(\Omega)} p_{X,Y}(x, y) \right) \\ &\implies E(g(X)) = \sum_{x \in X(\Omega)} g(x) p_X(x) \end{aligned}$$

De manera análoga,  $E(g(Y)) = \sum_{y \in Y(\Omega)} g(y) p_Y(y)$

**Ejemplo 2.10** ( $E(aX + bY) = aE(X) + bE(Y)$ ).

$$\begin{aligned} E(aX + bY) &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} (ax + by) \cdot p_{X,Y}(x, y) \\ \implies E(aX + bY) &= a \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} x \cdot p_{X,Y}(x, y) + b \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} y \cdot p_{X,Y}(x, y) \end{aligned}$$

28/02/2024

**Definición 2.9 (Independencia de v.a.d.).** En  $(\Omega, \mathcal{F}, P)$ , sean  $X$  e  $Y$  v.a.d.,  $X$  e  $Y$  son independientes

$$\iff \forall (x, y) \in X(\Omega) \times Y(\Omega) : X = x \text{ y } Y = y \text{ son sucesos independientes.}$$

$$\iff \forall (x, y) \in X(\Omega) \times Y(\Omega) : P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y)$$

**Teorema 2.7.** En  $(\Omega, \mathcal{F}, P)$ , sean  $X$  e  $Y$  v.a.d.,  $X$  e  $Y$  son independientes

$$\iff \exists g, h: \mathbb{R} \rightarrow \mathbb{R} : p_{X,Y}(x, y) = g(x) \cdot h(y)$$

**Demostración.** (  $\implies$  ) Trivial:  $g(x) = p_X(x) \wedge h(y) = p_Y(y)$ .

(  $\impliedby$  ) Suponemos que  $\exists g, h: \mathbb{R} \longrightarrow \mathbb{R} : p_{X,Y}(x, y) = g(x) \cdot h(y)$ , veamos las funciones marginales.

$$\implies p_X(x) = \sum_{y \in Y(\Omega)} p_{X,Y}(x, y) = \sum_{y \in Y(\Omega)} g(x) \cdot h(y) = g(x) \sum_{y \in Y(\Omega)} h(y)$$

$$\text{Análogamente } p_Y(y) = h(y) \sum_{x \in X(\Omega)} g(x)$$

$$\implies p_X(x) \cdot p_Y(y) = \left( g(x) \sum_{z \in Y(\Omega)} h(z) \right) \left( h(y) \sum_{w \in X(\Omega)} g(w) \right)$$

$$\implies p_X(x) \cdot p_Y(y) = g(x) \cdot h(y) \sum_{z \in Y(\Omega)} \sum_{w \in X(\Omega)} g(w) \cdot h(z) = g(x) \cdot h(y) = p_{X,Y}(x, y)$$

$$\forall (x, y) \in X(\Omega) \times Y(\Omega) : P(X = x \wedge Y = y) = p_X(x) \cdot p_Y(y) = P(X = x) \cdot P(Y = y)$$

■

**Ejemplo 2.11.** Sean  $X$  e  $Y$  dos v.a.d. tales que

$$p_{X,Y}(x, y) = e^{-(\lambda+\mu)} \frac{\lambda^x \mu^y}{x!y!} \text{ con } x, y \in \mathbb{Z} \text{ y } \lambda, \mu > 0$$

$\implies$  Se puede interpretar como  $X \sim \text{POISSON}(\lambda)$  e  $Y \sim \text{POISSON}(\mu)$  independientes.

**Observación 2.6.** Si  $X$  e  $Y$  son independientes  $\implies E(X \cdot Y) = E(X) \cdot E(Y)$ .

Sin embargo, la implicación recíproca no es cierta.

(Motivo de excomuni3n)

**Definici3n 2.10 (Covarianza).** En  $(\Omega, \mathcal{F}, P)$ , sean  $X$  e  $Y$  v.a.d.,  $\text{cov}(X, Y)$  es la covarianza de  $X$  e  $Y$

$$\iff \boxed{\text{cov}(X, Y) = E[(X - E(X)) \cdot (Y - E(Y))] = E(X \cdot Y) - E(X) \cdot E(Y)}$$

29/02/2024

**Observaci3n 2.7.**

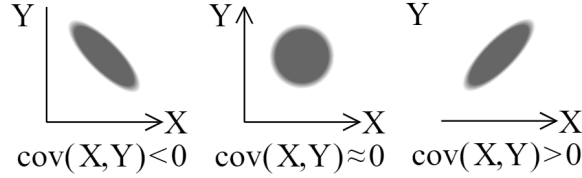
1. C3lculo de la covarianza

$$\text{cov}(X, Y) = \sum_{i,j} x_i y_j P(X = x_i \wedge Y = y_j) - \left( \sum_i x_i P(X = x_i) \right) \left( \sum_j y_j P(Y = y_j) \right)$$

2. Signo de la covarianza (y coeficiente de correlaci3n)

$$\text{cov}(X, Y) = \sum_{i,j} (x_i - E(X))(y_j - E(Y)) P(X = x_i \wedge Y = y_j)$$

Entonces, si la covarianza es positiva,  $X$  e  $Y$  tienden a crecer juntas. Si es negativa, tienden a decrecer juntas. Si es 0, no hay relación lineal entre  $X$  e  $Y$ .



### 3. Cálculo fundamental

$$\begin{aligned}
 V(X + Y) &= E((X + Y)^2) - (E(X + Y))^2 \\
 &= E(X^2 + 2XY + Y^2) - (E(X) + E(Y))^2 \\
 &= E(X^2) + 2E(XY) + E(Y^2) - (E(X))^2 - 2E(Y)E(X) - (E(Y))^2 \\
 &\implies \boxed{V(X + Y) = V(X) + V(Y) + 2\text{cov}(X, Y)}
 \end{aligned}$$

Pero cuidado:  $V(X - Y) = V(X) + V(Y) - 2\text{cov}(X, Y)$

De forma más general:

$$\begin{aligned}
 V(aX + bY) &= V(aX) + V(bY) + 2\text{cov}(aX, bY) \\
 &= a^2V(X) + b^2V(Y) + 2ab\text{cov}(X, Y) \\
 &\implies \boxed{V(aX + bY) = a^2V(X) + b^2V(Y) + 2ab\text{cov}(X, Y)}
 \end{aligned}$$

4. Si  $X$  e  $Y$  son independientes  $\implies \text{cov}(X, Y) = 0$

**Definición 2.11 (coeficiente de correlación).** Sean  $X$  e  $Y$  dos v.a.d.,  $\rho$  es su coeficiente de correlación  $\iff \rho := \frac{\text{cov}(X, Y)}{\sqrt{V(X) \cdot V(Y)}} \implies \rho$  no tiene unidades y  $|\rho(X, Y)| \leq 1$ .

**Proposición 2.2** (Desigualdad de Cauchy-Schwarz). Sean  $X$  e  $Y$  dos v.a.d.

$$\implies E(X \cdot Y)^2 \leq E(X^2) \cdot E(Y^2)$$

La igualdad se da cuando una variable es transformación lineal de la otra, i.e.  $Y = aX + b$ .

**Demostración.** Definimos  $W = sX + Y$ ,  $W^2 \geq 0$  con probabilidad 1.

$$\begin{aligned}
 \implies 0 \leq E(W^2) &= E((sX + Y)^2) = E(s^2X^2 + Y^2 + 2sXY) \\
 &= s^2E(X^2) + E(Y^2) + 2sE(XY) \\
 &= E(X^2) \cdot s^2 + 2E(XY) \cdot s + E(Y^2)
 \end{aligned}$$

Vemos que el resultado es una parábola si se toma como función de  $s$ .

Como  $\forall s \in \mathbb{R} : E(X) \geq 0$  y  $E(X^2) \geq 0$ , sabemos que la parábola o bien toca el eje  $X$  una única vez, o no lo hace nunca. Esto es equivalente a pedir que el valor del discriminante

sea menor o igual que 0.

$$4E(XY)^2 - 4E(X^2)E(Y^2) \leq 0 \implies E(XY)^2 \leq E(X^2)E(Y^2)$$

■

Por tanto,  $\overline{\text{cov}(X, Y)^2} = E((X - E(X))(Y - E(Y)))^2 \leq \overline{V(X) \cdot V(Y) \cdot \text{cov}(X, Y)}$

Además  $\rho(aX + b, cY + d) = \text{sgn}(ac) \cdot \rho(X, Y)$ .

04/03/2024

### 2.3.1 Detalle sobre independencia

**Teorema 2.8.** Sean  $X$  e  $Y$  dos v.a.d. en  $(\Omega, \mathcal{F}, P)$  y  $g, h: \mathbb{R} \rightarrow \mathbb{R}$  dos funciones

$$X \text{ e } Y \text{ independientes} \iff E(g(X) \cdot h(Y)) = E(g(X)) \cdot E(h(Y))$$

**Demostración.** ( $\implies$ )  $E(g(X) \cdot h(Y)) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} g(x) \cdot h(y) \cdot P(X = x \wedge Y = y)$

Como  $X$  e  $Y$  son independientes,  $P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y)$ , entonces

$$E(g(X) \cdot h(Y)) = \left( \sum_{x \in X(\Omega)} g(x) P(X = x) \right) \left( \sum_{y \in Y(\Omega)} h(y) P(Y = y) \right) = E(g(X)) \cdot E(h(Y))$$

( $\impliedby$ ) Sean  $\hat{x}, \hat{y} \in \mathbb{R}$ , queremos probar que  $P(X = \hat{x} \wedge Y = \hat{y}) = P(X = \hat{x}) \cdot P(Y = \hat{y})$

$$\text{Definimos } g(x) := \begin{cases} 1 & \text{si } x = \hat{x} \\ 0 & \text{si } x \neq \hat{x} \end{cases} \quad \wedge \quad h(y) := \begin{cases} 1 & \text{si } y = \hat{y} \\ 0 & \text{si } y \neq \hat{y} \end{cases}$$

$$\implies E(g(X) \cdot h(Y)) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} g(x) h(y) P(X = \hat{x} \wedge Y = \hat{y}) = P(X = \hat{x} \wedge Y = \hat{y})$$

Como  $E(g(X)) = P(X = \hat{x})$  y  $E(h(Y)) = P(Y = \hat{y})$  y  $E(g(X)h(Y)) = E(g(X))E(h(Y))$

$$\overline{P(X = \hat{x} \wedge Y = \hat{y})} = E(g(X) \cdot h(Y)) = E(g(X)) \cdot E(h(Y)) = \overline{P(X = \hat{x}) \cdot P(Y = \hat{y})}$$

■

¿Qué pasaría con  $(X_1, \dots, X_n)$  para  $n = 2$ ?

