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

01/02/2024

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 Ω. 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 Ω de un experimento aleatorio, $\mathcal{F} \subset \mathcal{P}(\Omega)$ es su espcio de sucesos

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

Observación 1.1. De la definición se deduce:

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

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

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

Proposición 1.1. *De la definición se deduce:*

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

Ejemplo 1.1. En un experimento aleatorio con espacio muestral finito, tomamos $\Omega = \{\omega_1, \dots, \omega_N\} \wedge \mathcal{F} = \mathcal{P} \to 2^N$. 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: $\forall j \in \mathbb{N}_N : p_j = \dfrac{1}{N} \implies \forall A \in \mathcal{F} : P(A) = \dfrac{|A|}{N} = \dfrac{\text{“Casos favorables”}}{\text{“Total de casos”}}$

05/02/2024

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) $(\Omega = \{\omega_1, \dots, \omega_N\} \implies |\Omega| = N) \wedge \left(\mathcal{F} = P(\Omega) \to |\mathcal{F}| = 2^N\right)$.
   Dato: $p_1, \dots, p_N \geq 0 \implies \displaystyle\sum_{j=1}^{N} p_j = 1$. Asignamos $\forall j \in \mathbb{N}_N : p_j := P(\{\omega_j\})$.
   Definimos $\forall A \subset \Omega : P(A) = \displaystyle\sum_{\omega \in A} P(\omega)$.

3. Lanzas n veces una moneda. Dato: $p \in (0,1)$.
$$\implies \Omega = \{111\cdots1, \dots, 000\cdots0\} \wedge |\Omega| = 2^N \wedge \text{escogemos } \mathcal{F} = P(\Omega)$$
$$\implies \forall \omega \in \Omega : P(\omega) = p^{\#\text{unos de } \omega}(1-p)^{\#\text{ceros de } \omega}$$
   Comprobamos:
$$\sum_{\omega \in \Omega} P(\omega) = \sum_{k=0}^{n} \left(\sum_{\substack{\omega \in \Omega \\ \#0\text{s de } \omega = k}} P(\omega) \right) = \sum_{k=0}^{n} p^k (1-p)^{n-k} \left(|\{\omega \in \Omega : \#1\text{s de } \omega = k\}| \right)$$
$$= \sum_{k=0}^{n} p^k (1-p)^{n-k} \binom{n}{k} = (p + 1 - p)^n = 1$$

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

1.2 Probabilidad condicionada y total, independencia y Bayes

Tienes $(\Omega, \mathcal{F}, P)$ y un suceso $A \in \mathcal{F} \to P(A)$. Llega “nueva información”: ha ocurrido el suceso $B \in F \longrightarrow$ ¿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

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

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* $Q_B \colon \mathcal{F} \longrightarrow [0,1] : Q_B(A) = P(A|B)$*, 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.

$$\implies Q_B\left(\bigcup_{j=1}^{\infty} A_j\right) = P\left(\bigcup_{j=1}^{\infty} A_j \,\middle|\, B\right) = \frac{P\left(\bigcup_{j=1}^{\infty} A_j \cap B\right)}{P(B)}$$

$$= \frac{1}{P(B)} P\left(\bigcup_{j=1}^{\infty} (A_j \cap B)\right) = \frac{1}{P(B)} \sum_{j=1}^{\infty} P(A_j \cap B) = \sum_{j=1}^{\infty} Q_B(A_j)$$

■

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

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

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

- A, B disjuntos $\implies$ no independientes.
- $A_1, \dots, A_N \in \mathcal{F}$ independientes $\iff \forall J \subset \mathbb{N}_N : P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j)$

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

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.

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* $\Omega : (\forall i, j : i \neq j : B_i \cap B_j = \phi) \wedge \left(\bigcup_{j=1}^{\infty} B_j = \Omega \right)$

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

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

■

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?

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

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

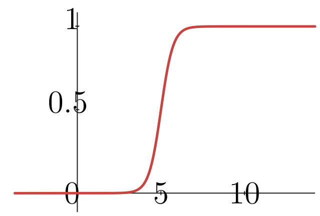
Este es el resultado que esperábamos, prestemos atención a otro caso más contraintituitivo.

$$\begin{aligned} P(U_1|6b4n) &= \frac{P(6b4n|U_1)P(U_1)}{P(6b4n|U_1)P(U_1)+P(6b4n|U_2)P(U_2)} \\ &= \frac{\binom{10}{6}0.8^6 0.2^4 \cdot {}^1\!/_2}{\binom{10}{6}0.8^6 0.2^4 \cdot {}^1\!/_2 + \binom{10}{6}0.8^4 0.2^6 \cdot {}^1\!/_2} \approx 90\% \end{aligned}$$

Si dibujamos la gráfica de la función

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

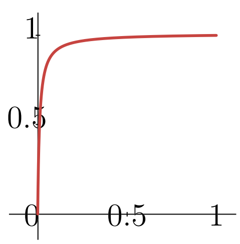
podemos ver que el cambio es muy brusco. Es decir, una pequeña diferencia en la evidencia puede cambiar mucho la probabilidad de que se haya usado una urna u otra.



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

$$\implies \{f(0.5) = 98.95\% \land f({}^1\!/_{100}) = 48.97\% \land f({}^1\!/_{1000}) = 8.68\%\}$$

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). $(\Omega, \mathcal{F}, P) \land A_1, \dots, A_n$ sucesos independientes tal que $\forall j \in \mathbb{N}_n : P(A_j) = \frac{1}{n}$. ¿Qué sabemos sobre $P\left(\bigcup_{j=1}^n A_j\right)$?

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

$n=2$: $P(A\cup B) = P(A)+P(B)-P(A\cap B) = {}^3\!/_4$

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

n general: $P\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} P(A_j) - \sum_{1\leq i<y\leq n} P(A_i \cap A_j) + \cdots$ (Inclusión exclusión)

$$\implies P\left(\bigcup_{j=1}^{n} A_j\right) = \binom{n}{1}\frac{1}{n} - \binom{n}{2}\frac{1}{n^2} + \binom{n}{3}\frac{1}{n^3} + \cdots = \sum_{j=1}^{n}\binom{n}{j}\frac{1}{n^j}(-1)^{j+1}$$

$$\implies P\left(\bigcup_{j=1}^{n} A_j\right) = 1 - \sum_{j=0}^{n}\binom{n}{j}\left(\frac{-1}{n}\right)^j = 1 - \left(1 - \frac{1}{n}\right)^n \xrightarrow{n\to\infty} 1 - \frac{1}{e}$$

1.2.2 Continuidad de la probabilidad: detalle técnico

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

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

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

$$P\left(\bigcup_{j=1}^{n} A_j\right) = P\left(A_1 \cup \bigcup_{j=1}^{n-1}(A_{j+1} \setminus A_j)\right) = P(A_1) + \sum_{j=1}^{n-1}(P(A_{j+1}) - P(A_j)) = P(A_n)$$

$$\implies P\left(\bigcup_{j=1}^{\infty} A_j\right) = P(A_1) + \sum_{j=1}^{\infty}(P(A_{j+1}) - P(A_j)) = \lim_{n\to\infty} P(A_n)$$

■

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

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

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

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

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

■

12/02/2024

2 Variables aleatorias discretas

Definición 2.1. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidades, $X\colon \Omega \longrightarrow \mathbb{R}$ 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 $(\Omega, \mathcal{F}, P)$, $p_X\colon \mathbb{R} \longrightarrow [0,1]$ es su función de masa

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

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

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

13/02/2024

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

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

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 $\{1, \cdots, N\}, N \geq 2$ ($X \sim \text{UNIF}\,(N)$).

   $$\iff S = \{1, \cdots, N\} \wedge \Pi_j = {}^1\!/\!_N, \dots, {}^1\!/\!_N$$

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

2. X sigue una distribución de **Bernoulli** con parámetro p ($X \sim \mathrm{BER}\,(p)$)
$$\iff \begin{cases} p_X(x) = 0 \iff x \neq 0, 1 \\ p_X(1) = p \wedge p_X(0) = 1 - p \end{cases} \iff \begin{cases} P(X = 1) = p \\ P(X = 0) = 1 - p \end{cases}$$
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 $n \geq 1 \wedge p \in (0, 1)$ ($X \sim \mathrm{BIN}\,(n, p)$)
$$\iff S = \{0, 1, \ldots, n\} \wedge \forall j \in \{0, 1, \ldots, n\} : P(X = j) = \binom{n}{j} p^j (1 - p)^{n-j}$$
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:
$$n! \sim n^n e^{-n} \sqrt{2\pi n} \implies \binom{n}{n/2} \sim \frac{n^n e^{-n} \sqrt{2\pi n}}{(n/2)^{(n/2)} e^{-(n/2)} \sqrt{2\pi (n/2)} (n/2)^{(n/2)} e^{-(n/2)} \sqrt{2\pi (n/2)}}$$
$$\implies \frac{n^n \sqrt{n}}{(n/2)^{(n/2)} \sqrt{2\pi (n/2)} (n/2)^{(n/2)} \sqrt{(n/2)}} = \frac{n^n \sqrt{n}}{(n/2)^n \sqrt{2\pi} (n/2)} = \frac{n^n \sqrt{2}}{(n/2)^n \sqrt{\pi n}} = 2^n \sqrt{\frac{2}{\pi n}}$$
$$\implies P\left(X = \frac{n}{2}\right) = \binom{n}{n/2} \frac{1}{2^n} \approx 2^n \sqrt{\frac{2}{\pi n}} \frac{1}{2^n} = \sqrt{\frac{2}{\pi n}}$$

4. X sigue una distribución **geométrica** de parámetro $p \in (0, 1)$ ($X \sim \mathrm{GEOM}\,(p)$).
$$\iff S = \{1, 2, \ldots\} \wedge \forall j \geq 1 : P(X = j) = p(1 - p)^{j-1}$$
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 $S = \{0, 1, 2, \ldots\}$. 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 $\lambda > 0$ ($X \sim \mathrm{POISSON}\,(\lambda)$)
$$\iff S = \{0, 1, \ldots\} \wedge \forall j \geq 0 : P(X = j) = e^{-\lambda} \frac{\lambda^j}{j!}$$
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* $X \sim \mathrm{BIN}\,(n,p)$ *una v.a.d.*

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

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

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

■

Ejemplo 2.2 (¿Hay más ejemplos?).

• Binomial negativa • Hipergeométrica

• Sea cualquier serie convergente $\displaystyle\sum_{n=1}^{\infty} a_n = s$

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

14/02/2024

2.1 Funciones transformadoras de variables aleatorias (discretas)

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

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

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 ($Y = x^2$).