1. Modelo  $\rightarrow$  función de masa conjunta  $p_{X_1, \dots, X_n}(x_1, \dots, x_n)$

$$\begin{cases} \sum_{x_1 \in X_1(\Omega)} \cdots \sum_{x_n \in X_n(\Omega)} p_{X_1, \dots, X_n}(x_1, \dots, x_n) = 1 \\ p_{X_1, \dots, X_n}(x_1, \dots, x_n) \geq 0 \end{cases}$$

2. Marginales  $p_{X_1}(x_1) = \sum_{x_2 \in X_2(\Omega)} \cdots \sum_{x_n \in X_n(\Omega)} p_{X_1, \dots, X_n}(x_1, \dots, x_n)$

3. Independencia (la función de masa conjunta se factoriza)

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n) \iff \text{independencia completa}$$

Pero puede haber otras nociones de independencia (ej: 2 a 2).

4. **Matriz varianzas-covarianzas y matriz correlaciones** respectivamente

$$V = \begin{pmatrix} V(X_1) & \cdots & \text{cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \cdots & V(X_n) \end{pmatrix} \wedge \Sigma = \begin{pmatrix} 1 & \cdots & \rho(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \rho(X_n, X_1) & \cdots & 1 \end{pmatrix}$$

Ambas son simétricas y definidas positivas.

**Ejemplo 2.12.** Queremos modelizar experimentos del tipo lanzar 18 veces un dado y sumar los resultados obtenidos.

$$S_n = X_1 + \cdots + X_n \implies \begin{cases} E(S_n) = \sum_{i=1}^n E(X_i) \\ V(S_n) = \sum_{i=1}^n V(X_i) + 2 \sum_{i < j} \text{cov}(X_i, X_j) \end{cases}$$

Si suponemos las  $X_i$  independientes e idénticas  $\implies \forall i \in \mathbb{N}_n : E(X_i) =: \mu \wedge V(X_i) =: \sigma^2$

$$\implies E(S_n) = n\mu \wedge V(S_n) = n\sigma^2$$

Si definimos  $Z_n := \frac{1}{n}S_n = \frac{1}{n}(X_1 + \cdots + X_n) \implies E(Z_n) = \mu \wedge V(Z_n) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$   
 $\implies Z_n$  no es aleatoria si  $n \rightarrow \infty$  (ley de los grandes números)

05/03/2024

## 2.4 Funciones generatrices de probabilidad

### 2.4.1 Series de potencias

Sea  $(a_n)_{n=0}^\infty$  una sucesión y  $f(x) := \sum_{n=0}^\infty a_n x^n$  una función, ¿en qué valores de  $x$  está definida?

Sabemos que existe  $R \in [0, \infty)$  radio de convergencia tal que

$$\sum_{n=0}^\infty a_n x^n \rightarrow \begin{cases} \text{converge} & \text{si } |x| < R \\ \text{diverge} & \text{si } |x| > R \end{cases} \iff \frac{1}{R} = \limsup_{n \rightarrow \infty} |a_n|^{1/n}$$

**Ejemplo 2.13.**

$$\sum_{n=0}^\infty \frac{x^n}{n!} = e^x \quad \wedge \quad \sum_{n=0}^\infty x^n = \frac{1}{1-x}, |x| < 1 \quad \wedge \quad \sum_{n=0}^m \binom{m}{n} x^n = (1+x)^m$$

### 2.4.2 Funciones generatrices

$$(a_n)_{n=0}^{\infty} \longleftrightarrow f(x) = \sum_{n=0}^{\infty} a_n x^n \implies \begin{cases} x \cdot f(x) \longleftrightarrow (0, a_0, \dots) \\ x \cdot f'(x) \longleftrightarrow (0a_0, 1a_1, 2a_2, \dots) \end{cases}$$

$$f(x) \cdot g(x) = \sum_{n=0}^{\infty} a_n x^n \sum_{n=0}^{\infty} b_n x^n = \sum_{n=0}^{\infty} \left( \sum_{k=0}^n a_k b_{n-k} \right) x^n$$

**Definición 2.12 (Función generatriz de probabilidad).** Sea  $X$  una variable aleatoria discreta que toma valores en  $\{0, 1, 2, \dots\}$  donde  $\forall j \geq 0 : p_j = P(X = j)$ ,  $G_X(s)$  es su función

generatriz de probabilidad  $\iff \boxed{G_X(s) = \sum_{n=0}^{\infty} p_n s^n}$

**Ejemplo 2.14.**

1.  $X \sim \text{BER}(p) \implies G_X(s) = (1-p) + ps$

2.  $X \sim \text{BIN}(n, p) \implies G_X(s) = \sum_{j=0}^n \binom{n}{j} (1-p)^{n-j} (ps)^j = (1-p + ps)^n$

3.  $X \sim \text{GEOM}(p) \implies G_X(s) = \sum_{j=1}^{\infty} (1-p)^{j-1} ps^j = \frac{ps}{1 - (1-p)s}$

**Demostración.**

$$G_X(s) = \sum_{j=1}^{\infty} (1-p)^{j-1} ps^j = p \sum_{k=0}^{\infty} (1-p)^k s^{k+1} = ps \sum_{k=0}^{\infty} ((1-p)s)^k = \frac{ps}{1 - (1-p)s}$$

■

4.  $X \sim \text{POISSON}(\lambda) \implies G_X(s) = \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} s^j = e^{\lambda(s-1)}$

**Demostración.**  $G_X(s) = \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} s^j = e^{-\lambda} \sum_{j=0}^{\infty} \frac{(\lambda s)^j}{j!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}$

■

¿Para qué?

1. Cálculo de momentos con  $(p_n)_{n=0}^{\infty}$

$$G_X(s) = \sum_{n=0}^{\infty} p_n s^n \implies G'_X(s) = \sum_{n=1}^{\infty} n p_n s^{n-1} \implies G'_X(1) = \sum_{n=1}^{\infty} n p_n = E(X)$$

Si seguimos derivando, obtenemos

$$\implies G''_X(s) = \sum_{n=2}^{\infty} n(n-1) p_n s^{n-2} = \sum_{n=2}^{\infty} n^2 p_n s^{n-2} - \sum_{n=2}^{\infty} n p_n s^{n-2}$$

$$\begin{aligned} \implies G_X''(1) &= \sum_{n=2}^{\infty} n^2 p_n - \sum_{n=2}^{\infty} n p_n = E(X^2) - E(X) \\ \implies V(X) &= E(X^2) - E(X)^2 = G_X''(1) + G_X'(1)(1 - G_X'(1)) \end{aligned}$$

**Ejemplo 2.15.**

$$\begin{aligned} \text{(a) } X \sim \text{BER}(p) &\implies G_X(s) = (1 - p) + ps \\ &\implies G_X'(s) = p = E(X) \wedge G_X''(s) = 0 \implies V(X) = p(1 - p) \\ \text{(b) } X \sim \text{BIN}(n, p) &\implies G_X(s) = (1 - p + ps)^n \\ &\implies G_X'(s) = n(1 - p + ps)^{n-1}p \implies G_X'(1) = np = E(X) \\ &\implies G_X''(s) = n(n-1)(\dots)^{n-2}p^2 \implies G_X''(1) = n(n-1)p^2 \\ &\implies V(X) = n(n-1)p^2 + np(1 - np) = np(1 - p) \end{aligned}$$

2. Suma de independientes

**Teorema 2.9.** Sean  $X, Y$  dos v.a.d. independientes con valores en  $\{0, 1, 2, \dots\}$  y con funciones generatrices de probabilidad  $G_X(s), G_Y(s)$  respectivamente

$$\implies G_{X+Y}(s) = G_X(s) \cdot G_Y(s)$$

**Demostración.**

$$G_{X+Y}(s) = \sum_{n=0}^{\infty} P(X + Y = n) s^n = \sum_{n=0}^{\infty} \left( \sum_{k=0}^n P(X = k \wedge Y = n - k) \right) s^n$$

Por otro lado,

$$\begin{aligned} G_X(s) \cdot G_Y(s) &= \left( \sum_{n=0}^{\infty} P(X = n) s^n \right) \left( \sum_{n=0}^{\infty} P(Y = n) s^n \right) \\ &= \sum_{n=0}^{\infty} \left( \sum_{k=0}^n P(X = k) P(Y = n - k) \right) s^n \implies G_{X+Y}(s) = G_X(s) \cdot G_Y(s) \end{aligned}$$

■

Otra manera:

$$G_X(s) = E(s^X) \wedge G_{X+Y}(s) = E(s^{X+Y}) = E(s^X \cdot s^Y) = E(s^X) \cdot E(s^Y) = G_X(s) \cdot G_Y(s)$$

■

**Corolario 2.1.** Sean  $X_1, X_2, \dots, X_n$  v.a.d. independientes con valores en  $\{0, 1, 2, \dots\}$  y con



funciones generatrices de probabilidad  $G_{X_1}(s), G_{X_2}(s), \dots, G_{X_n}(s)$  respectivamente

$$\implies G_{X_1+X_2+\dots+X_n}(s) = \prod_{i=1}^n G_{X_i}(s)$$

Si las  $X_i$  son “idénticas”  $\implies G_{X_1+X_2+\dots+X_n}(s) = (G_{X_1}(s))^n$

06/03/2024

**Teorema 2.10** (Unicidad). Sean  $X, Y$  dos v.a.d. con valores en  $\{0, 1, 2, \dots\}$  y con funciones generatrices de probabilidad  $G_X(s), G_Y(s)$  respectivamente

$$\implies G_X(s) = G_Y(s) \iff \forall n \geq 0 : P(X = n) = P(Y = n)$$

**Ejemplo 2.16.** Sean  $X \sim \text{POISSON}(\lambda) \wedge Y \sim \text{POISSON}(\mu)$  independientes con  $\lambda, \mu > 0$ . Definimos  $Z = X + Y$ .

$$\implies \forall x \geq 0 : P(Z = k) = P(X + Y = k) = \sum_{j=0}^k P(X = j \wedge Y = k - j) = \dots$$

Pero, a través de funciones generatrices obtenemos:

$$G_X(s) = e^{\lambda(s-1)} \wedge G_Y(s) = e^{\mu(s-1)} \implies G_Z(s) = e^{(\lambda+\mu)(s-1)} \implies Z \sim \text{POISSON}(\lambda + \mu)$$

**Ejemplo 2.17.**

1. Sean  $I_1, I_2, \dots, I_n$  v.a.d. independientes con  $\forall k \in \mathbb{N}_n : I_k \sim \text{BER}(p)$  y definimos  $Z = I_1 + I_2 + \dots + I_n$ .

$$\implies G_Z(s) = [(1-p) + ps]^n \implies Z \sim \text{BIN}(n, p)$$

2. Sean  $X \sim \text{BIN}(n, p) \wedge Y \sim \text{POISSON}(\lambda)$  independientes y definimos  $Z = X + Y$ .

$$\implies G_Z(s) = ((1-p) + ps)^n \cdot e^{\lambda(s-1)}$$

### 3 Variables aleatorias continuas

Hasta ahora en  $(\Omega, \mathcal{F}, P)$ , una variable aleatoria  $X$  discreta era una función  $X: \Omega \rightarrow \mathbb{R}$  tal que  $\exists N \subseteq \mathbb{N} : |X(\Omega)| = |N|$  y  $P(X = k) = P(X^{-1}(k))$ .

**Definición 3.1 (Variable aleatoria).** Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad, la función  $X: \Omega \rightarrow \mathbb{R}$  es una variable aleatoria  $\iff \forall x \in \mathbb{R} : \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$

**Proposición 3.1.**  $X$  es v.a.d.  $\implies X$  es variable aleatoria.

**Demostración.** Puedo describir el suceso  $\{X \leq x\}$  como unión numerable de sucesos

$$\{\omega \in \Omega : X(\omega) \leq x\} = \bigcup_{y \in X(\Omega)} \{\omega \in \Omega : X(\omega) = y\}$$

Como la unión numerable de sucesos es un suceso,  $\{X \leq x\} \in \mathcal{F}$ . ■

**Definición 3.2 (Función de distribución).** Sea  $X$  una variable aleatoria,  $F_X: \mathbb{R} \rightarrow [0, 1]$  es su función de distribución  $\iff \forall x \in X(\Omega) : \boxed{F_X(x) = P(X \leq x)}$

Sea  $X$  una v.a.d. que toma los valores  $x_1, x_2, \dots$  con probabilidades  $p_1, p_2, \dots$  y con función de masa  $p_X \implies F_X(x) = \sum_{x_i \leq x} P(X = x_i)$ . Es decir, es la función de masa acumulada.

**Lema 3.1.** En  $(\Omega, \mathcal{F}, P)$ , sea  $X$  una variable aleatoria con función de distribución  $F_X$ .

1.  $\lim_{x \rightarrow \infty} F_X(x) = 1 \wedge \lim_{x \rightarrow -\infty} F_X(x) = 0$
- $\implies$  2.  $F_X$  es no decreciente
3.  $F_X$  es continua por la derecha

**Demostración.**

1. Definimos  $A_n := \{\omega \in \Omega : X(\omega) \leq n\}$  que es creciente según  $n \rightarrow \infty$ .

$$\implies \lim_{n \rightarrow \infty} F_X(n) = \lim_{n \rightarrow \infty} P(X \leq n) = \lim_{n \rightarrow \infty} P\left(\bigcup_{j=1}^n A_j\right) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(\Omega) = 1$$

2.  $x < y \implies \{X \leq x\} \subseteq \{X \leq y\} \implies F_X(x) \leq F_X(y)$ .

3. Definimos  $A_h := \{\omega \in \Omega : X(\omega) \leq x + h\}$ .

$$\implies \lim_{h \rightarrow 0^+} F_X(x + h) = \lim_{h \rightarrow 0^+} P(A_h) = P\left(\lim_{h \rightarrow 0^+} A_h\right) = P(A_0) = P(X \leq x) = F_X(x)$$

■

**Teorema 3.1.** Sea  $F: U \subset \mathbb{R} \rightarrow [0, 1]$  una función que cumpla los puntos del lema anterior.

$$\implies \exists! X \text{ variable aleatoria} : F_X = F$$

**Demostración.** Suponemos que  $\exists X, Y: \Omega \rightarrow \mathbb{R}$  v.a. tales que  $F_X = F = F_Y$ .

$$\implies \forall x \in X(\Omega) = U = Y(\Omega) : P(X \leq x) = F_X(x) = F(x) = F_Y(x) = P(Y \leq x)$$

Por tanto,  $\forall x \in U : P(X \leq x) = P(Y \leq x) \implies X = Y$ . ■

**Moraleja:** Una variable aleatoria queda determinada por su función de distribución.

13/03/2024

Sea  $X$  una variable aleatoria con función de distribución  $F_X$  y  $a, b \in \mathbb{R}$ .

$$\implies P(a < X \leq b) = P(\{x \leq b\} \setminus \{x \leq a\}) = P(x \leq b) - P(x \leq a) = F_X(b) - F_X(a)$$

**Definición 3.3 (Variable aleatoria continua).** Sea  $X$  una variable aleatoria con función de distribución  $F_X$ ,  $X$  es continua (v.a.c.)

$$\iff \exists f_X: \mathbb{R} \rightarrow \mathbb{R} \geq 0 : \left( F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(y) dy \right) \wedge \left( \int_{-\infty}^{\infty} f_X(y) dy = 1 \right)$$

$f_X$  se denomina la **función de densidad** de  $X$ .

**Observación 3.1.**

$$1. \forall a \in \mathbb{R} : P(X = a) = 0$$

**Demostración.** Por continuidad de la probabilidad:

$$\begin{aligned} \overline{P(X = a)} &= \lim_{h \rightarrow 0^+} P(a - h < X \leq a + h) = \lim_{h \rightarrow 0^+} F_X(a + h) - F_X(a - h) \\ &= \lim_{h \rightarrow 0^+} \left( \int_{-\infty}^a f_X(y) dy - \int_{-\infty}^{a-h} f_X(y) dy \right) = \lim_{h \rightarrow 0^+} \int_{a-h}^{a+h} f_X(y) dy = \overline{0} \end{aligned}$$
■

$$2. \text{ Cálculo de probabilidades: } \forall a \leq b : P(a \leq X \leq b) = \int_a^b f_X(y) dy$$

$$3. \forall x \in \mathbb{R} : f_X(x) = \begin{cases} \frac{d}{dx} F_X(x), & \text{si } F_X \text{ es derivable en } x \\ 0, & \text{en otro caso} \end{cases}$$

**Ejemplo 3.1.**

$$1. \text{ Para cualquier } f \geq 0 \text{ tal que } \int_{-\infty}^{\infty} f(x) dx = c \in \mathbb{R} \text{ tenemos una v.a.c.}$$

$$2. X \sim U(0, 1) \iff f_X(u) = \begin{cases} 1, & \text{si } u \in [0, 1] \\ 0, & \text{en otro caso} \end{cases} \implies F_X(u) = \begin{cases} 0, & \text{si } u < 0 \\ u, & \text{si } u \in [0, 1] \\ 1, & \text{si } u > 1 \end{cases}$$

$$3. X \sim \text{EXP}(\lambda) \iff f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0 \\ 0, & \text{en otro caso} \end{cases} = \lambda e^{-\lambda x} \cdot \mathbb{1}_{\{x \geq 0\}}(x) \\ \implies \forall x > 0 : F_X(x) = \int_0^x \lambda e^{-\lambda y} dy = [-e^{-\lambda y}]_0^x = 1 - e^{-\lambda x} \implies \lim_{x \rightarrow \infty} F_X(x) = 1$$

$$4. X \sim N(\mu, \sigma^2) \iff f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\text{La guay es } X \sim N(0, 1) \iff \phi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \text{ Veamos que } \int_{-\infty}^{\infty} \phi(x) dx = 1.$$

**Demostración.**

$$\begin{aligned} I &:= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \implies I^2 = \left( \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right) \left( \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \right) \\ \implies I^2 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dx dy = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left[ -e^{-\frac{r^2}{2}} \right]_0^{\infty} d\theta = \frac{1}{2\pi} \int_0^{2\pi} 1 d\theta = 1 \end{aligned}$$

■

18/03/2024

### 3.1 Funciones / Transformaciones de v.a.c.

Sea  $X$  una v.a.c. con función de densidad  $f_X$  y  $g: X(\Omega) \rightarrow \mathbb{R}$  una función real y  $Y := g(X)$ :

- $Y$  es variable aleatoria  $\iff \forall y \in \mathbb{R} : \{\omega \in \Omega : g(X(\omega)) \leq y\} \in \mathcal{F}$ .  
Esto es cierto para las  $g$  “habituales” (continuas, monótonas, etc.), para más detalle, hay que esperar a teoría de la medida.
- ¡Cuidado!  $Y$  puede no ser continua.
- $Y$  v.a.  $\implies Y$  tiene función de distribución  $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = ???$

**Ejemplo 3.2.**  $g(x) = ax + b$  con  $a \neq 0 \wedge b \in \mathbb{R}$ .

$$\implies F_Y(y) = P(aX + b \leq y) = \begin{cases} P\left(X \leq \frac{y-b}{a}\right), & \text{si } a > 0 \\ P\left(X \geq \frac{y-b}{a}\right) = 1 - P\left(X < \frac{y-b}{a}\right), & \text{si } a < 0 \end{cases}$$

$$\text{Para } X \text{ v.a. continua, } F_Y(y) = \begin{cases} F_X\left(\frac{y-b}{a}\right), & \text{si } a > 0 \\ 1 - F_X\left(\frac{y-b}{a}\right) & \text{si } a < 0 \end{cases}.$$

$$\implies f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \frac{1}{a} f_X\left(\frac{y-b}{a}\right), & \text{si } a > 0 \\ -\frac{1}{a} f_X\left(\frac{y-b}{a}\right), & \text{si } a < 0 \end{cases} = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

**Teorema 3.2.** En  $(\Omega, \mathcal{F}, P)$ , sea  $X$  una v.a.c. con función de densidad  $f_X$  y  $g: \mathbb{R} \rightarrow \mathbb{R}$  una función diferenciable y estrictamente creciente.

$$\implies f_Y(y) = f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))}$$

De manera similar, si  $g$  es estrictamente decreciente  $\implies f_Y(y) = -f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))}$

**Demostración.**

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

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) = f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))}$$

En el caso de  $g$  decreciente tenemos:

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

$$f_Y(y) = -\frac{d}{dy} F_X(g^{-1}(y)) = -f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) = -f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))}$$

■

20/03/2024

**Ejercicio 3.1.1.** Sea  $X$  una variable aleatoria con función de densidad  $f_X(x)$  y  $Y = g(X)$  con  $g(x) = x^2$ . ¿Cuál es la función de densidad de  $Y$ ?

$$F_Y(y) = P(X^2 \leq y) = \begin{cases} y < 0 \implies 0 \\ y \geq 0 \implies P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{cases}$$

$$\implies f_Y(y) = \begin{cases} y \leq 0 \implies 0 \\ y > 0 \implies \frac{d}{dy} F_X(\sqrt{y}) - \frac{d}{dy} F_X(-\sqrt{y}) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) \end{cases}$$

**Ejemplo 3.3.** Sea  $X \sim N(\mu, \sigma^2) \implies \forall x \in \mathbb{R} : f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$  e  $Y = e^X$  (se denomina lognormal). Calculamos la derivada y la inversa de  $g(x) = e^x$ .

$$\implies g'(x) = e^x \quad \wedge \quad g^{-1}(y) = \ln y \quad \wedge \quad \frac{d}{dy} g^{-1}(y) = \frac{1}{y}$$

Como  $g$  es estrictamente creciente, aplicamos el teorema anterior:

$$f_Y(y) = \begin{cases} 0, & \text{si } y \leq 0 \\ \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{y} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}}, & \text{si } y > 0 \end{cases}$$

**Ejemplo 3.4.** Sea  $X \sim \text{EXP}(\lambda)$ ,  $\lambda > 0$  y  $Y = 3X + 2$ . Calculamos la función de densidad de  $Y$ .

$$g'(x) = 3 \quad \wedge \quad \frac{d}{dy} g^{-1}(y) = \frac{d}{dy} \left( \frac{y-2}{3} \right) = \frac{1}{3}$$

Como  $g$  es estrictamente creciente, aplicamos el teorema anterior:

$$f_Y(y) = f_X \left( \frac{y-2}{3} \right) \cdot \frac{1}{3} = \begin{cases} 0, & \text{si } y < 2 \\ \frac{1}{3} \lambda e^{-\lambda \frac{y-2}{3}}, & \text{si } y \geq 2 \end{cases}$$

## 3.2 Esperanzas de v.a.c.

**Definición 3.4 (Esperanza).** Sea  $X$  una v.a.c. con función de densidad  $f_X$ .  $E(X)$  es la esperanza de  $X \iff \boxed{E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) dx}$

Siempre que haya convergencia absoluta  $\left( \iff \int_{-\infty}^{\infty} |x| \cdot f_X(x) dx < \infty \right)$ .