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

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*)

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

Pero ojo, solo si la serie es absolutamente convergente.

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

20/02/2024

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 $(1+x)^n = \sum_{j=0}^{n} \binom{n}{j} x^j$

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

Si $x = \frac{p}{1-p} \implies \frac{p}{1-p} n \left(1 + \frac{p}{1-p}\right)^{n-1} = \sum_{j=0}^{n} \binom{n}{j} j \left(\frac{p}{1-p}\right)^j$

$$\implies \frac{p}{1-p} n \left(\frac{1}{1-p}\right)^{n-1} = \sum_{j=0}^{n} \binom{n}{j} j \left(\frac{p}{1-p}\right)^j$$

$$\implies np = (1-p)^n \sum_{j=0}^{n} \binom{n}{j} j \left(\frac{p}{1-p}\right)^j = \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j} = E(X)$$

$\blacksquare$

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

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

$\blacksquare$

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

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

$\blacksquare$

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

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

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 $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ 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.
Como $P(g(X) = y) = \displaystyle\sum_{x\in g^{-1}(y)} P(X = x)$

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

Como $\forall y \in g(X(\Omega)) : \exists \left|g^{-1}(y)\right|$ 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
$$\sum_{j\geq 1}(ax_j + b)p_j = a\sum_{j\geq 1} x_jp_j + b\sum_{j\geq 1} p_j = aE(X) + b$$
4. En general $E(g(X)) \neq g(E(X))$ **(Motivo de excomunión)**

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

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

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

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

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

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

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. $\boxed{V(aX + b)} = E[(aX + b - E(aX + b))^2] = E\left[(aX - aE(X))^2\right] = \boxed{a^2 V(X)}$
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. $X \sim \text{UNIF}(\{1, \ldots, N\}) \implies E(X) = \dfrac{N+1}{2} \wedge \boxed{V(X) = \dfrac{N^2 - 1}{12}}$

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

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

Demostración. ■

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

Demostración. ■

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

Demostración.

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

■

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

Observación 2.4. Algunos momentos tienen nombre propio:
1. $\mu_1 = 0$ 2. $\mu_2 = V(X)$ 3. μ_3 es la **asimetría** de X 4. μ_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 $Y_t(\omega) = t \cdot \mathbb{1}_{\{x \geq t\}}(\omega) = \begin{cases} t \text{ con probabilidad } P(x \geq t) \\ 0 \text{ con probabilidad } 1 - P(x \geq t) \end{cases}$

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

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

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

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

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

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

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

Definimos $\alpha := \sqrt{t} \implies \forall \alpha > 0 : P\left(|X - E(X)| \geq \alpha\right) \leq \dfrac{V(X)}{\alpha^2}$
y para la desigualdad equivalente definimos $\lambda := \dfrac{\alpha}{\sigma(X)}$

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

■

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 Ω*

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

(Siempre que la serie sea absolutamente convergente).

Demostración.

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

■

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.

$$\implies E(X) = E(X|C) \cdot p + E(X|\times) \cdot (1 - p)$$

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

$$\implies E(X) = p\sum_{k=1}^{\infty} kp^{k-1}(1-p) + (1-p)\sum_{k=1}^{\infty} k(1-p)^{k-1}p$$

$$\implies E(X) = \frac{1}{1-p}\cdot p + \frac{1}{p}\cdot(1-p) = \frac{1}{p(1-p)} - 2$$

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

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

$$\implies E(X) = \frac{p^2+1-2p+p^2}{p(1-p)} = \frac{2p(p-1)+1}{p(1-p)} = \frac{1}{p(1-p)} - 2$$

27/02/2024

2.3 Varias variables aleatorias

En $(\Omega, \mathcal{F}, P)$, sea $\mathcal{X} = X_1, X_2, \ldots, X_n$ 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 $\displaystyle\sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} p_{X,Y}(x,y) = 1$ 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* $g\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ *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 2.2 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 $g\colon \mathbb{R} \longrightarrow \mathbb{R}$

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

$$\implies E(g(X)) = \sum_{x \in X(\Omega)} g(x) p_X(x)$$

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

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

$$E(aX + bY) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} (ax + by) \cdot p_{X,Y}(x,y)$$

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

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*

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

Demostración. ($\Longrightarrow$) Trivial: $g(x) = p_X(x) \wedge h(y) = p_Y(y)$.
($\Longleftarrow$) Suponemos que $\exists\, g, h \colon \mathbb{R} \longrightarrow \mathbb{R} : p_{X,Y}(x,y) = g(x) \cdot h(y)$, veamos las funciones marginales.

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

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

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

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

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

■

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

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

$\Longrightarrow$ 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 $\Longrightarrow E(X \cdot Y) = E(X) \cdot E(Y)$.
Sin embargo, la implicación recíproca no es cierta. **(Motivo de excomunión)**

Definición 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

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

29/02/2024

Observación 2.7.

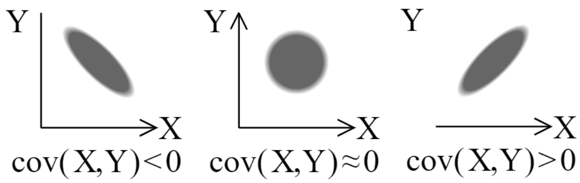
1. Cálculo de la covarianza

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

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

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

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

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

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

De forma más general:

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

$$\implies \boxed{V(aX+bY) = a^2V(X)+b^2V(Y)+2ab\operatorname{cov}(aX,bY)}$$

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

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

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

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

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.

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

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 $E\left(X^2\right) \geq 0$, 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.

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

■

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

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

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 $g, h\colon \mathbb{R} \longrightarrow \mathbb{R}$ dos funciones*

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

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

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

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

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

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

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

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

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

■

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

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

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

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

3. Independencia (la función de masa conjunta se factoriza)
$$p_{X_1,\dots,X_n}(x_1,\dots,x_n) = p_{X_1}(x_1)\cdot\dots\cdot p_{X_n}(x_n) \iff \text{ independencia completa}$$
Pero puede haber otras nociones de independencia (ej: 2 a 2).

4. **Matriz varianzas-covarianzas** y **matriz correlaciones** respectivamente
$$V = \begin{pmatrix} V(X_1) & \cdots & \operatorname{cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(X_n, X_1) & \cdots & V(X_n) \end{pmatrix} \wedge \Sigma = \begin{pmatrix} 1 & \cdots & \rho(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \rho(X_n, X_1) & \cdots & 1 \end{pmatrix}$$
Ambas son simétricas y definidas positivas.

Ejemplo 2.12. Queremos modelizar experimentos del tipo lanzar 18 veces un dado y sumar los resultados obtenidos.
$$S_n = X_1 + \dots + X_n \implies \begin{cases} E(S_n) = \sum_{i=1}^{n} E(X_i) \\ V(S_n) = \sum_{i=1}^{n} V(X_i) + 2\sum_{i<j} \operatorname{cov}(X_i, X_j) \end{cases}$$
Si suponemos las X_i independientes e idénticas $\implies \forall i \in \mathbb{N}_n : E(X_i) =: \mu \wedge V(X_i) =: \sigma^2$
$$\implies E(S_n) = n\mu \wedge V(S_n) = n^2\sigma^2$$
Si definimos $Z_n := \frac{1}{n}S_n = \frac{1}{n}(X_1 + \dots + X_n) \implies E(Z_n) = \mu \wedge V(Z_n) = \frac{\sigma^2}{n} \xrightarrow{n\to\infty} 0$
$$\implies Z_n \text{ no es aleatoria si } n \to \infty \text{ (ley de los grandes números)}$$

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

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

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

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

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

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

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

$$\implies V(X) = E(X^2) - E(X)^2 = G''_X(1) + G'_X(1)\,(1 - G'_X(1))$$

Ejemplo 2.15.

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

$$\implies G'_X(s) = p = E(X) \wedge G''_X(s) = 0 \implies V(X) = p(1-p)$$

(b) $X \sim \text{BIN}\,(n,p) \implies G_X(s) = (1-p+ps)^n$

$$\implies G'_X(s) = n(1-p+ps)^{n-1}p \implies G'_X(1) = np = E(X)$$

$$\implies G''_X(s) = n(n-1)(\cdots)^{n-2}p^2 \implies G''_X(1) = n(n-1)p^2$$

$$\implies V(X) = n(n-1)p^2 + np(1-np) = np(1-p)$$

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,

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

■

Otra manera:

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

■

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 $G_{X_1}(s), G_{X_2}(s), \ldots, G_{X_n}(s)$ respectivamente

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

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

06/03/2024

Teorema 2.10 (Unicidad)**.** *Sean X, Y dos v.a.d. con valores en $\{0, 1, 2, \ldots\}$ 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 $Z = X + Y$.

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

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 $I_1, I_2, \ldots, I_n$ v.a.d. independientes con $\forall k \in \mathbb{N}_n : I_k \sim \text{BER}\,(p)$ y definimos $Z = I_1 + I_2 + \cdots + I_n$.
$$\implies G_Z(s) = [(1-p) + ps]^n \implies Z \sim \text{BIN}\,(n, p)$$

2. Sean $X \sim \text{BIN}\,(n, p) \wedge Y \sim \text{POISSON}\,(\lambda)$ independientes y definimos $Z = X + Y$.
$$\implies G_Z(s) = ((1-p) + ps)^n \cdot e^{\lambda(s-1)}$$

12/03/2024

3 Variables aleatorias continuas

Hasta ahora en $(\Omega, \mathcal{F}, P)$, una variable aleatoria X discreta era una función $X\colon \Omega \longrightarrow \mathbb{R}$ 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 $X\colon \Omega \longrightarrow \mathbb{R}$ 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, $F_X\colon \mathbb{R} \longrightarrow [0,1]$ 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 $x_1, x_2, \cdots$ con probabilidades $p_1, p_2, \ldots$ y con función de masa $p_X \implies F_X(x) = \displaystyle\sum_{x_i \leq x} P(X = x_i)$. 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.*

$$\implies \begin{array}{ll} 1. & \lim_{x\to\infty} F_X(x) = 1 \wedge \lim_{x\to-\infty} F_X(x) = 0 \\ 2. & F_X \textit{ es no decreciente} \\ 3. & F_X \textit{ es continua por la derecha} \end{array}$$

Demostración.