**Teorema 3.3.** Sea  $X$  una v.a.c. con función de densidad  $f_X$  y  $g: \mathbb{R} \rightarrow \mathbb{R}$  una función diferenciable y estrictamente creciente  $\implies E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx$

**Demostración.**

$$E(g(X)) = \int_{-\infty}^{\infty} y \cdot f_Y(y) dy = \int_{-\infty}^{\infty} y \cdot f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))} dy$$

Por el cambio de variable  $y = g(x) \implies g^{-1}(y) = x \wedge dy = g'(x) dx$ .

$$\implies E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \frac{1}{g'(x)} \cdot g'(x) dx = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx$$

■

**Ejemplo 3.5.**

1. Sea  $X \sim \text{UNIF}(a, b)$ , entonces  $f_X(x) = \frac{1}{b-a} \cdot \mathbb{1}_{[a,b]}(x)$ .

$$\implies \boxed{E(X)} = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \left[ \frac{x^2}{2} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \boxed{\frac{a+b}{2}}$$

$$\begin{aligned}
\Rightarrow \overline{V(X)} &= E(X^2) - E(X)^2 = \int_a^b x^2 \cdot \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \frac{1}{b-a} \left[\frac{x^3}{3}\right]_a^b - \left(\frac{a+b}{2}\right)^2 \\
&= \frac{b^3 - a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2 = \frac{b^2 + ab + a^2}{3} - \frac{a^2 + 2ab + b^2}{4} \\
&= \frac{4b^2 - 4ab + 4a^2 - 3a^2 - 6ab - 3b^2}{12} = \frac{b^2 - 2ab + a^2}{12} = \overline{\frac{(b-a)^2}{12}}
\end{aligned}$$

2. Sea  $X \sim \text{EXP}(\lambda)$ ,  $\lambda > 0$ , entonces  $f_X(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}_{[0, \infty)}(x)$ .

$$\begin{aligned}
\Rightarrow \overline{E(X)} &= \int_0^\infty x \cdot \lambda e^{-\lambda x} dx = [-x e^{-\lambda x}]_0^\infty + \int_0^\infty e^{-\lambda x} dx = 0 + \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^\infty = \overline{\frac{1}{\lambda}} \\
\Rightarrow \overline{V(X)} &= E(X^2) - E(X)^2 = \int_0^\infty x^2 \cdot \lambda e^{-\lambda x} dx - \left(\frac{1}{\lambda}\right)^2 = \overline{\frac{1}{\lambda^2}} \text{ porque} \\
&\int_0^\infty x^2 \cdot \lambda e^{-\lambda x} dx = [-x^2 e^{-\lambda x}]_0^\infty + 2 \int_0^\infty x e^{-\lambda x} dx = 0 + 2 \cdot \frac{1}{\lambda^2} = \frac{2}{\lambda^2}
\end{aligned}$$

3. Sea  $X \sim N(0, 1)$ , entonces  $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ .

$$\begin{aligned}
\overline{E(X)} &= \int_{-\infty}^\infty x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \overline{0} \text{ (integrando impar en regi3n centrada en el origen).} \\
\overline{V(X)} &= E(X^2) - E(X)^2 = \int_{-\infty}^\infty x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty x^2 \cdot e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \left( \left[-x e^{-\frac{x^2}{2}}\right]_{-\infty}^\infty + \int_{-\infty}^\infty e^{-\frac{x^2}{2}} dx \right) = \frac{1}{\sqrt{2\pi}} (0 + \sqrt{2\pi}) = \overline{1}
\end{aligned}$$

4. Sea  $X \sim N(\mu, \sigma^2)$ , entonces  $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ .

$$E(X) = \int_{-\infty}^\infty x \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Con el cambio de variable  $t = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma t + \mu \wedge dx = \sigma dt$ .

$$\begin{aligned}
\Rightarrow \overline{E(X)} &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty (\sigma t + \mu) e^{-\frac{t^2}{2}} \sigma dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty (\sigma t + \mu) e^{-\frac{t^2}{2}} dt \\
&= \frac{1}{\sqrt{2\pi}} \left( \sigma \int_{-\infty}^\infty t e^{-\frac{t^2}{2}} dt + \mu \int_{-\infty}^\infty e^{-\frac{t^2}{2}} dt \right) = \frac{1}{\sqrt{2\pi}} (0 + \mu \sqrt{2\pi}) = \overline{\mu} \\
\Rightarrow V(X) &= E[(X - E(X))^2] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx
\end{aligned}$$

Con el cambio de variable  $t = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma t + \mu \wedge dx = \sigma dt$ .

$$V(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^\infty (\sigma t)^2 e^{-\frac{t^2}{2}} \sigma dt = \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^\infty t^2 e^{-t^2} dt = \frac{4\sigma^2}{\sqrt{\pi}} \int_0^\infty t^2 e^{-t^2} dt$$

Consideramos  $\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt$  y el cambio de variable  $t^2 = u \implies 2t dt = du$ .

$$\begin{aligned} \implies \overline{V(X)} &= \frac{4\sigma^2}{\sqrt{\pi}} \int_0^\infty u e^{-u} \frac{1}{2\sqrt{u}} du = \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty u^{\frac{3}{2}-1} e^{-u} du = \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2} \cdot \Gamma\left(\frac{1}{2}\right) = \frac{\sigma^2}{\cancel{\sqrt{\pi}}} \cdot \cancel{\sqrt{\pi}} = \overline{\sigma^2} \end{aligned}$$

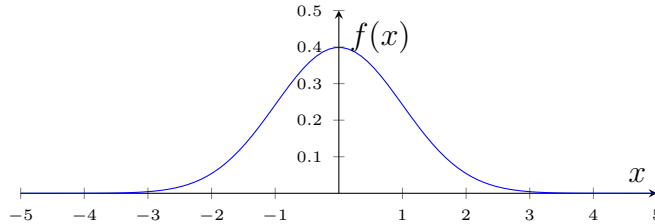
21/03/2024

### 3.2.1 Calculando con la normal ( $N(\mu, \sigma^2)$ )

Sea  $X \sim N(\mu, \sigma^2)$  con parámetros  $\mu \in \mathbb{R}$ ,  $\sigma^2 > 0$

$$\implies f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \wedge \quad F(x) = \int_{-\infty}^x f(y) dy$$

El caso particular de  $N(0, 1)$  se denota por  $\phi(x) := f(x)$  y  $\Phi(x) := F(x)$ .



¡Basta con  $N(0, 1)$ !

$$X \sim N(0, 1) \implies Y := \sigma X + \mu \sim N(\mu, \sigma^2)$$

$$Y \sim N(\mu, \sigma^2) \implies X := \frac{Y - \mu}{\sigma} \sim N(0, 1)$$

Podemos tipificar cualquier v.a.c.  $X$  con esperanza  $E(X)$  y varianza  $V(X)$ .

$$Y := \frac{X - E(X)}{\sqrt{V(X)}} \implies E(Y) = 0 \quad \wedge \quad V(Y) = 1$$

¿Qué cálculos queremos hacer con  $Y \sim N(\mu, \sigma^2)$ ?

$$E(Y) = E(\mu + \sigma X) = \mu + \sigma E(X) = \mu \quad \wedge \quad V(Y) = V(\mu + \sigma X) = \sigma^2 V(X) = \sigma^2$$

$$E(Y^7) = E((\mu + \sigma X)^7) = \sum_{k=0}^7 \binom{7}{k} \mu^7 \sigma^{7-k} E(X^{7-k})$$

$$\text{Caso particular de } X \sim N(0, 1) \implies E(X^k) = \begin{cases} 0, & \text{si } k \text{ es impar} \\ \frac{1}{2^{k/2}} \frac{k!}{(k/2)!} = \frac{(k-1)!!}{k!!}, & \text{si } k \text{ es par} \end{cases}$$

¿Cómo calculamos las probabilidades de  $X \sim N(0, 1)$ ?

$$\forall a, b \in \mathbb{R} : P(a \leq X \leq b) = \Phi(b) - \Phi(a) \quad (\text{se calcula numéricamente})$$



$$\forall a \in \mathbb{R} : P(|X| \leq a) = P(-a \leq X \leq a) = \Phi(a) - \Phi(-a) = 2\Phi(a) - 1$$

¿Qué hago si me dan  $P(X \leq a)$  y me piden  $a$ ?

$\Phi(a) = P(X \leq a) \implies a = \Phi^{-1}(P(X \leq a))$  que también se hace numericamente.

**Observación 3.2.**  $\text{erf}(x) = \frac{2}{\sqrt{2}} \int_0^x e^{-y^2} dy \implies \text{erf}(x) = 2\Phi(\sqrt{2}x) - 1$

Para  $X \sim N(\mu, \sigma^2)$ , se tiene  $P(Y \leq a) = P(\mu + \sigma X \leq a) = P\left(X \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right)$ .