1. Definimos $A_n := \{\omega \in \Omega : X(\omega) \leq n\}$ que es creciente según $n \to \infty$.
$$\implies \lim_{n\to\infty} F_X(n) = \lim_{n\to\infty} P(X \leq n) = \lim_{n\to\infty} P\left(\bigcup_{j=1}^{n} A_j\right) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(\Omega) = 1$$
2. $x < y \implies \{X \leq x\} \subseteq \{X \leq y\} \implies F_X(x) \leq F_X(y)$.
3. Definimos $A_h := \{\omega \in \Omega : X(\omega) \leq x + h\}$.
$$\implies \lim_{h\to 0^+} F_X(x+h) = \lim_{h\to 0^+} P(A_h) = P\left(\lim_{h\to 0^+} A_h\right) = P(A_0) = P(X \leq x) = F_X(x)$$

∎

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

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

Demostración. Suponemos que $\exists X, Y : \Omega \longrightarrow \mathbb{R}$ 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.)

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

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:

   $$\boxed{P(X = a)} = \lim_{h \to 0^+} P(a - h < X \leq a + h) = \lim_{h \to 0^+} F_X(a + h) - F_X(a - h)$$
   $$= \lim_{h \to 0^+} \left(\int_{-\infty}^{a} f_X(y)\,\mathrm{d}y - \int_{-\infty}^{a-h} f_X(y)\,\mathrm{d}y \right) = \lim_{h \to 0^+} \int_{a-h}^{a+h} f_X(y)\,\mathrm{d}y = \boxed{0}$$

   ■

2. Cálculo de probabilidades: $\forall a \leq b : P(a \leq X \leq b) = \displaystyle\int_a^b f_X(y)\,\mathrm{d}y$

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

Ejemplo 3.1.

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

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

3. $X \sim \mathrm{EXP}(\lambda) \iff f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0 \\ 0, & \text{en otro caso} \end{cases} = \lambda e^{-\lambda x} \cdot \mathbb{1}_{\{x \geq 0\}}(x)$

$$\implies \forall x > 0 : F_X(x) = \int_0^x \lambda e^{-\lambda y} \,\mathrm{d}y = \left[-e^{-\lambda y}\right]_0^x = 1 - e^{-\lambda x} \implies \lim_{x \to \infty} F_X(x) = 1$$

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

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

   Demostración.

$$I := \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x \implies I^2 = \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \,\mathrm{d}y\right)$$

$$\implies I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(x^2+y^2\right)} \,\mathrm{d}x \,\mathrm{d}y = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r \,\mathrm{d}r \,\mathrm{d}\theta$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \left[-e^{-\frac{r^2}{2}}\right]_0^{\infty} \mathrm{d}\theta = \frac{1}{2\pi} \int_0^{2\pi} 1 \,\mathrm{d}\theta = 1$$

$\blacksquare$

18/03/2024

3.1 Funciones / Transformaciones de v.a.c.

Sea X una v.a.c. con función de densidad f_X y $g \colon X(\Omega) \longrightarrow \mathbb{R}$ 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 $a \neq 0 \wedge b \in \mathbb{R}$.

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

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

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

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

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

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

***Demostración*.**

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

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

En el caso de g decreciente tenemos:

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

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

$\blacksquare$

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?

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

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

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

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

Como g es estrictamente creciente, aplicamos el teorema anterior:

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

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

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

Como g es estrictamente creciente, aplicamos el teorema anterior:

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

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 $\iff$ $\boxed{E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, \mathrm{d}x}$

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

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

Demostración.

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

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

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

■

Ejemplo 3.5.

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

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

$$\Longrightarrow \boxed{V(X)} = E\left(X^2\right) - E(X)^2 = \int_a^b x^2 \cdot \frac{1}{b-a}\,\mathrm{d}x - \left(\frac{a+b}{2}\right)^2 = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{b^3-a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2 = \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4}$$

$$= \frac{4b^2-4ab+4a^2-3a^2-6ab-3b^2}{12} = \frac{b^2-2ab+a^2}{12} = \boxed{\frac{(b-a)^2}{12}}$$

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

$$\Longrightarrow \boxed{E(X)} = \int_0^\infty x \cdot \lambda e^{-\lambda x}\,\mathrm{d}x = \left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}\,\mathrm{d}x = 0 + \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^\infty = \boxed{\frac{1}{\lambda}}$$

$$\Longrightarrow \boxed{V(X)} = E\left(X^2\right) - E(X)^2 = \int_0^\infty x^2 \cdot \lambda e^{-\lambda x}\,\mathrm{d}x - \left(\frac{1}{\lambda}\right)^2 = \boxed{\frac{1}{\lambda^2}} \text{ porque}$$

$$\int_0^\infty x^2 \cdot \lambda e^{-\lambda x}\,\mathrm{d}x = \left[-x^2e^{-\lambda x}\right]_0^\infty + 2\int_0^\infty xe^{-\lambda x}\,\mathrm{d}x = 0 + 2\cdot\frac{1}{\lambda^2} = \frac{2}{\lambda^2}$$

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

$\boxed{E(X)} = \displaystyle\int_{-\infty}^\infty x\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,\mathrm{d}x = \boxed{0}$ (integrando impar en región centrada en el origen).

$$\boxed{V(X)} = E(X^2) - E(X)^2 = \int_{-\infty}^\infty x^2 \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty x^2 \cdot e^{-\frac{x^2}{2}}\,\mathrm{d}x$$

$$= \frac{1}{\sqrt{2\pi}}\left(\left[-xe^{-\frac{x^2}{2}}\right]_{-\infty}^\infty + \int_{-\infty}^\infty e^{-\frac{x^2}{2}}\,\mathrm{d}x\right) = \frac{1}{\sqrt{2\pi}}\left(0+\sqrt{2\pi}\right) = \boxed{1}$$

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

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

Con el cambio de variable $t = \frac{x-\mu}{\sigma} \implies x = \sigma t + \mu \wedge \mathrm{d}x = \sigma\,\mathrm{d}t$.

$$\Longrightarrow \boxed{E(X)} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^\infty (\sigma t+\mu)e^{-\frac{t^2}{2}}\sigma\,\mathrm{d}t = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty (\sigma t+\mu)e^{-\frac{t^2}{2}}\,\mathrm{d}t$$

$$= \frac{1}{\sqrt{2\pi}}\left(\sigma\int_{-\infty}^\infty te^{-\frac{t^2}{2}}\,\mathrm{d}t + \mu\int_{-\infty}^\infty e^{-\frac{t^2}{2}}\,\mathrm{d}t\right) = \frac{1}{\sqrt{2\pi}}\left(0+\mu\sqrt{2\pi}\right) = \boxed{\mu}$$

$$\Longrightarrow V(X) = E\left[(X-E(X))^2\right] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^\infty (x-\mu)^2e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x$$

Con el cambio de variable $t = \frac{x-\mu}{\sqrt{2}\sigma} \implies x = \sqrt{2}\sigma t + \mu \wedge \mathrm{d}x = \sqrt{2}\sigma\,\mathrm{d}t$.

$$V(X) = \frac{1}{\sqrt{\cancel{2}\pi\cancel{\sigma}}}\int_{-\infty}^\infty \left(\sqrt{2}\sigma t\right)^2 e^{-t^2}\cancel{\sqrt{2}\sigma}\,\mathrm{d}t = \frac{2\sigma^2}{\sqrt{\pi}}\int_{-\infty}^\infty t^2e^{-t^2}\,\mathrm{d}t = \frac{4\sigma^2}{\sqrt{\pi}}\int_0^\infty t^2e^{-t^2}\,\mathrm{d}t$$

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

$$\implies \boxed{V(X)} = \frac{4\sigma^2}{\sqrt{\pi}}\int_0^\infty ue^{-u}\frac{1}{2\sqrt{u}}\,\mathrm{d}u = \frac{2\sigma^2}{\sqrt{\pi}}\int_0^\infty u^{\frac{3}{2}-1}e^{-u}\,\mathrm{d}u = \frac{2\sigma^2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right)$$
$$= \frac{\cancel{2}\sigma^2}{\sqrt{\pi}}\cdot\frac{1}{\cancel{2}}\cdot\Gamma\left(\frac{1}{2}\right) = \frac{\sigma^2}{\cancel{\sqrt{\pi}}}\cdot\cancel{\sqrt{\pi}} = \boxed{\sigma^2}$$

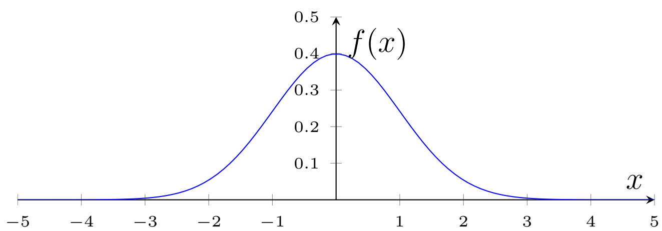
21/03/2024

3.2.1 Calculando con la normal

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

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

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



¡Basta con $\mathbf{N}(0,1)$!

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

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

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

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

$$E(Y) = E(\mu + \sigma X) = \mu + \sigma E(X) = \mu \ \wedge \ V(Y) = V(\mu + \sigma X) = \sigma^2 V(X) = \sigma^2$$
$$E\left(Y^7\right) = E\left((\mu + \sigma X)^7\right) = \sum_{k=0}^{7}\binom{7}{k}\mu^7\sigma^{7-k}E(X^{7-k})$$

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

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

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

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

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

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

Observación 3.2. $\operatorname{erf}(x) = \dfrac{2}{\sqrt{2}} \displaystyle\int_0^x e^{-y^2}\, \mathrm{d}y \implies \operatorname{erf}(x) = 2\Phi(\sqrt{2}x) - 1$

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

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. • $\displaystyle\lim_{x,y\to\infty} F_{X,Y}(x,y) = 1$ • $\forall \hat{x} \leq x, \hat{y} \leq y : F_{X,Y}(\hat{x},\hat{y}) \leq F_{X,Y}(x,y)$

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

¿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

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

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

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

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

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

$$\implies f_{X,Y} = \begin{cases} 0, & \text{si } (x,y) \notin \mathcal{R} \\ \frac{1}{ab}, & \text{si } (x,y) \in \mathcal{R} \end{cases}$$

$$\implies \forall A \subset \mathcal{R} : P((x,y) \in A) = \iint_A \frac{1}{ab}\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{ab}\,\text{Área}(A)$$

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 ρ)

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

Notación matricial

$$\vec{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \mathbb{X} = \begin{pmatrix} X \\ Y \end{pmatrix} \quad \Sigma := \begin{pmatrix} \sigma^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \implies \Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

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

3.3.1 Normal multidimensional

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

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

3.3.2 Marginales e independencia

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

Teorema 3.4. *Sean X, Y dos variables aleatorias con función de densidad conjunta $f_{X,Y}$*

$$X \text{ e } Y \text{ independientes} \iff \exists\, h, g : \forall (x,y) \in \mathbb{R} : f_{X,Y}(x,y) = h(x) \cdot g(y)$$