**Un famoso resultado: El percentil 95**

$$P(|X| \leq a) = \frac{95}{100} \implies 2\Phi(a) - 1 = \frac{95}{100} \implies a = \Phi\left(\frac{1 + 0.95}{2}\right) \approx 1.96$$

02/04/2024

### 3.3 Modelos multidimensionales (vectores aleatorios)

**Definición 3.5 (Función de distribución conjunta).** Sean  $X$  e  $Y$  dos variables aleatorias,  $F_{X,Y}$  es su función de distribución conjunta

$$\iff \forall (x, y) \in \mathbb{R}^2 : F_{X,Y}(x, y) = P(X \leq x \wedge Y \leq y)$$

**Observación 3.3.** •  $\lim_{x, y \rightarrow \infty} F_{X,Y}(x, y) = 1$  •  $\forall \hat{x} \leq x, \hat{y} \leq y : F_{X,Y}(\hat{x}, \hat{y}) \leq F_{X,Y}(x, y)$

Funciones de densidad marginales:  $\forall x \in \mathbb{R} : F_X(x) = P(X \leq x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y)$

¿Independencia de  $X$  e  $Y$ ?  $X, Y$  independientes  $\iff F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)$

**Definición 3.6 (Función de densidad conjunta).** Sean  $X$  e  $Y$  dos variables aleatorias,  $f_{X,Y}$  es su función de densidad conjunta

$$\iff \forall (x, y) \in \mathbb{R}^2 : F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) du dv$$

**Observación 3.4.** •  $\forall (x, y) \in \mathbb{R}^2 : f_{X,Y}(x, y) \geq 0$  •  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(u, v) du dv = 1$

Cálculo de probabilidades:  $\forall A \subset \mathbb{R}^2 : P((X, Y) \in A) = \iint_A f_{X,Y}(x, y) dx dy$

Si  $F_{X,Y}(x, y)$  es el dato,  $f_{X,Y}(x, y) = \begin{cases} \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) & \text{si la derivada existe} \\ 0 & \text{si no existe} \end{cases}$

**Ejemplo 3.6.** Sean  $a, b \in \mathbb{R}$  y  $(X, Y) \sim \text{UNIF}(\mathcal{R} := [0, a] \times [0, b])$

$$\begin{aligned} \implies f_{X,Y} &= \begin{cases} 0, & \text{si } (x, y) \notin \mathcal{R} \\ \frac{1}{ab}, & \text{si } (x, y) \in \mathcal{R} \end{cases} \\ \implies \forall A \subset \mathcal{R} : P((x, y) \in A) &= \iint_A \frac{1}{ab} dx dy = \frac{1}{ab} \text{Área}(A) \end{aligned}$$

03/04/2024

**Ejemplo 3.7 (Normal bidimensional).** Sean  $X, Y$  dos variables aleatorias con función de densidad conjunta dependiente de cinco parámetros:  $\mu_1 \in \mathbb{R}$ ,  $\mu_2 \in \mathbb{R}$ ,  $\sigma_1 > 0$ ,  $\sigma_2 > 0$ ,  $\rho \in (-1, 1)$ ;  $(X, Y)$  siguen una distribución normal bidimensional

$$\iff f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)}$$

Para  $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1$  se tiene la normal bidimensional estándar (dependiente de  $\rho$ )

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)}$$

**Notación matricial**

$$\begin{aligned} \vec{\mu} &= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \mathbb{X} = \begin{pmatrix} X \\ Y \end{pmatrix} \quad \Sigma := \begin{pmatrix} \sigma^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \implies \Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \\ \implies f_{\mathbb{X}}(\vec{x}) &= \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1}(\vec{x} - \vec{\mu})} \end{aligned}$$

### 3.3.1 Normal multidimensional

$\mathbb{X} = (X_1, \dots, X_n)$   $\vec{\mu} = (\mu_1, \dots, \mu_n)$   $\Sigma$  simétrica definida positiva  $n \times n$

$$\implies f_{\mathbb{X}}(\vec{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1}(\vec{x} - \vec{\mu})}$$

### 3.3.2 Marginales e independencia

$$\forall x \in \mathbb{R} : f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \quad \wedge \quad \forall y \in \mathbb{R} : f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

**Teorema 3.4.** Sean  $X, Y$  dos variables aleatorias con función de densidad conjunta  $f_{X,Y}$   
 $X$  e  $Y$  independientes  $\iff \exists h, g : \forall (x, y) \in \mathbb{R} : f_{X,Y}(x, y) = h(x) \cdot g(y)$

**Ejemplo 3.8.** 1.  $f(x, y) = \begin{cases} e^{-x-y} & \text{si } x, y \geq 0 \\ 0 & \text{en otro caso} \end{cases} \implies f(x, y) = e^{-x-y} \cdot \mathbb{1}_{\{x>0 \wedge y>0\}}(x, y)$   
 $\implies f(x, y) = (e^{-x} \cdot \mathbb{1}_{\{x \geq 0\}}(x)) \cdot (e^{-y} \cdot \mathbb{1}_{\{y \geq 0\}}(y)) \implies$  son independientes

2.  $f(x, y) = \begin{cases} 2e^{-x-y} & \text{si } 0 < x < y \\ 0 & \text{en otro caso} \end{cases} \implies f(x, y) = 2e^{-x-y} \cdot \mathbb{1}_{\{0 < x < y\}}(x, y)$

Por tanto, no son independientes porque si lo fueran, tendríamos

$$P(c < X < d \wedge b < Y < a) = P(c < X < d) \cdot P(b < Y < a)$$

04/04/2024

### 3.4 Condicionando

Sean  $X$  e  $Y$  dos variables aleatorias con función de densidad conjunta  $f_{X,Y}$  y  $f_X(x)$ .

En general  $\forall A, B \subset \mathbb{R}^2 : P((X, Y) \in A | (X, Y) \in B) = \frac{P((X, Y) \in A \wedge (X, Y) \in B)}{P((X, Y) \in B)}$  con  $P((X, Y) \in B) > 0$ . Sin embargo a veces la información “nueva”  $((X, Y) \in B)$  es muy precisa, por ejemplo  $X = 3$ , entonces la fórmula no vale porque  $P(X = 3) = 0$ .

**Definición 3.7 (Función de densidad condicionada).** Sean  $X, Y$  dos variables aleatorias con función de densidad conjunta  $f_{X,Y}$  y funciones de densidad marginales  $f_X(x)$  y  $f_Y(y)$ .

- $f_{Y|X}(y|x) = f_{Y|X=x}(y)$  es la función de densidad de  $Y$  condicionada a  $X = x$

$$\iff f_{Y|X=x}(y) := \frac{f_{X,Y}(x, y)}{f_X(x)} \quad (\text{allá donde } f_X(x) > 0)$$

- $f_{X|Y}(x|y) = f_{X|Y=y}(x)$  es la función de densidad de  $X$  condicionada a  $Y = y$

$$\iff f_{X|Y=y}(x) := \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad (\text{allá donde } f_Y(y) > 0)$$

Una comprobación necesaria es que  $\int_{-\infty}^{\infty} f_{Y|X=x}(y) dy = 1$  y  $\int_{-\infty}^{\infty} f_{X|Y=y}(x) dx = 1$ :

$$\int_{-\infty}^{\infty} f_{Y|X=x}(y) dy = \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \frac{1}{f_X(x)} f_X(x) = 1$$

**Ejemplo 3.9.**  $f_{X,Y}(x, y) = \begin{cases} 2e^{-x}e^{-y} & \text{si } 0 < x < y \\ 0 & \text{en otro caso} \end{cases}$

$$\begin{aligned} \implies \forall x > 0 : f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_x^{\infty} 2e^{-x}e^{-y} dy = 2e^{-x} \int_x^{\infty} e^{-y} dy \\ &= 2e^{-x}e^{-x} = 2e^{-2x} \\ \forall y > 0 : \overline{f_Y(y)} &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = \int_0^y 2e^{-x}e^{-y} dx = 2e^{-y} \int_0^y e^{-x} dx = \overline{2e^{-y}(1 - e^{-y})} \\ f_{Y|X=3}(y) &= \frac{f_{X,Y}(3, y)}{f_X(3)} = \frac{2e^{-3}e^{-y}}{2e^{-6}} = e^{3-y} \cdot \mathbb{1}_{\{y>3\}}(y) \\ \implies f_{X|Y=y}(x) &= \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{2e^{-x}e^{-y}}{2e^{-y}(1 - e^{-y})} = \frac{e^{-x}}{1 - e^{-y}} \cdot \mathbb{1}_{\{0 < x < y\}}(x) \end{aligned}$$

¡Cuidado! No se nos puede olvidar el soporte. (Motivo de excomuni3n)

**Nota sobre independencia:**  $X, Y$  indep.  $\implies f_{X|Y=y}(x) = f_X(x) \wedge f_{Y|X=x}(y) = f_Y(y)$

### 3.5 Transformaciones / cambio de variables

Sean  $X, Y$  dos v.a. con funci3n de densidad conjunta  $f_{X,Y} \wedge T: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$   
 $(x, y) \longmapsto (u(x, y), v(x, y))$

Definimos las variables aleatorias  $U := u(X, Y)$  y  $V := v(X, Y)$ .

Por ejemplo, si tenemos  $T$  lineal con matriz asociada  $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies \begin{cases} U = aX + cY \\ V = bX + dY \end{cases}$   
pero tambi3n hay transformaciones no lineales como  $U = X + Y \wedge V = \frac{X}{X+Y}$ .

Definimos  $D := \{(x, y) \in \mathbb{R}^2 : f_{X,Y}(x, y) > 0\}$  y sea  $B \subset T(D)$ .

$$\begin{aligned} \implies P((U, V) \in B) &= \iint_B f_{U,V}(u, v) du dv = \iint_{T^{-1}(B)} f_{X,Y}(x, y) dx dy \\ &= \iint_B f_{X,Y}(x(u, v), y(u, v)) |J(u, v)| du dv \end{aligned}$$

**Teorema 3.5.** Sean  $X, Y$  dos variables aleatorias con funci3n de densidad conjunta  $f_{X,Y}$  y sea  $T: \mathbb{R}^2 \longrightarrow \mathbb{R}^2 : (x, y) \mapsto T(x, y) = (u(x, y), v(x, y))$  una biyecci3n de  $D$  en  $T(D)$  de clase  $C^1$  con inversa de clase  $C^1$ .

$$\implies f_{U,V}(u, v) = \begin{cases} f_{X,Y}(x(u, v), y(u, v)) |J(u, v)| & \text{si } (u, v) \in T(D) \\ 0 & \text{en otro caso} \end{cases}$$

09/04/2024

**Ejemplo 3.10.** Sean  $X, Y$  v.a. con  $\forall x, y > 0 : f_{X,Y}(x, y) := \frac{1}{4}e^{-\frac{1}{2}(x+y)}$ .

Sea  $T(x, y) = (\frac{x-2}{2}, y)$ . Definimos  $(U, V) := T(X, Y)$ .

$$\implies \forall x, y > 0 : T(x, y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix} \implies \begin{cases} x = 2u + 2 \\ y = v \end{cases}$$

$$\implies T^{-1}(u, v) = (2u + 2, v) \wedge J(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} = 2$$

$$\implies f_{U,V}(u, v) = \begin{cases} 2\frac{1}{4}e^{-\frac{1}{2}[2u+v+v]} = \frac{1}{2}e^{-(u+v)} & \text{si } v > 0 \wedge u > -\frac{v}{2} \\ 0 & \text{en otro caso} \end{cases}$$

**Ejemplo 3.11.** Sean  $X, Y$  v.a. con  $\forall x, y > 0 : f_{X,Y}(x, y) := e^{-x-y}$ .

Sea  $T(x, y) = \left(x + y, \frac{x}{x + y}\right)$ . Definimos  $(U, V) := T(X, Y)$ .

$$\implies T^{-1}(u, v) = (u \cdot v, u(1 - v)) \wedge J(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 1 - v & -u \end{vmatrix} = -u$$

$$\implies f_{U,V}(u, v) = \begin{cases} |-u| e^{-u \cdot v - u(1-v)} = ue^{-u} & \text{si } u > 0 \wedge 0 < v < 1 \\ 0 & \text{en otro caso} \end{cases}$$

$$\implies \forall v \in (0, 1) : f_V(v) = \int_{-\infty}^{\infty} f_{U,V}(u, v) du = \int_0^{\infty} ue^{-u} du = [-e^{-u}(u + 1)]_0^{\infty} = 1$$

$$\implies V \sim \text{UNIF}(0, 1)$$

10/04/2024

Sean  $X, Y$  dos variables con función de densidad conjunta  $f_{X,Y}$ , consideramos una función  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$  “razonable”<sup>(\*)</sup> y definimos  $Z := g(X, Y)$ . Queremos calcular  $E(Z), V(Z)$ .

(\*)  $g$  cumple que  $\forall z \in \mathbb{R} : \{\omega \in \Omega : g(X(\omega), Y(\omega)) \leq z\} \in \mathcal{F}$

**Teorema 3.6.** Sean  $X, Y$  dos variables aleatorias con  $f_{X,Y}$  y  $Z := g(X, Y)$  con  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$

$$\implies E(Z) = \int_{-\infty}^{\infty} z \cdot f_Z(z) dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f_{X,Y}(x, y) dx dy$$

**Observación 3.5.** 1.  $\forall \alpha, \beta \in \mathbb{R} : E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$

$$2. E(X \cdot Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot f_{X,Y}(x, y) dx dy$$

$$\text{cov}(X, Y) = E(X \cdot Y) - E(X)E(Y) \quad \wedge \quad \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{V(X)V(Y)}}$$

$$\implies X, Y \text{ indep} \not\iff \rho(X, Y) = 0$$

3.  $X, Y$  indep.  $\iff f_{X,Y}$  se factoriza

$$\implies X, Y \text{ indep.} \iff \forall g, h: \mathbb{R} \longrightarrow \mathbb{R} : E(g(X)h(Y)) = E(g(X))E(h(Y))$$

4. Esperanza condicionada y total

$$E(Y|X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X=x}(y) dy = \int_{-\infty}^{\infty} y \cdot \frac{f_{X,Y}(x, y)}{f_X(x)} dy$$

$$\implies E(Y) = \int_{-\infty}^{\infty} E(Y|X = x) \cdot f_X(x) dx$$

**¿Y qué hay de  $f_Z(z)$ ?** Definimos  $A_g(z) := \{(x, y) \in \mathbb{R}^2 : g(x, y) \leq z\}$

$$\forall z \in \mathbb{R} : F_Z(z) = P(Z \leq z) = P(g(X, Y) \leq z) = P((X, Y) \in A_g(z)) = \iint_{A_g(z)} f_{X,Y}(x, y) dx dy$$

$$\implies \forall z \in \mathbb{R} : f_Z(z) = \begin{cases} \frac{\partial}{\partial z} F_Z(z) & \text{si la derivada existe} \\ 0 & \text{en otro caso} \end{cases}$$

**Ejemplo 3.12 (El caso de la suma).** Sean  $X, Y$  dos v.a. con  $f_{X,Y}$  y  $Z := X + Y$ . Entonces  $A_+(z) := \{(x, y) \in \mathbb{R}^2 : x + y \leq z\}$ .

$$\implies \forall z \in \mathbb{R} : F_Z(z) = \iint_{A_+(z)} f_{X,Y}(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x, y) dy dx$$

Mediante el cambio de variables  $u = x \wedge v = x + y$  se tiene

$$T(x, y) = (x, x + y) \implies T^{-1}(u, v) = (u, v - u) \implies J(u, v) = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

$$\implies F_Z(z) = \int_{-\infty}^z \int_{-\infty}^{\infty} f_{X,Y}(u, v-u) du dv \implies \boxed{f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(u, z-u) du}$$

**Caso particular:** Si  $X, Y$  independientes  $\implies f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$

$$\implies \forall z \in \mathbb{R} : \boxed{f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(u) \cdot f_Y(z-u) du =: (f_X * f_Y)(z)}$$

11/04/2024

### 3.6 Convolución

Sean  $X$  e  $Y$  dos v.a.d. indep. que toman valores en  $\mathbb{Z}_{\geq 0}$ . Sea  $Z = X + Y$ .

$$\implies P(Z = k) = \sum_{j=0}^k P(X = j \wedge Y = k - j) = \sum_{j=0}^k P(X = j) \cdot P(Y = k - j)$$

Ahora, sean  $X$  e  $Y$  dos v.a.c. indep. con funciones de densidad  $f_X$  y  $f_Y$  respectivamente.

$$\implies \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x) dx = (f_X * f_Y)(z)$$

**Ejemplo 3.13.** Sean  $X, Y \sim N(0, 1)$  independientes. Queremos  $f_Z$  con  $Z := X + Y$ .

$$\begin{aligned} \forall z \in \mathbb{R} : f_Z(z) &= \int_{-\infty}^{\infty} f_X(u) \cdot f_Y(z-u) du = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-u)^2}{2}} du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[z^2+2u^2-2uz]} du \stackrel{(*)}{=} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{4}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\sqrt{2}u - \frac{z}{\sqrt{2}}\right)^2} du \end{aligned}$$

$$(*) \text{ porque } -\frac{1}{2}(z^2 + 2u^2 - 2uz) = -\frac{1}{2}\left(2u^2 - 2uz + \frac{z^2}{2} + \frac{z^2}{2}\right) = -\frac{1}{2}\left[\left(\sqrt{2}u - \frac{z}{\sqrt{2}}\right)^2 + \frac{z^2}{2}\right].$$

Por el cambio de variables  $w = \sqrt{2}u - \frac{z}{\sqrt{2}} \implies dw = \sqrt{2} du$ , se tiene

$$\forall z \in \mathbb{R} : f_Z(z) = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{4}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{w^2}{2}} dw \stackrel{1}{=} \frac{1}{2\sqrt{\pi}} e^{-\frac{z^2}{4}} \implies Z \sim N\left(0, (\sqrt{2})^2\right)$$

**Moraleja:** la suma de normales independientes es una normal.

**Ejemplo 3.14 (Normal bidimensional).** Sean  $X, Y$  dos variables aleatorias con función

de densidad conjunta  $\forall x, y \in \mathbb{R} : f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)}$ .

$$\implies f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)} dy$$

Por tanto,  $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  y  $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$  (ambas normales estándar).

Ahora veamos que  $\rho(X, Y) := \frac{\text{cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \rho$ .

$$E(X \cdot Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)} dx dy$$

Usando esperanza total,  $E(g(X, Y)) = \int_{-\infty}^{\infty} E(g(X, Y)|X = x) \cdot f_X(x) dx$

$$E(X \cdot Y) = \int_{-\infty}^{\infty} E(X \cdot Y|X = x) \cdot f_X(x) dx = \int_{-\infty}^{\infty} x \cdot E(Y|X = x) \cdot f_X(x) dx$$

Necesitamos calcular  $E(Y|X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X=x}(y) dy$ .

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}} = \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}(y-\rho x)^2}$$

Es decir,  $Y|X = x \sim N(\rho x, 1 - \rho^2)$ .

$$\begin{aligned} \implies E(Y|X = x) &= \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}(y-\rho x)^2} dy = \rho x \\ \implies \overline{E(X \cdot Y)} &= \int_{-\infty}^{\infty} x \cdot \rho x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \rho \underbrace{\int_{-\infty}^{\infty} x^2 f_X(x) dx}_{E(X^2)} = \overline{\rho} \end{aligned}$$

17/04/2024

### 3.7 Fuera de menú

Tenemos  $X_1, X_2$  siguiendo una distribución normal bidimensional con  $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$  conocidos.

$$\implies f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)}$$

Podemos escribir  $X_1$  y  $X_2$  como transformaciones de  $Z_1, Z_2 \sim N(0, 1)$ .

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \sigma_1 Z_1 \\ \sigma_2(\rho Z_1 + \sqrt{1-\rho^2} Z_2) \end{pmatrix}$$

Esta transformación facilita muchos cálculos como  $E(X_1 \cdot X_2)$ .

Tenemos  $\mathbb{X}^T = (X_1, \dots, X_n)$  normal  $n$ -dimensional con  $\vec{\mu} = (\mu_1, \dots, \mu_n)$  y  $\mathbb{V}$  matriz de covarianzas definida positiva.

$$\implies f_{\mathbb{X}}(\vec{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbb{V}}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \mathbb{V}^{-1}(\vec{x}-\vec{\mu})}$$

Si ahora tenemos en cuenta que  $A$  definida positiva  $\iff \exists R$  tal que  $A = R^T R$  con  $B$  no singular, podemos escribir  $\mathbb{V} = U^T U$ .

**Teorema 3.7** (de representación).  $\mathbb{X} \sim N(\vec{\mu}, \mathbb{V}) \iff \mathbb{X} = \vec{\mu} + U \cdot \mathbb{Z}$  con  $\mathbb{Z} \sim N(\vec{0}, I)$



## 4 Convergencia de variables aleatorias

Partimos de una sucesión de variables aleatorias en un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$ ,  $(X_n)_{n \in \mathbb{N}}$ , y queremos estudiar las series

$$(S_n)_{n \in \mathbb{N}} = \left( \sum_{i=1}^n X_i \right)_{n \in \mathbb{N}} \quad \wedge \quad (Z_n)_{n \in \mathbb{N}} = \left( \frac{1}{n} \sum_{i=1}^n X_i \right)_{n \in \mathbb{N}}$$

Interesan, específicamente, los límites cuando  $n \rightarrow \infty$  de estas series.

1. Sabemos que significa  $\lim_{n \rightarrow \infty} E(S_n)$  y  $\lim_{n \rightarrow \infty} V(S_n)$ .

Pero, ¿qué significa  $\lim_{n \rightarrow \infty} S_n = S$ ? Requerimos de técnicas más avanzadas que se denominan **modos de convergencia**.

2. Casi siempre asumiremos que  $\{X_n\}_{n \in \mathbb{N}}$  son **independientes e idénticamente distribuidas (i.i.d.)**.