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

$$\implies f(x,y) = \left(e^{-x} \cdot \mathbb{1}_{\{x\geq 0\}}(x)\right) \cdot \left(e^{-y} \cdot \mathbb{1}_{\{y \geq 0\}}(y)\right) \implies \text{son independientes}$$

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 $\forall A, B \subset \mathbb{R}^2 : P((X,Y) \in A | (X,Y) \in B) = \dfrac{P((X,Y) \in A \,\wedge\, (X,Y) \in B)}{P((X,Y) \in B)}$ 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 $\displaystyle\int_{-\infty}^{\infty} f_{Y|X=x}(y)\,\mathrm{d}y = 1$ y $\displaystyle\int_{-\infty}^{\infty} f_{X|Y=y}(x)\,\mathrm{d}x = 1$:

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

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

$$\begin{aligned} \implies \forall x > 0 : f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y = \int_x^{\infty} 2e^{-x}e^{-y}\,\mathrm{d}y = 2e^{-x}\int_x^{\infty} e^{-y}\,\mathrm{d}y \\ &= 2e^{-x}e^{-x} = 2e^{-2x} \end{aligned}$$

$$\forall y > 0 : \boxed{f_Y(y)} = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x = \int_0^y 2e^{-x}e^{-y}\,\mathrm{d}x = 2e^{-y}\int_0^y e^{-x}\,\mathrm{d}x = \boxed{2e^{-y}(1-e^{-y})}$$

$$f_{Y|X=3}(y) = \frac{f_{X,Y}(3,y)}{f_X(3)} = \frac{2e^{-3}e^{-y}}{2e^{-6}} = e^{3-y}\cdot \mathbb{1}_{\{y>3\}}(y)$$

$$\implies f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{2e^{-x}e^{-y}}{2e^{-y}(1-e^{-y})} = \frac{e^{-x}}{1-e^{-y}}\cdot \mathbb{1}_{\{0<x<y\}}(x)$$

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

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 función de densidad conjunta $f_{X,Y} \wedge T\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$
$(x,y) \longmapsto (u(x,y), v(x,y))$

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 también hay transformaciones no lineales como $U = X + Y \wedge V = \frac{X}{X+Y}$.

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

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

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

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

09/04/2024

Ejemplo 3.10. Sean X, Y v.a. con $\forall x, y > 0 : f_{X,Y}(x, y) := \dfrac{1}{4}e^{-\frac{1}{2}(x+y)}$.
Sea $T(x, y) = \left(\frac{x-2}{2}, y\right)$. Definimos $(U, V) := T(X, Y)$.

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

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

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

Ejemplo 3.11. Sean X, Y v.a. con $\forall x, y > 0 : f_{X,Y}(x, y) := e^{-x-y}$.
Sea $T(x, y) = \left(x + y, \dfrac{x}{x + y}\right)$. Definimos $(U, V) := T(X, Y)$.

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

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

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

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

10/04/2024

Sean X, Y dos variables con función de densidad conjunta $f_{X,Y}$, consideramos una función $g\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ “razonable”$^{(*)}$ 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 $g\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$*

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

Observación 3.5.

1. $\forall \alpha, \beta \in \mathbb{R} : E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$
2. $E(X \cdot Y) = \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y$

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

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

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

4. Esperanza condicionada y total
$$E(Y|X=x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X=x}(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} y \cdot \frac{f_{X,Y}(x,y)}{f_X(x)}\,\mathrm{d}y$$
$$\implies E(Y) = \int_{-\infty}^{\infty} E(Y|X=x) \cdot f_X(x)\,\mathrm{d}x$$

¿Y qué hay de $\boldsymbol{f_Z(z)}$? Definimos $A_g(z) := \{(x,y) \in \mathbb{R}^2 : g(x,y) \leq z\}$
$$\forall z \in \mathbb{R} : F_Z(z) = P(Z \leq z) = P(g(X,Y) \leq z) = P((X,Y) \in A_g(z)) = \iint_{A_g(z)} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
$$\implies \forall z \in \mathbb{R} : f_Z(z) = \begin{cases} \frac{\partial}{\partial z} F_Z(z) & \text{si la derivada existe} \\ 0 & \text{en otro caso} \end{cases}$$

Ejemplo 3.12 (El caso de la suma). Sean X, Y dos v.a. con $f_{X,Y}$ y $Z := X+Y$. Entonces $A_+(z) := \{(x,y) \in \mathbb{R}^2 : x+y \leq z\}$.
$$\implies \forall z \in \mathbb{R} : F_Z(z) = \iint_{A_+(z)} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x$$
Mediante el cambio de variables $u = x \wedge v = x+y$ se tiene
$$T(x,y) = (x, x+y) \implies T^{-1}(u,v) = (u, v-u) \implies \mathrm{J}(u,v) = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$
$$\implies F_Z(z) = \int_{-\infty}^{z}\int_{-\infty}^{\infty} f_{X,Y}(u, v-u)\,\mathrm{d}u\,\mathrm{d}v \implies \boxed{f_Z(z)} = \frac{\mathrm{d}}{\mathrm{d}z} F_Z(z) = \boxed{\int_{-\infty}^{\infty} f_{X,Y}(u, z-u)\,\mathrm{d}u}$$
Caso particular: Si X, Y independientes $\implies f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$
$$\implies \forall z \in \mathbb{R} : \boxed{f_{X+Y}(z)} = \int_{-\infty}^{\infty} f_X(u) \cdot f_Y(z-u)\,\mathrm{d}u =: \boxed{(f_X * f_Y)(z)}$$

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.

$$\implies \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x)\,\mathrm{d}x = \left(f_X * f_Y\right)(z)$$

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

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

$(*)$ 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 $w = \sqrt{2}u - \frac{z}{\sqrt{2}} \implies \mathrm{d}w = \sqrt{2}\,\mathrm{d}u$, se tiene

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

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

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

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

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

$$E(X\cdot Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\cdot y \cdot \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(x^2-2\rho xy+y^2\right)}\,\mathrm{d}x\,\mathrm{d}y$$

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

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

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

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

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

$$\implies E(Y|X=x) = \int_{-\infty}^{\infty} y \cdot \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}(y-\rho x)^2} \,\mathrm{d}y = \rho x$$

$$\implies \boxed{E(X \cdot Y)} = \int_{-\infty}^{\infty} x \cdot \rho x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x = \rho \underbrace{\int_{-\infty}^{\infty} x^2 f_X(x) \,\mathrm{d}x}_{E(X^2)} = \boxed{\rho}$$

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 $Z_1, Z_2 \sim \mathrm{N}\left(0,1\right)$.

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

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

Tenemos $\mathbb{X}^T = (X_1, \ldots, X_n)$ normal n-dimensional con $\vec{\mu} = (\mu_1, \ldots, \mu_n)$ 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)**.** $\mathbb{X} \sim \mathrm{N}\left(\vec{\mu}, \mathbb{V}\right) \iff \mathbb{X} = \vec{\mu} + U \cdot \mathbb{Z}$ *con* $\mathbb{Z} \sim \mathrm{N}\left(\vec{0}, I\right)$

22/04/2024

4 Convergencia de variables aleatorias

Sea $(X_n)_{n\in\mathbb{N}}$ una sucesión de variables aleatorias en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, 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 $n \to \infty$ de estas series.

1. Sabemos que significa $\lim\limits_{n\to\infty} E(S_n)$ y $\lim\limits_{n\to\infty} V(S_n)$ pero, ¿qué significa $\lim\limits_{n\to\infty} S_n = S$? Requerimos de técnicas más avanzadas que se denominan **modos de convergencia**.
2. Asumiremos que $\{X_n\}$ son **independientes e idénticamente distribuidas (i.i.d.)**.
   - Todas las X_i tienen la misma distribución. Por tanto, es habitual definir una $\boldsymbol{X}$ **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 $\dfrac{Z_n - E(X)}{\sigma(Z_n)/\sqrt{n}} \approx \mathrm{N}(0,1)$.

4.1 Medias y varianzas de las sumas y las medias

Las medias de S_n y Z_n resultan sencillas de calcular.

$$E(S_n) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) \overset{(\star)}{=} nE(X)$$

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

($\star$) Si X_i tienen la misma media.

Nótese que, en la mayoría de casos, $E(S_n) \xrightarrow{n\to\infty} \infty$ mientras que $E(Z_n)$ tenderá a al promedio de las medias de las X_i.

Las varianzas de S_n y Z_n son más intricadas de calcular y debemos tener en cuenta, no solo el caso de que las X_i tengan la misma varianza, sino también si son incorreladas.

$$
\begin{aligned}
V(S_n) &= V\left(\sum_{i=1}^{n} X_i\right) = V(X_1) + V\left(\sum_{i=2}^{n} X_i\right) + 2\operatorname{cov}\left(X_1, \sum_{i=2}^{n} X_i\right) \\
&= V(X_1) + 2\sum_{i=2}^{n} \operatorname{cov}(X_1, X_i) + V\left(\sum_{i=2}^{n} X_i\right) \\
&= V(X_1) + 2\sum_{i=2}^{n} \operatorname{cov}(X_1, X_i) + V(X_2) + 2\sum_{i=3}^{n} \operatorname{cov}(X_2, X_i) + V\left(\sum_{i=3}^{n} X_i\right) \\
&= \sum_{i=1}^{n} V(X_i) + 2\sum_{1\le i<j\le n} \operatorname{cov}(X_i, X_j) \overset{(*)_1}{=} \sum_{i=1}^{n} V(X_i) \overset{(*)_2}{=} n(\sigma(X))^2 \\
V(Z_n) &= V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right) \overset{(*)_1}{=} \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) \overset{(*)_2}{=} \frac{(\sigma(X))^2}{n} \xrightarrow{n\to\infty} 0
\end{aligned}
$$

$(*)_1$ Si X_i incorreladas, $(*)_2$ Si (además) X_i tienen la misma varianza.

De forma similar a como ocurre con las medias, (en la mayoría de casos) la varianza de S_n tiende a infinito mientras que la de Z_n tiende a 0 cuando $n \to \infty$.

4.2 Convergencia cuadrática

Definición 4.1 (Convergencia cuadrática). Sea $(X_n)_{n\in\mathbb{N}}$ una sucesión de v.a. y X v.a. en el mismo espacio de probabilidad $(\Omega, \mathcal{F}, P)$, $\boxed{X_n \xrightarrow{\text{cuad.}} X \iff E\left(|X_n - X|^2\right) \xrightarrow{n\to\infty} 0}$

Teorema 4.1. *Sean $(X_n)_{n\in\mathbb{N}}$ v.a. i.i.d. con X v.a. de referencia $\implies Z_n \xrightarrow{cuad.} E(X)$*

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

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

■

Basta con que sean incorreladas y tengan la misma media y varianza porque es lo único que se ha usado en la demostración. Podemos incluso ofrecer un teorema más general que no requiera que tengan la misma media y varianza.