- Todas las  $X_i$  tienen la misma distribución. Por tanto, es habitual definir una  **$X$  de referencia** que tenga la misma distribución que todas las  $X_i$ .
- Las  $X_i$  son completamente independientes, es decir, para todo subconjunto finito  $I \subset \mathbb{N}$ ,  $\{X_i\}_{i \in I}$  son independientes.

3. Descubriremos que, para  $n$  grande y  $X$  de referencia,  $Z_n$  se comporta como  $E(X)$ . Además, veremos el teorema central del límite, que nos dice que  $\frac{Z_n - E(X)}{\sigma(Z_n)/\sqrt{n}} \approx N(0, 1)$ .

### 4.1 Medias y varianzas de las sumas y las medias

$$E(S_n) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = nE(X) \iff \forall i \in \mathbb{N} : E(X_i) = E(X)$$

$$E(Z_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X)$$

$$V(S_n) = V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j) \stackrel{*1}{=} \sum_{i=1}^n V(X_i) \stackrel{*2}{=} n(\sigma(X))^2$$

$$V(Z_n) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \stackrel{*1}{=} \frac{1}{n^2} \sum_{i=1}^n V(X_i) \stackrel{*2}{=} \frac{(\sigma(X))^2}{n} \xrightarrow{n \rightarrow \infty} 0$$

\*<sub>1</sub> Si  $X_i$  incorreladas, \*<sub>2</sub> Si  $X_i$  tienen la misma varianza.

## 4.2 Convergencia cuadrática

**Definición 4.1 (Convergencia cuadrática).** Sean  $(X_n)_{n \in \mathbb{N}}$  v.a. en  $(\Omega, \mathcal{F}, P)$  y  $X$  v.a. en  $(\Omega, \mathcal{F}, P)$ ,  $(X_n)_{n \in \mathbb{N}}$  converge cuadráticamente a  $X$   $\left( X_n \xrightarrow{\text{cuad.}} X \right)$

$$\iff E(|X_n - X|^2) \xrightarrow{n \rightarrow \infty} 0$$

**Teorema 4.1.** Sean  $(X_n)_{n \in \mathbb{N}}$  v.a. i.i.d. en  $(\Omega, \mathcal{F}, P)$  con  $X$  v.a. de referencia

$$\implies Z_n \xrightarrow{\text{cuad.}} E(X)$$

**Demostración.** Si denominamos  $\mu = E(X) \wedge V(X) = \sigma^2$ , entonces

$$V(Z_n) = \frac{\sigma^2}{n} \implies E(|Z_n - \mu|^2) = V(Z_n) \xrightarrow{n \rightarrow \infty} 0$$

■

Quitando hipótesis: basta con que sean incorreladas y tengan la misma media y varianza.

**Teorema 4.2.** Sean  $(X_n)_{n \in \mathbb{N}}$  v.a. en  $(\Omega, \mathcal{F}, P)$  con medias y varianzas  $E(X_i) = \mu_i$  y  $V(X_i) = \sigma_i^2$ . Suponemos que  $\mu_n = \frac{1}{n} \sum_{i=1}^n \mu_i \rightarrow \mu \in \mathbb{R}$  y  $V(Z_n) = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \rightarrow 0$ .

$$Z_n \xrightarrow{\text{cuad.}} \mu$$

**Demostración.**

$$\begin{aligned} E(|Z_n - \mu|^2) &= E(|Z_n - \mu_n + \mu_n - \mu|^2) \\ &= E(|Z_n - \mu_n|^2 + |\mu_n - \mu|^2 + 2(Z_n - \mu_n)(\mu_n - \mu)) \\ &= E((Z_n - \mu_n)^2) + (\mu_n - \mu)^2 + 2(\mu_n - \mu)E(Z_n - \mu_n) \end{aligned}$$

■

## 4.3 Convergencia en probabilidad / Ley débil de los grandes números

J. Bernoulli (1713) *Ars Conjectandi*: Si tienes un dado regular, cuantas más veces lo lances, más se aproximará la frecuencia relativa de un número a su probabilidad.

## 5 Ejercicios

### 5.1 Hoja 1

### 5.2 Hoja 2

**7. b**  $X \sim \text{GEOM}(p) \iff P(X > n+m | X > m) = P(X > n)$

**Solución:** Lo que nos está diciendo la caracterización es que una distribución geométrica no tiene memoria, la probabilidad de no tener éxito en los próximos  $n$  intentos no depende de los intentos anteriores.

**Demostración.** ( $\implies$ ) Suponemos que  $X \sim \text{GEOM}(p)$

$$\implies P(X > n+m | X > m) = \frac{P(X > n+m \wedge X > m)}{P(X > m)} = \frac{P(X > n+m)}{P(X > m)}$$

Como  $P(X > m) = (1-p)^m$  (por eso se llama geométrica), obtenemos

$$\boxed{P(X > n+m | X > m) = \frac{(1-p)^{n+m}}{(1-p)^m} = (1-p)^n = \boxed{P(X > n)}}$$

( $\impliedby$ ) Suponemos que  $P(X > n+m | X > m) = P(X > n)$

$$\implies \frac{P(X > n+m \wedge X > m)}{P(X > m)} = \frac{P(X > n)}{P(X > m)}$$

$$\implies P(X > n+m \wedge X > m) = P(X > n) \cdot P(X > m)$$

■

**12.** Sea  $X$  una v.a.d,  $X \sim \text{BINNEG}(n, p) \iff P(X = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}$ .

Esto significa que  $X$  es la suma de  $n$  v.a.d. independientes, con distribución  $\text{GEOM}(p)$ .

Comprobemos que  $\sum_{k=n}^{\infty} \binom{k-1}{n-1} p^n (1-p)^{k-n} = 1$

$$\sum_{k=n}^{\infty} \binom{k-1}{n-1} p^n (1-p)^{k-n} = \sum_{l=0}^{\infty} \binom{l+n-1}{n-1} p^n (1-p)^l = p^n \sum_{l=0}^{\infty} \binom{l+n-1}{n-1} (1-p)^l$$

Como sabemos que  $\frac{1}{(1-x)^{m+1}} = \sum_{l=0}^{\infty} \binom{l+m}{m} x^l$ , podemos tomar  $x = 1-p$  y  $m = n-1$ :

$$\implies \sum_{k=n}^{\infty} P(X = k) = \frac{p^n}{(1-(1-p))^{n-1+1}} = \frac{p^n}{p^n} = \boxed{1}$$

**20.** Cada día compramos 1 cromo de  $n$  totales que hay, con reposición. ¿Cuántos días esperamos hasta tener todos los cromos?

**Solución:** Sea  $T$  una v.a.d. igual a la cantidad de días hasta que terminamos la colección, queremos calcular  $E(T)$ . Se puede utilizar el modelo de distribución geométrica.

Si definimos  $T_i$  como la cantidad de días que esperamos hasta tener el cromo  $i$ -ésimo nuevo

sabiendo que tienes los  $i - 1$  anteriores, entonces:

$$\begin{aligned} \implies T_1 = 1 \wedge T_2 &\sim \text{GEOM} \left( \frac{n-1}{n} \right) \wedge T_3 \sim \text{GEOM} \left( \frac{n-2}{n} \right) \wedge \dots \\ \implies \forall i \in \mathbb{N}_n : T_i &\sim \text{GEOM} \left( 1 - \frac{i-1}{n} \right) \implies E(T_i) = \frac{n}{n-i} \end{aligned}$$

Además,  $T = T_1 + T_2 + \dots + T_n$ . Por linealidad de la esperanza:

$$\begin{aligned} E(T) &= \sum_{i=1}^n E(T_i) = \sum_{i=0}^{n-1} \frac{n}{n-i} = n \sum_{i=0}^{n-1} \frac{1}{n-i} = n \sum_{k=1}^n \frac{1}{k} = nH_n \sim \ln n - \gamma + O\left(\frac{1}{n}\right) \\ \implies E(T) &= nH_n \approx n \ln n \end{aligned}$$

### 5.3 Hoja 3

8. Sea  $X \sim N(0, 1)$ . Definimos  $Y := e^X$ .  $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$  es la función de densidad de  $X$ .

Queremos calcular  $E(Y)$  y  $V(Y)$ .

$$\begin{aligned} \implies E(Y) &= E(e^X) = \int_{-\infty}^{\infty} e^x \phi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{x-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2x+1}{2}+1} dx \\ &= e^{\frac{1}{2}} \left( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-1)^2}{2}} dx \right) \xrightarrow{\text{cambio de variable}} e^{\frac{1}{2}} \\ \implies V(Y) &= E(Y^2) - E(Y)^2 = E(e^{2X}) - e = e^2 - e = e(e-1) \end{aligned}$$

Si  $X \sim N(\mu, \sigma^2)$  en su lugar y  $Z \sim N(0, 1)$

$$\begin{aligned} \implies E(e^X) &= E(e^{\mu+\sigma Z}) = e^\mu E(e^{\sigma Z}) = e^{\mu+\frac{\sigma^2}{2}} \\ \implies V(e^X) &= E(e^{2X}) - E(e^X)^2 = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} = e^{2\mu+\sigma^2} (e^{\sigma^2} - 1) \end{aligned}$$

11. Sea  $X$  una v.a. con función de distribución  $F_X$  no decreciente con inversa. Definimos  $Y := F_X(X)$ . Queremos ver que  $Y \sim \text{UNIF}([0, 1])$ .

$$\forall y \in (0, 1) : P(Y \leq y) = P(F_X(X) \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y$$

12. Sea  $F$  una función de distribución y  $U \sim \text{UNIF}([0, 1])$ . Definimos  $X := F^{-1}(U)$  y queremos ver que  $X \sim F$ .

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$$

Lo que nos dice este resultado es que cualquier variable aleatoria es una transformación de una variable aleatoria uniforme (Método de inversión).