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

Demostración. (Buen ejercicio de examen)

$$
\begin{aligned}
E\left(|Z_n-\mu|^2\right) &= E\left(|Z_n-\mu_n+\mu_n-\mu|^2\right) \\
&= E\left(|Z_n-\mu_n|^2+|\mu_n-\mu|^2+2(Z_n-\mu_n)(\mu_n-\mu)\right) \\
&= E\left((Z_n-\mu_n)^2\right)+(\mu_n-\mu)^2+2(\mu_n-\mu)\cancelto{0}{E(Z_n-\mu_n)} \\
&= E\left((Z_n-\mu_n)^2\right)+(\mu_n-\mu)^2 \xrightarrow{n\to\infty} 0
\end{aligned}
$$

■

Es decir, si el promedio de las medias tiende a un número finito y la varianza del promedio tiende a 0, entonces el promedio tiende al límite del promedio de las medias.

23/04/2024

4.3 Convergencia en probabilidad (ley débil)

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.

Definición 4.2 (Convergencia en probabilidad). Sea $(X_n)_{n\in\mathbb{N}}$ una sucesión de v.a. y X v.a. en el mismo espacio de probabilidad $(\Omega,\mathcal{F},P)$,

$$X_n \xrightarrow{P} X \iff \forall\varepsilon>0: \lim_{n\to\infty} P\left(|X_n-X|>\varepsilon\right)=0$$

Teorema 4.3 (Ley débil de los grandes números)**.** *Sean $(X_n)_{n\in\mathbb{N}}$ v.a. i.i.d. en $(\Omega,\mathcal{F},P)$ con X v.a. de referencia tal que $\mu := E(X)<\infty \wedge \sigma^2 := V(X)<\infty \implies Z_n := \frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{P} \mu$*

Demostración. Dado $\varepsilon>0$, por la desigualdad de Chebyshev

$$P\left(|Z_n-\mu|>\varepsilon\right) \le \frac{V(Z_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n\to\infty} 0$$

■

Observación 4.1.

1. Hay una ley fuerte de los grandes números.
2. Hipótesis: No hace falta que sean independientes, basta con que sean incorreladas. Tampoco hace falta que sean idénticas, basta con que tengan la misma media y varianza.
3. Se podría incluso hacer una variante del teorema para variables aleatorias que no tengan la misma media y varianza.
4. Existe la posibilidad de adaptar el teorema para que no haga falta que sean incorreladas, solo que tengan una correlación "pequeña".

Teorema 4.4. *Sean* $(X_n)_{n\in\mathbb{N}}$ *v.a. incorreladas en* $(\Omega, \mathcal{F}, P)$ *con* $E(X_i) = \mu_i$ *y* $V(X_i) = \sigma_i^2$. *Suponemos* $V(Z_n) = \dfrac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 \xrightarrow{n\to\infty} 0 \implies Z_n \xrightarrow{P} \dfrac{1}{n}\sum_{i=1}^{n}\mu_i$

Demostración. Por la desigualdad de Chebyshev

$$P\left(\left|Z_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| > \varepsilon\right) \le \frac{V(Z_n)}{\varepsilon^2} = \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{\infty}\sigma_i^2 \xrightarrow{n\to\infty} 0$$

■

Es decir, si la varianza del promedio tiende a 0, entonces el promedio tiende al promedio de las medias.

Usos en estadística: Sean $X_1, \ldots, X_n$ v.a. i.i.d. con $X \sim \mathrm{BER}(p)$ de referencia. Queremos estimar p (desconocido), ¿cuánto de grande tiene que ser n para estar razonablemente seguros que el estimador $\bar{X}_n := \dfrac{1}{n}\sum_{i=1}^{n} X_i$ está cerca de p?

$$P\left(\left|\bar{X}_n - p\right| > \varepsilon\right) \le \frac{p(1-p)}{n\varepsilon^2} \le \frac{1}{4n\varepsilon^2} \le \delta \implies n \ge \frac{1}{4\varepsilon^2\delta}$$

Es decir, si queremos estar δ seguros de que el estimador está a una distancia menor que ε de p, necesitamos que n sea mayor que $\dfrac{1}{4\varepsilon^2\delta}$.

4.4 Cálculo de la distribución de la suma y el promedio

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

1. $\boldsymbol{X \sim \mathbf{BER}(p)} \implies S_n \sim \mathrm{BIN}(n,p)$ y $Z_n \in \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$ con las mismas probabilidades que una $\mathrm{BIN}(n,p)$.

2. $\boldsymbol{X \sim \mathbf{POISSON}(\lambda)} \implies S_n \sim \mathrm{POISSON}(n\lambda)$. Ya vimos que la suma de dos poisson es poisson. Ahora, por inducción asumimos que se cumple para $n-1$.

$$\begin{aligned}P(S_n = k) &= P\left(\sum_{i=1}^{n} X_i = k\right) = \sum_{i=0}^{k} P\left(\sum_{i=1}^{n-1} X_i = k - i \wedge X_n = i\right)\\ &= \sum_{i=0}^{k} P\left(\sum_{i=1}^{n-1} X_i = k - i\right) P(X_n = i) = \sum_{i=0}^{k} \frac{(n-1)^{k-i}\lambda^{k-i}}{(k-i)!} e^{-(n-1)\lambda}\frac{\lambda^i}{i!}e^{-\lambda}\\ &= e^{-n\lambda}\lambda^k \sum_{i=0}^{k}\frac{(n-1)^{k-i}}{(k-i)!i!} = \frac{\lambda^k}{k!}e^{-n\lambda}\sum_{i=0}^{k}\binom{k}{i}(n-1)^{k-i} = \frac{(n\lambda)^k}{k!}e^{-n\lambda}\end{aligned}$$

3. $\boldsymbol{X \sim \mathbf{GEOM}(p)} \implies S_n \sim$ binomialnegativa(n, p).

4. $\boldsymbol{X \sim \mathbf{N}(\mu, \sigma^2)} \implies S_n \sim \mathrm{N}(n\mu, n\sigma^2)$.

5. $\boldsymbol{X \sim \mathbf{EXP}\,(\lambda)} \implies S_n \sim \text{Gamma}(n, \lambda)$.

Técnicas generales (Funciones generatrices)

24/04/2024

4.5 Convergencia en distribución

Definición 4.3 (Convergencia en distribución). Sea $(X_n)_{n\in\mathbb{N}}$ una sucesión de v.a. y X v.a. en el mismo espacio de probabilidad $(\Omega, \mathcal{F}, P)$,

$$\left(X_n \xrightarrow{d} X\right) \iff \forall t \in \mathbb{R}^{(*)} : P\left(X_n \le t\right) \xrightarrow{n\to\infty} F_X(t) := P(X \le t)$$

$(*)$ F_X debe ser continua en t.

4.5.1 Teorema del límite central / central del límite

Vamos a tipificar $S_n := \displaystyle\sum_{i=1}^{n} X_i$ y $Z_n := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ con $(X_n)_{n\in\mathbb{N}}$ i.i.d. con X de referencia.

$$W_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} \quad \wedge \quad V_n := \frac{Z_n - \mu}{\sigma/\sqrt{n}}$$

Teorema 4.5 (del límite central)**.** *Sean $(X_n)_{n\in\mathbb{N}}$ v.a. i.i.d. en $(\Omega, \mathcal{F}, P)$ con X de referencia.*

$$\implies \forall t \in \mathbb{R} : P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le t\right) \xrightarrow{n\to\infty} \Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{x^2}{2}}dx$$

Es decir, este teorema nos dice que $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} U$ *donde* $U \sim \mathrm{N}\,(0, 1)$

Ejemplo 4.1. Sea $X \sim \mathrm{BER}\,(p)$ la v.a. de referencia de $(X_n)_{n\in\mathbb{N}}$ i.i.d., entonces $S_n \sim \mathrm{BIN}\,(n, p)$ y $Z_n \in \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$ con las mismas probabilidades que una $\mathrm{BIN}\,(n, p)$.

Si $p = 1/2$ y $n = 1000$ (lanzamos una moneda regular 1000 veces) y nos piden:

$$P(480 \le S_{1000} \le 530) = \sum_{j=480}^{530} \binom{1000}{j} \left(\frac{1}{2}\right)^{1000} \approx 87.578\%$$

Pero podemos aproximar la respuesta usando la normal por el teorema del límite central:

$$\begin{aligned} P(480 \le S_{1000} \le 530) &= P\left(\frac{480 - 500}{\sqrt{250}} \le \frac{S_{1000} - 500}{\sqrt{250}} \le \frac{530 - 500}{\sqrt{250}}\right) \\ &\approx \Phi\left(\frac{30}{\sqrt{250}}\right) - \Phi\left(\frac{-20}{\sqrt{250}}\right) \approx 86.816\% \end{aligned}$$

Ejemplo 4.2. Lanzamos un dado 10000 veces y nos piden $P(3400 \le S \le 3500)$.

Tenemos $X \sim \mathrm{UNIF}\,(1, 6) \implies E(X) = \dfrac{7}{2} \wedge V(X) = \dfrac{35}{12}$ y, por tanto, $S = \sum_{i=1}^{1000} X_i \implies E(S) = 3500 \wedge V(S) = 1000\frac{35}{12}$.

$$P(3400 \le S \le 3500) = P\left(\frac{-100}{\sqrt{1000 \cdot {}^{35}/_{12}}} \le \frac{S - 3500}{\sqrt{1000 \cdot {}^{35}/_{12}}} \le 0\right)$$

$$\approx \Phi(0) - \Phi\left(\frac{-100}{\sqrt{1000 \cdot {}^{35}/_{12}}}\right) \approx 46.79\%$$

Ejemplo 4.3 (Intervalo de confianza). Sea $X \sim \mathrm{BER}\,(p)$ de referencia con p desconocido de $(X_n)_{n\in\mathbb{N}}$ i.i.d. y $\overline{X}_n$ el promedio de las X_i.

Fijamos $\alpha \in \mathbb{R}$ "pequeño" y definimos $z_{\alpha/2} := \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$.

$$\implies P\left(-z_{\alpha/2} \le \frac{\overline{X}_n - p}{\sqrt{p(1-p)/n}} \le z_{\alpha/2}\right) \approx 1 - \alpha$$

$$=P\left(-\frac{-z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}} \le \overline{X}_n - p \le \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right)$$

$$=P\left(\overline{X}_n - \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}} \le p \le \overline{X}_n + \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right)$$

Tenemos confianza $1 - \alpha$ de que p está en el intervalo $\left(\overline{x}_n - \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}, \overline{x}_n + \frac{z_{\alpha/2}\sqrt{p(1-p)}}{\sqrt{n}}\right)$ donde $\overline{x}_n$ es el valor observado de $\overline{X}_n$. Sin embargo, este intervalo depende de p, pero podemos acotarlo por $\left(\overline{x}_n - \frac{z_{\alpha/2}}{2\sqrt{n}}, \overline{x}_n + \frac{z_{\alpha/2}}{2\sqrt{n}}\right)$.

Por tanto, el error sería $\frac{z_{\alpha/2}}{2\sqrt{n}}$. Si queremos que sea menor que ε, necesitamos que $n \ge \frac{\left(z_{\alpha/2}\right)^2}{4\varepsilon^2}$.

25/04/2024

Ejercicio 4.5.1 (4 b del examen). Sea $(X, Y) \sim$ 2-dim $\mathrm{N}\,(0, 1)$ con correlación ρ. Calcula la varianza de $X \cdot Y$. Tenemos que
$V(X\cdot Y) = E\left(X^2 \cdot Y^2\right) - E\left(X \cdot Y\right)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2y^2\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)}\,\mathrm{d}x\,\mathrm{d}y - \rho^2$

porque $E(X \cdot Y) = \rho$. Completando cuadrados, obtenemos

$$V(X \cdot Y) = \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} x^2e^{-\frac{x^2}{2}}\left(\int_{-\infty}^{\infty} y^2e^{-\frac{1}{2}\left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right)^2}\mathrm{d}y\right)\mathrm{d}x - \rho^2$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2e^{-\frac{x^2}{2}}\left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}y^2e^{-\frac{1}{2}\left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right)^2}\mathrm{d}y\right)\mathrm{d}x - \rho^2$$

El término entre paréntesis es $E\left(W^2\right)$ donde $W \sim \mathrm{N}\left(\rho x, 1-\rho^2\right)$ y como $V(W) = E(W^2) - E(W)^2 = 1-\rho^2$, tenemos $E(W^2) = V(W) + E(W)^2 = 1-\rho^2+(\rho x)^2$.

$$
\begin{aligned}
\Longrightarrow \boxed{V(X\cdot Y)} &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}}\left(1-\rho^2+\rho^2x^2\right)\mathrm{d}x - \rho^2 \\
&= (1-\rho^2)\int_{-\infty}^{\infty} x^2\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,\mathrm{d}x + \rho^2\int_{-\infty}^{\infty} x^4\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,\mathrm{d}x - \rho^2 \\
&= (1-\rho^2) + \rho^2 E\left((\mathrm{N}(0,1))^4\right) - \rho^2 = 1-\rho^2+\rho^2\left(0^4+6\cdot 0^2 1^2 + 3(1^4)\right) - \rho^2 \\
&= 1-\rho^2+3\rho^2-\rho^2 = \boxed{1+\rho^2}
\end{aligned}
$$

29/06/2024

4.5.2 Variaciones del TCL

Si consideramos $(X_n)_{n\in\mathbb{N}}$ v.a. independientes no idénticas con $E(X_i) = \mu_i \wedge V(X_i) = \sigma_i^2$

$$
\Longrightarrow \text{La suma tipificada es } T_n := \frac{\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mu_i}{\sqrt{\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2}} = \frac{\sum_{i=1}^n X_i - \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}}
$$

Ahora, según las condiciones de Lyapunov, definimos $S_n^\alpha := \sum_{i=1}^n \sigma_i^\alpha$, entonces

$$
\exists\delta>0 : \frac{1}{S_n^{2+\delta}}\sum_{i=1}^n E\left(|X_i-\mu_i|^{2+\delta}\right) \xrightarrow{n\to\infty} 0 \Longrightarrow T_n \xrightarrow{d} \mathrm{N}(0,1)
$$

Teorema 4.6 (Cota de Berry-Esseen)**.** *Sea $(X_n)_{n\in\mathbb{N}}$ una sucesión de v.a. i.i.d. con $E(X) = \mu \wedge V(X) = \sigma^2 < \infty$ y $E\left(|X|^3\right) < \infty$.*

$$
\Longrightarrow \sup_{t\in\mathbb{R}}\left|P\left(\frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \le t\right) - \Phi(t)\right| \le C\frac{E\left(|X|^3\right)}{\sigma^3}\frac{1}{\sqrt{n}} \text{ con } C \text{ constante universal}
$$

4.6 Funciones generatrices de momentos

Recordamos: Sea X v.a. que toma valores en $\{0,1,\dots\}$ con probabilidades $p_0, p_1, \dots,$ entonces la función generatriz de probabilidad de X es $G_X(s) := \sum_{n=0}^{\infty} p_n s^n$.

Definición 4.4. Sea X v.a., su función generatriz de momentos es $M_X(t) := E\left(e^{tX}\right)$.

Observación 4.2.
- Si X toma valores en $\{0,1,\dots\}$, entonces $M_X(t) = G_X(e^t)$.

$$
\text{porque } M_X(t) = E\left(e^{tX}\right) = \sum_{n=0}^{\infty} e^{tn}p_n = \sum_{n=0}^{\infty} p_n(e^t)^n = G_X(e^t)
$$

- Si X es discreta $\implies E\left(e^{tX}\right) = \displaystyle\sum_{x\in X(\Omega)} e^{tx} P(X=x)$.

  Si X es continua $\implies E\left(e^{tX}\right) = \displaystyle\int_{-\infty}^{\infty} e^{tx} f_X(x)\,\mathrm{d}x$.

Ejemplo 4.4.

- $X \sim \mathrm{EXP}\,(\lambda) \implies M_X(t) = \displaystyle\int_0^{\infty} e^{tx}\lambda e^{-\lambda x}\,\mathrm{d}x = \begin{cases} \frac{\lambda}{\lambda - t} & \text{si } t<\lambda \\ \infty & \text{si } t \geq \lambda \end{cases}$

  porque $E\left(e^{tX}\right) = \displaystyle\int_0^{\infty} e^{tx}\lambda e^{-\lambda x}\,\mathrm{d}x = \lambda \int_0^{\infty} e^{-(\lambda-t)x}\,\mathrm{d}x = \begin{cases} \frac{\lambda}{\lambda - t} & \text{si } t<\lambda \\ \infty & \text{si } t \geq \lambda \end{cases}$

- $X \sim \mathrm{N}\,(0,1) \implies M_X(t) = \displaystyle\int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx-\frac{x^2}{2}}\,\mathrm{d}x$

  $\implies M_X(t) = \displaystyle\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} e^{\frac{t^2}{2}}\,\mathrm{d}x = e^{\frac{t^2}{2}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2}\,\mathrm{d}x = e^{\frac{t^2}{2}}$

30/04/2024

Ejemplo 4.5. X sigue una distribución de Cauchy $\iff \forall x \in \mathbb{R} : f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$.

$$\implies M_X(t) = \int_{-\infty}^{\infty} e^{tx}\frac{1}{\pi}\frac{1}{1+x^2}\,\mathrm{d}x \text{ que solo converge cuando } t=0$$

Es decir, muchas veces M_X genera problemas porque e^{tX} se hace grande.
Por tanto, se define la transformada de Fourier de X como $\hat{f}_X(t) := \displaystyle\int_{-\infty}^{\infty} e^{itx} f_X(x)\,\mathrm{d}x$. En este caso, $\forall t,x \in \mathbb{R} : |e^{itx}| = 1$.

4.6.1 ¿Por qué ese nombre?

Si tenemos en cuenta el desarrollo de Taylor de $e^z = \displaystyle\sum_{n=0}^{\infty}\frac{z^n}{n!}$, entonces

$$\begin{aligned}\forall t \in \mathbb{R} : M_X(t) = E\left(e^{tX}\right) &= E\left(1 + tx + \frac{t^2x^2}{2} + \frac{t^3x^3}{3!} + \dots\right) \\ &= 1 + tE(X) + \frac{t^2E(X^2)}{2} + \frac{t^3E(X^3)}{3!} + \dots = \sum_{n=0}^{\infty}\frac{E\left(X^n\right)t^n}{n!}\end{aligned}$$

Es decir, $M_X(t)$ es función generatriz (exponencial) de la sucesión de momentos.

Teorema 4.7. *Sea X v.a. con función generatriz de momentos $M_X(t)$.*

$$M_X(t) \text{ converge en } |t|<\delta \text{ para cierto } \delta \implies \forall k \in \mathbb{Z}_{\geq 0} : E\left(X^k\right) = M_X^{(k)}(0)$$

Ejemplo 4.6. Sea $X \sim \mathrm{N}(0,1) \implies M_X(t) = e^{\frac{t^2}{2}}$.

$$\begin{aligned}
M'_X(t) &= te^{\frac{t^2}{2}} && \implies M'_X(0) = 0 \\
M''_X(t) &= e^{\frac{t^2}{2}} + t^2 e^{\frac{t^2}{2}} && \implies M''_X(0) = 1 \\
M'''_X(t) &= 2te^{\frac{t^2}{2}} + t^3 e^{\frac{t^2}{2}} && \implies M'''_X(0) = 0 \\
M^{(4)}_X(t) &= 3e^{\frac{t^2}{2}} + 3t^2 e^{\frac{t^2}{2}} + t^4 e^{\frac{t^2}{2}} && \implies M^{(4)}_X(0) = 3
\end{aligned}$$

4.6.2 La gracia

- Sea X v.a. con función generatriz de momentos $M_X(t)$. Definimos $Y := aX + b$ con $a, b \in \mathbb{R} \implies M_Y(t) = E\left(e^{tY}\right) = E\left(e^{t(aX+b)}\right) = e^{tb} E\left(e^{(at)X}\right) = e^{tb} M_X(at)$.
- Sean X e Y v.a. independientes con funciones generatrices de momentos $M_X(t)$ y $M_Y(t) \implies M_{X+Y}(t) = E\left(e^{t(X+Y)}\right) = E\left(e^{tX} e^{tY}\right) = E\left(e^{tX}\right) E\left(e^{tY}\right) = M_X(t) M_Y(t)$.

  Si, en su lugar, tenemos $X_1, \ldots, X_n$ v.a. independientes con funciones generatrices de momentos $M_{X_i}(t)$, entonces $M_{X_1+\cdots+X_n}(t) = M_{X_1}(t) \cdots M_{X_n}(t)$.