13. Sea  $X \sim N(0, 1)$ .

$$\implies \text{a) } \Phi(1.25) \wedge \text{b) } 1 - \Phi(-0.4) = \Phi(0.4) \wedge \text{c) } 2\Phi(1.35) - 1$$

14. Sea  $X \sim N(\mu, \sigma^2)$  con  $\mu = 100 \wedge \sigma = 15$ . Si definimos  $Z \sim N(0, 1)$ :

$$P(X > 120) = P(\mu + \sigma Z > 120) = P\left(Z > \frac{120 - 100}{15}\right) = P\left(Z > \frac{4}{3}\right) = 1 - \Phi\left(\frac{4}{3}\right)$$

15. Sea  $X \sim N(0, 1)$ . Queremos  $a$  tal que  $P(|X| > a) = 0.95$ .

$$P(|X| > a) = 2P(X > a) = 2\Phi(a) - 1 = 0.95 \iff \Phi(a) = 0.975 \iff a = \Phi^{-1}(0.975)$$

Si  $X \sim N(\mu, \sigma^2)$  en su lugar:

$$\begin{aligned} P(|X| > a) &= P(-a < X < a) = P(-a < \mu + \sigma Z < a) \\ &= P\left(-\frac{a + \mu}{\sigma} < Z < \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right) - \Phi\left(-\frac{a + \mu}{\sigma}\right) = 0.95 \end{aligned}$$

19. Sea  $T \sim \text{lognormal}(\mu, \sigma^2)$  una variable aleatoria que mide la longitud de una conferencia (en minutos).

$$\begin{aligned} \begin{cases} 0.60 = P(T > 40) = P(e^{\mu + \sigma Z} > 40) = P\left(Z > \frac{\ln 40 - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{\ln 40 - \mu}{\sigma}\right) \\ 0.55 = P(T > 50) = P(e^{\mu + \sigma Z} > 50) = P\left(Z > \frac{\ln 50 - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{\ln 50 - \mu}{\sigma}\right) \end{cases} \\ \implies \begin{cases} \frac{\ln 40 - \mu}{\sigma} = \Phi^{-1}(0.40) \implies \mu = \ln 40 - \Phi^{-1}(0.40)\sigma \\ \frac{\ln 50 - \mu}{\sigma} = \Phi^{-1}(0.45) \implies \mu = \ln 50 - \Phi^{-1}(0.45)\sigma \end{cases} \\ \implies \sigma = \frac{\ln 50 - \ln 40}{\Phi^{-1}(0.45) - \Phi^{-1}(0.40)} \end{aligned}$$

## 5.4 Hoja 4

4. Sean  $X, Y$  dos variables aleatorias independientes idénticas ( $f := f_X \equiv f_Y$ ).

Definimos  $M := \max X, Y \wedge m := \min X, Y$ .

$$\begin{aligned} F_m(z) &= P(\min X, Y \leq z) = 1 - P(\min X, Y > z) = 1 - P(X > z \wedge Y > z) \\ &= 1 - P(X > z) \cdot P(Y > z) = 1 - P(X > z)^2 = 1 - (1 - F(z))^2 \\ \implies f_M(z) &= \frac{d}{dz} F_M(z) = 2(1 - F(z))f(z) \end{aligned}$$

Análogamente  $F_m(z) = F(z)^2 \implies f_m(z) = 2F(z)f(z)$

Si ahora tenemos  $X_1, \dots, X_n$  independientes con  $F_i, f_i$ ,  $M := \max\{X_1, \dots, X_n\}$ ,  $m := \min\{X_1, \dots, X_n\}$ :

$$\implies F_m(z) = 1 - \prod_{i=1}^n (1 - F_i(z)) \wedge F_M(z) = \prod_{i=1}^n F_i(z)$$

6. Sean  $X, Y \sim \text{EXP}(\lambda)$  independientes  $\implies P(\max\{X, Y\} \leq aX) = \begin{cases} 0 & \text{si } a \leq 1 \\ \frac{a}{1+a} & \text{si } a > 1 \end{cases}$

**Demostración.**

$$\implies \forall x, y > 0 : f_{X,Y}(x, y) = \lambda^2 e^{-\lambda x} e^{-\lambda y}$$

■

7. Sean  $X, Y$  con función de densidad conjunta  $f_{X,Y}$  y  $Z := \frac{Y}{X}$

$$\implies f_Z(z) = \int_{-\infty}^{\infty} |x| f_{X,Y}(x, z-x) dx$$

**Demostración.**

$$F_Z(z) = P(Z \leq z) = P\left(\frac{Y}{X} \leq z\right) = \begin{cases} P(Y \leq zX) & \text{si } X > 0 \\ P(Y \geq zX) & \text{si } X < 0 \end{cases}$$

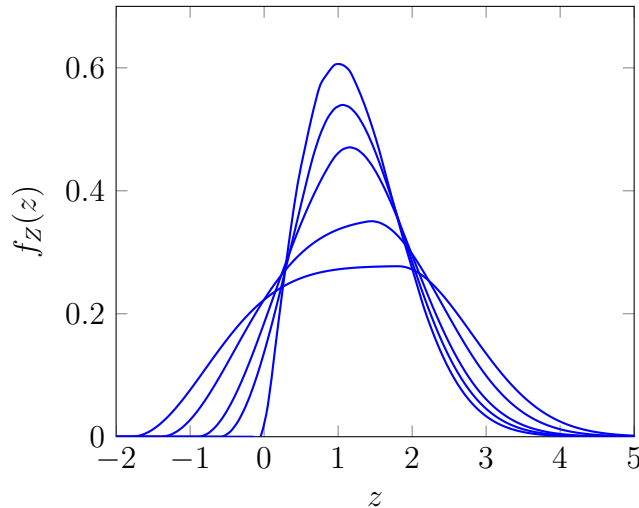
■

8. Sean  $X$  e  $Y$  dos v.a. indep. con  $f_X = xe^{-\frac{x^2}{2}} \cdot \mathbb{1}_{\{x>0\}}(x)$  e  $Y \sim \text{UNIF}([- \varepsilon, \varepsilon])$

$$\implies f_Y(y) = \frac{1}{2\varepsilon} \cdot \mathbb{1}_{\{-\varepsilon < y < \varepsilon\}}(y). \text{ Definimos } Z := X + Y.$$

**Nota:** En un escenario de la vida real,  $Y$  representa un error.

$$\begin{aligned} \forall z \geq -\varepsilon : f_Z(z) &= \int_{-\infty}^{\infty} f_X(u) f_Y(z-u) du = \int_{-\infty}^{\infty} ue^{-\frac{u^2}{2}} \cdot \mathbb{1}_{\{u>0\}}(u) \cdot \frac{1}{2\varepsilon} \cdot \mathbb{1}_{\{-\varepsilon < z-u < \varepsilon\}}(u) du \\ &= \begin{cases} \frac{1}{2\varepsilon} \int_{-\infty}^{\infty} ue^{-\frac{u^2}{2}} \cdot \mathbb{1}_{\{0 < u < z+\varepsilon\}}(u) du & z < \varepsilon \\ \frac{1}{2\varepsilon} \int_{-\infty}^{\infty} ue^{-\frac{u^2}{2}} \cdot \mathbb{1}_{\{z-\varepsilon < u < z+\varepsilon\}}(u) du & z \geq \varepsilon \end{cases} = \begin{cases} \frac{1}{2\varepsilon} \int_0^{z+\varepsilon} ue^{-\frac{u^2}{2}} du & z < \varepsilon \\ \frac{1}{2\varepsilon} \int_{z-\varepsilon}^{z+\varepsilon} ue^{-\frac{u^2}{2}} du & z \geq \varepsilon \end{cases} \\ \implies \forall z \geq -\varepsilon : f_Z(z) &= \begin{cases} \frac{1}{2\varepsilon} \left[ -e^{-\frac{u^2}{2}} \right]_0^{z+\varepsilon} & z < \varepsilon \\ \frac{1}{2\varepsilon} \left[ -e^{-\frac{u^2}{2}} \right]_{z-\varepsilon}^{z+\varepsilon} & z \geq \varepsilon \end{cases} = \begin{cases} \frac{1}{2\varepsilon} \left( 1 - e^{-\frac{(z+\varepsilon)^2}{2}} \right) & z < \varepsilon \\ \frac{1}{2\varepsilon} \left( e^{-\frac{(z-\varepsilon)^2}{2}} - e^{-\frac{(z+\varepsilon)^2}{2}} \right) & z \geq \varepsilon \end{cases} \end{aligned}$$



**13.** Sean  $(X_1, X_2) \sim \text{normal bidimensional con } \mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ . Hacemos el cambio de variable  $Z_1 := \frac{X_1 - \mu_1}{\sigma_1} \wedge Z_2 := \frac{X_2 - \mu_2}{\sigma_2}$ :

**14.** Sean  $X, Y \sim N(0, 1)$  independientes. Definimos  $W := 2X - Y$  y queremos ver que  $W \sim N(0, 5)$ .

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

Hacemos el cambio de variable  $U = X \wedge W = 2X - Y$  y sale

Otro camino, por convolución  $W = 2X + (-Y)$ :

$$\begin{aligned} f_W(w) &= \int_{-\infty}^{\infty} f_{X,Y}(x, w-2x) dx = \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2+(w-2x)^2}{2}} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{5x^2-2wx+w^2}{2}} dx = \frac{1}{2\pi} e^{-\frac{w^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{5x^2}{2}+wx} dx \\ \int_{-\infty}^{\infty} e^{-\frac{5x^2}{2}+wx} dx &= \int_{-\infty}^{\infty} e^{-\frac{5}{2}\left(x-\frac{w}{5}\right)^2+\frac{w^2}{10}} dx = e^{\frac{w^2}{10}} \int_{-\infty}^{\infty} e^{-\frac{5}{2}\left(x-\frac{w}{5}\right)^2} dx \\ &= e^{\frac{w^2}{10}} \sqrt{\frac{2\pi}{5}} = e^{\frac{w^2}{10}} \sqrt{\frac{2\pi}{5}} \implies f_W(w) = \frac{1}{\sqrt{10\pi}} e^{-\frac{w^2}{10}} \end{aligned}$$

**Experimento:** Sean  $X, Y \sim N(0, 1)$  independientes. Definimos  $Z := \sqrt{X^2 + Y^2}$ .

$$E(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sqrt{x^2 + y^2} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx dy$$

Mediante el cambio a polares:

$$\begin{aligned} \begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases} &\implies \begin{cases} dx = \cos \theta dr - r \sin \theta d\theta \\ dy = \sin \theta dr + r \cos \theta d\theta \end{cases} \implies dx dy = r dr d\theta \\ E(Z) &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} r^2 e^{-\frac{r^2}{2}} dr d\theta \end{aligned}$$

Desarrollando por partes:

$$\int_0^{\infty} r^2 e^{-\frac{r^2}{2}} dr = \left[ -r e^{-\frac{r^2}{2}} \right]_0^{\infty} + \int_0^{\infty} e^{-\frac{r^2}{2}} dr = \int_0^{\infty} e^{-\frac{r^2}{2}} dr = \sqrt{2\pi} \implies E(Z) = \frac{1}{\sqrt{2\pi}}$$

**Contenido adicional:**

Crear un modelo de normal bidimensional con  $\rho$  dado a partir de un modelo de normal bidimensional  $(X_1, X_2)$  con  $\rho = 0$ :

$$\begin{aligned} \begin{cases} Y_1 = X_1 \\ Y_2 = \rho X_1 + \sqrt{1-\rho^2} X_2 \end{cases} &\implies E(Y_1) = E(Y_2) = 0 \wedge V(Y_1) = V(Y_2) = 1 \\ \rho(Y_1, Y_2) &= \frac{E(Y_1 Y_2) - E(Y_1)E(Y_2)}{\sqrt{V(Y_1)V(Y_2)}} = \frac{E(X_1(\rho X_1 + \sqrt{1-\rho^2} X_2))}{\sqrt{1 \cdot 1}} = \rho \end{aligned}$$