4.6.3 El problema de los momentos

Supongamos que para X v.a. tenemos todos sus momentos, ¿podemos recuperar X?

La respuesta es que no. Por ejemplo, si X sigue una distribución dada por la función de densidad $\forall x > 0 : f_X(x) = \dfrac{1}{x\sqrt{2\pi}} e^{-\frac{1}{2}\ln^2(x)}$.

Entonces, la familia de variables aleatorias $\{X_a : a \in (-1,1)\}$ dada por la familia de funciones de densidad $\{f_{X_a}(x) = (1 + a\sin(2\pi \ln x) f_X(x) : a \in (-1,1)\}$ tiene los mismos momentos.

Teorema 4.8. *(Unicidad) Sea $M_X(t) = E\left(e^{tX}\right) < \infty$ para $|t| < \delta$ para cierto $\delta > 0$.*

$$\implies \exists! X \text{ v.a. con función generatriz de momentos } M_X$$

En este caso, además, $\forall k \in \mathbb{Z}_{\geq 0} : E\left(X^k\right) < \infty$ y $\forall t : |t| < \delta : M_X(t) = \displaystyle\sum_{k=0}^{\infty} \frac{E\left(X^k\right) t^k}{k!}$.

Ejemplo 4.7. Sean $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ y $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$ v.a. independientes.

$$\implies M_{X_1+X_2}(t) = M_{X_1}(t) M_{X_2}(t) = e^{\mu_1 t + \frac{1}{2}\sigma_1^2 t^2} e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2} = e^{(\mu_1+\mu_2)t + \frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}$$

$$\implies X_1 + X_2 \sim \mathrm{N}\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$$

Sean $a, b \in \mathbb{R}$, definimos $Z := aX_1 + bX_2$.

$$\begin{aligned} \implies M_Z(t) &= M_{aX_1+bX_2}(t) = M_{X_1}(at)M_{X_2}(bt) = e^{a\mu_1 t + \frac{1}{2}a^2\sigma_1^2 t^2} e^{b\mu_2 t + \frac{1}{2}b^2\sigma_2^2 t^2} \\ &= e^{(a\mu_1 + b\mu_2)t + \frac{1}{2}(a^2\sigma_1^2 + b^2\sigma_2^2)t^2} \end{aligned}$$

$$\implies Z \sim \mathrm{N}\left(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2\right)$$

06/05/2024

Demostración (del TCL). Sean $(X_n)_{n\in\mathbb{N}}$ v.a. i.i.d. con X de referencia y $E(X) = \mu \wedge V(X) = \sigma^2$.

Suponemos, <u>además</u>, que $M_X(t)$ está bien definida en $|t| < \delta$ para cierto $\delta > 0$. (esto no estaba entre las hipótesis del teorema).

Definimos $(U_n)_{n\in\mathbb{N}}$ como $\forall n \in \mathbb{N} : U_n := X_n - \mu \implies E(U_n) = 0 \wedge V(U_n) = \sigma^2$ y son i.i.d.

Definimos $S_n := \displaystyle\sum_{i=1}^{n} X_i \implies E(S_n) = n\mu \wedge V(S_n) = n\sigma^2$

$$\implies \tilde{S}_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} U_i \implies E(\tilde{S}_n) = 0 \wedge V(\tilde{S}_n) = 1$$

$$\implies M_{\tilde{S}_n}(t) = E\left(e^{t\tilde{S}_n}\right) = E\left(e^{\frac{t}{\sigma\sqrt{n}}\sum_{i=1}^{n} U_i}\right) = E\left(e^{\frac{t}{\sigma\sqrt{n}}U_1} \cdots e^{\frac{t}{\sigma\sqrt{n}}U_n}\right) = \left(M_U\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n$$

$$M_U(x) = \sum_{k=0}^{\infty} \frac{E(U^k)}{k!}x^k = 1 + \cancelto{0}{E(U)}x + \frac{E(U^2)}{2}x^2 + o(x^2) = 1 + \frac{\sigma^2}{2}x^2 + o(x^2)$$

$$\implies \forall t \in \mathbb{R} : M_{\tilde{S}_n}(t) = \left(1 + \frac{\cancel{\sigma^2}}{2}\frac{t^2}{\cancel{\sigma^2}n} + o\left(\frac{1}{n}\right)\right)^n = \left(1 + \frac{t^2/2}{n} + o\left(\frac{1}{n}\right)\right)^n \xrightarrow{n\to\infty} e^{\frac{t^2}{2}}$$

Como $e^{\frac{t^2}{2}}$ es la función generatriz de momentos de una $\mathrm{N}(0,1)$, entonces $\tilde{S}_n \xrightarrow{d} \mathrm{N}(0,1)$. ■

Para que la demostración anterior fuese válida, habría que primero haber demostrado el siguiente teorema.

Teorema 4.9. *Sean $Z_1, Z_2, \ldots$ v.a. con funciones generatrices de momentos $M_{Z_i}(t)$ definidas entorno al origen. Si $\forall t \in \mathbb{R} : \lim_{n\to\infty} M_{Z_n}(t) = M(t)$, entonces existsZ v.a. con función generatriz de momentos $M(t) : Z_n \xrightarrow{d} Z$*

4.7 Función característica

Definición 4.5 (Función característica). Sea X v.a., $\forall t \in \mathbb{R} : \phi_X(t) := E\left(e^{itX}\right)$.

¿Cómo se calcula? Suponemos que X tiene función de densidad $f_X(x)$.

$$\implies E\left(e^{itX}\right) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\,\mathrm{d}x \text{ que es una función de } \mathbb{R} \text{ a } \mathbb{C}$$

Ejemplo 4.8. Sea $X \sim \mathrm{BER}\,(p)$ con $p \in (0,1)$.

$$\implies \phi_X(t) = E\left(e^{itX}\right) = e^{it\cdot 0}(1-p) + e^{it\cdot 1}p = (1-p) + pe^{it}$$

¿Por qué esto mola?

1. $\forall t \in \mathbb{R} : |\phi_X(t)| \leq 1$.

2. Siempre está bien definida.

3. Sean X e Y v.a. independientes con funciones características $\phi_X(t)$ y $\phi_Y(t)$.
$$\implies \phi_{aX+b}(t) = E\left(e^{it(aX+b)}\right) = e^{itb}E\left(e^{iatX}\right) = e^{itb}\phi_X(at)$$
$$\implies \phi_{X+Y}(t) = E\left(e^{it(X+Y)}\right) = E\left(e^{itX}e^{itY}\right) = E\left(e^{itX}\right)E\left(e^{itY}\right) = \phi_X(t)\phi_Y(t)$$

4. (Unicidad) Sean X e Y con funciones características $\phi_X(t)$ y $\phi_Y(t)$.
$$\forall t \in \mathbb{R} : \phi_X(t) = \phi_Y(t) \iff X \text{ e } Y \text{ tienen la misma distribución}$$

5. (Continuidad) Sean $(Z_n)_{n\geq 1}$ v.a. con funciones características $\phi_{Z_n}(t)$.
$$z_n \xrightarrow{d} Z \iff \forall t \in \mathbb{R} : \lim_{n\to\infty} \phi_{Z_n}(t) = \phi_Z(t)$$

Ejemplo 4.9.

- $X \sim \mathrm{EXP}\,(\lambda) \implies \phi_X(t) = \int_0^{\infty} e^{itx}\lambda e^{-\lambda x}\,\mathrm{d}x = \dfrac{\lambda}{\lambda - it}$
- $X \sim \mathrm{N}\,(0,1) \implies \phi_X(t) = \int_{-\infty}^{\infty} e^{itx}\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,\mathrm{d}x = e^{-\frac{t^2}{2}}$

07/05/2024

$$\phi_X(t) = E\left(e^{itX}\right) = \int_{-\infty}^{\infty} e^{itx}f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty}\cos(tx)f_X(x)\,\mathrm{d}x + i\int_{-\infty}^{\infty}\sin(tx)f_X(x)\,\mathrm{d}x$$

Teorema 4.10. *Sea X v.a. con función generatriz de momentos $M_X(t)$ definida en un entorno del origen $\implies \phi_X(t) = M_X(it)$.*

Por ejemplo, $X \sim \mathrm{N}\left(\mu, \sigma^2\right) \implies \phi_X(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$.

Demostración (del TCL). Sean $(X_n)_{n\in\mathbb{N}}$ v.a. i.i.d. con X de referencia y $E(X) =$

$\mu \wedge V(X) = \sigma^2$.

$$\phi_{Z_n}(t) = \phi_{S_n}\left(\frac{t}{n}\right) = \phi_X\left(\frac{t}{n}\right)^n$$

Desarrollando por Taylor tenemos: $\phi_X(w) = 1 + itE(X) + o(w)$.

$$\implies \phi_{Z_n}(t) = \left(1 + \frac{it\mu}{n} + o\left(\frac{1}{n}\right)\right)^n \xrightarrow{n\to\infty} e^{it\mu}$$

Si X es una v.a. constante ($X = \mu$) $\implies \phi_X(t) = e^{it\mu}$. Por tanto, $Z_n \xrightarrow{d} \mu$.

Ahora, podemos demostar la ley débil de los grandes números.
Basta ver que $Z_n \xrightarrow{d} \mu \implies Z_n \xrightarrow{P} \mu$ cuando μ es constante (ejs 7 y 9 de la hoja 5).

Definimos $U_n := X_n - \mu \implies E(U_n) = 0 \wedge V(U_n) = \sigma^2$ y son i.i.d. Estudiemos

$$\tilde{S}_n := \frac{\sum_{j=1}^n X_j - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}}\sum_{j=1}^n U_j \implies \phi_{\tilde{S}_n}(t) = \left(\phi_U\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n$$

Por Taylor tenemos $\phi_U(w) = 1 + iwE(U) + \frac{(iw)^2}{2!}E(U^2) + o(w^2)$.

$$\implies \phi_{\tilde{S}_n}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \xrightarrow{n\to\infty} e^{-\frac{t^2}{2}}$$

■

4.7.1 Función característica para distribución de Cauchy: Variable compleja

Sea $X \sim$ Cauchy $\iff \forall x \in \mathbb{R} : f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}$. Queremos calcular su función característica.

$$\implies \phi_X(t) = \int_{-\infty}^{\infty} e^{itx}\frac{1}{\pi}\frac{1}{1+x^2}\,\mathrm{d}x = \cdots = e^{-|t|}$$

Veamos una técnica de variable compleja para calcular la integral. Consideramos $t > 0$.
Definimos $\forall z \in \mathbb{C} : g(z) := \dfrac{e^{itz}}{1+z^2}$ con un $t \in \mathbb{R}_{>0}$ fijo. Esta función tiene dos polos en i y $-i$ y hay un teorema en variable compleja que dice que si f es holomorfa en un entorno de un camino cerrado γ y f no tiene polos en el interior de γ, entonces $\int_\gamma f(z)\,\mathrm{d}z = 0$.
Si, en su defecto, f tiene un polo en z_0

$$\implies \int_\gamma f(z)\,\mathrm{d}z = 2\pi i \operatorname{Res}(f, z_0) \text{ donde } \operatorname{Res}(f, z_0) = \lim_{z\to z_0}(z - z_0)f(z)$$

Definimos $\Gamma_R \subset \mathbb{C}$ como el camino (orientado) que va de $-R$ a R por la recta real y luego vuelve a $-R$ por la semicircunferencia superior de radio R en $\mathbb{C}$ donde $R \in \mathbb{R}_{>1}$. Podemos separar este camino en dos partes: $\Gamma_R = \Gamma_R^1 \cup \Gamma_R^2$ donde Γ_R^1 es la parte de la recta real y Γ_R^2 es la semicircunferencia superior y se cumple que:

$$\int_{\Gamma_R} g(z)\,\mathrm{d}z = \int_{\Gamma_R^1} g(z)\,\mathrm{d}z + \int_{\Gamma_R^2} g(z)\,\mathrm{d}z = 2\pi i \operatorname{Res}(g, i)$$

Tenemos $\mathrm{Res}(g,i) = \lim_{z\to i} \cancel{(z-i)} \dfrac{e^{itz}}{(z+i)\cancel{(z-i)}} = \dfrac{e^{-t}}{2i} \implies \displaystyle\int_{\Gamma_R} g(z)\,\mathrm{d}z = 2\pi i \frac{e^{-t}}{2i} = \pi e^{-t}$.

Observamos que, claramente, $\displaystyle\lim_{R\to\infty} \int_{\Gamma_R^1} g(z)\,\mathrm{d}z = \int_{-\infty}^{\infty} e^{itx}\frac{1}{1+x^2}\,\mathrm{d}x = \pi\cdot\phi_X(t)$. Ahora

$$\left|\int_{\Gamma_R^2} g(z)\,\mathrm{d}z\right| \leq \int_{\Gamma_R^2} |g(z)|\,\mathrm{d}z \leq \pi R \max_{z\in\Gamma_R^2} |g(z)| = \pi R \max_{z\in\Gamma_R^2} \left|\frac{e^{itz}}{1+z^2}\right|$$

Como $z\in\Gamma_R^2 \implies |z| = R$, entonces $\left|e^{itz}\right| \leq \left|e^{itRe^{i\theta}}\right| = \left|e^{itR\cos\theta}e^{-tR\sin\theta}\right|$, y como $t>0$

$$-tR\sin\theta \leq 0 \implies \left|\frac{e^{itz}}{1+z^2}\right| \leq \frac{\left|e^{itR\cos\theta}\right|}{|1+z^2|} \leq \frac{1}{|1+z^2|}$$

Ahora, dado que $|1+z^2| \geq |z^2| - 1 = R^2-1$, tenemos $\left|\dfrac{e^{itz}}{1+z^2}\right| \leq \dfrac{1}{R^2-1}$.

$$\implies \left|\int_{\Gamma_R^2} g(z)\,\mathrm{d}z\right| \leq \pi R \max_{z\in\Gamma_R^2}\left|\frac{e^{itz}}{1+z^2}\right| \leq \frac{\pi R}{R^2-1} \implies \lim_{R\to\infty}\int_{\Gamma_R^2} g(z)\,\mathrm{d}z = 0$$

Concluimos que para todo t positivo, $\phi_X(t) = e^{-t}$:

$$\pi e^{-t} = \lim_{R\to\infty}\int_{\Gamma_R} g(z)\,\mathrm{d}z = \lim_{R\to\infty}\left[\int_{\Gamma_R^1} g(z)\,\mathrm{d}z + \int_{\Gamma_R^2} g(z)\,\mathrm{d}z\right] = \pi\phi_X(t) + 0 \implies \phi_X(t) = e^{-t}$$

Para $t<0$ se puede hacer un razonamiento análogo con un Γ que vaya por la semicircunferencia inferior y luego vuelva por la recta real. Finalmente llegamos a

$$\forall t>0 : \phi(t) = e^{-t} \;\wedge\; \forall t<0 : \phi(t) = e^{t} \implies \boxed{\forall t\in\mathbb{R} : \phi(t) = e^{-|t|}}$$

Como además, tenemos $\displaystyle\phi(t) = \int_{-\infty}^{\infty} e^{itx}\frac{1}{\pi}\frac{1}{1+x^2}\,\mathrm{d}x = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\cos(tx)}{1+x^2}\,\mathrm{d}x + i\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin(tx)}{1+x^2}\,\mathrm{d}x$

$$\implies \int_{-\infty}^{\infty}\frac{\cos(tx)}{1+x^2}\,\mathrm{d}x = \pi e^{-|t|} \;\wedge\; \int_{-\infty}^{\infty}\frac{\sin(tx)}{1+x^2}\,\mathrm{d}x = 0$$

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 GEOM$\,(p)$.

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

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

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

$$\implies T_1 = 1 \wedge T_2 \sim \text{GEOM}\left(\frac{n-1}{n}\right) \wedge T_3 \sim \text{GEOM}\left(\frac{n-2}{n}\right) \wedge \cdots$$

$$\implies \forall i \in \mathbb{N}_n : T_i \sim \text{GEOM}\left(1 - \frac{i-1}{n}\right) \implies E(T_i) = \frac{n}{n-i}$$

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

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

$$\implies E(T) = nH_n \approx n \ln n$$

5.3 Hoja 3

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

$$\implies E(Y) = E\left(e^X\right) = \int_{-\infty}^{\infty} e^x \phi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{x - \frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2x+1}{2}+1} dx$$

$$= e^{\frac{1}{2}} \left(\cancelto{1}{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-1)^2}{2}} dx} \right) = e^{\frac{1}{2}}$$

$$\implies V(Y) = E(Y^2) - E(Y)^2 = E(e^{2X}) - e = e^2 - e = e(e-1)$$

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

$$\implies E(e^X) = E\left(e^{\mu + \sigma Z}\right) = e^{\mu} E\left(e^{\sigma Z}\right) = e^{\mu + \frac{\sigma^2}{2}}$$

$$\implies V\left(e^X\right) = E\left(e^{2X}\right) - E\left(e^X\right)^2 = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2} = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right)$$

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

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

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 $X \sim \mathrm{N}(0,1)$.

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

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

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

15. Sea $X \sim \mathrm{N}(0,1)$. 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 $X \sim \mathrm{N}(\mu, \sigma^2)$ 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 $T \sim \mathrm{lognormal}(\mu, \sigma^2)$ una variable aleatoria que mide la longitud de una conferencia (en minutos).

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

$$\implies \begin{cases} \frac{\ln 40 - \mu}{\sigma} = \Phi^{-1}(0.40) \implies \mu = \ln 40 - \Phi^{-1}(0.40)\sigma \\ \frac{\ln 50 - \mu}{\sigma} = \Phi^{-1}(0.45) \implies \mu = \ln 50 - \Phi^{-1}(0.45)\sigma \end{cases}$$

$$\implies \sigma = \frac{\ln 50 - \ln 40}{\Phi^{-1}(0.45) - \Phi^{-1}(0.40)}$$

5.4 Hoja 4

4. Sean X, Y dos variables aleatorias independientes idénticas ($f := f_X \equiv f_Y$).
Definimos $M := \max X, Y \wedge m := \min X, Y$.

$$\begin{aligned} F_m(z) &= P(\min X, Y \leq z) = 1 - P(\min X, Y > z) = 1 - P(X > z \wedge Y > z) \\ &= 1 - P(X > z) \cdot P(Y > z) = 1 - P(X > z)^2 = 1 - (1 - F(z))^2 \end{aligned}$$

$$\implies f_M(z) = \frac{\mathrm{d}}{\mathrm{d}z} F_M(z) = 2\,(1 - F(z))\, f(z)$$

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

Si ahora tenemos $X_1, \ldots, X_n$ independientes con F_i, f_i, $\begin{matrix} M := \max\{X_1, \cdots, X_n\} \\ m := \min\{X_1, \cdots, X_n\} \end{matrix}$:

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

6. Sean $X, Y \sim \mathrm{EXP}(\lambda)$ 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 $Z := \dfrac{Y}{X}$

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

Demostración.

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

■

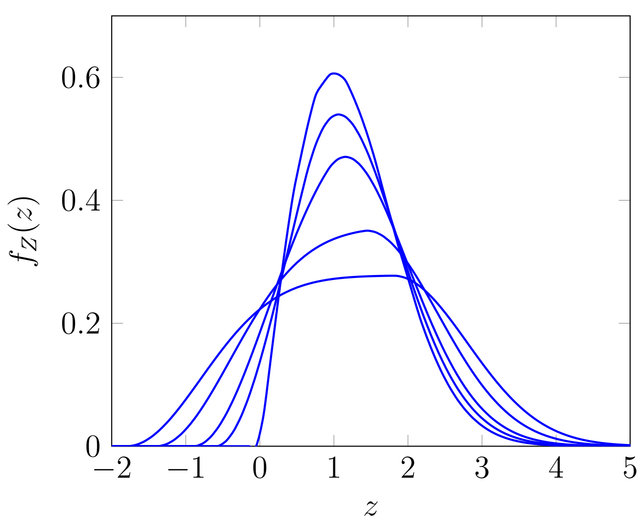
8. Sean X e Y dos v.a. indep. con $f_X = xe^{-\frac{x^2}{2}} \cdot \mathbb{1}_{\{x>0\}}(x)$ e $Y \sim \text{UNIF}\left([-\varepsilon, \varepsilon]\right)$
$\implies f_Y(y) = \frac{1}{2\varepsilon} \cdot \mathbb{1}_{\{-\varepsilon<y<\varepsilon\}}(y)$. Definimos $Z := X + Y$.

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

$$\forall z \ge -\varepsilon : f_Z(z) = \int_{-\infty}^{\infty} f_X(u) f_Y(z-u)\,\mathrm{d}u = \int_{-\infty}^{\infty} ue^{-\frac{u^2}{2}} \cdot \mathbb{1}_{\{u>0\}}(u) \cdot \frac{1}{2\varepsilon} \cdot \mathbb{1}_{\{-\varepsilon<z-u<\varepsilon\}}(u)\,\mathrm{d}u$$

$$= \begin{cases} \dfrac{1}{2\varepsilon} \displaystyle\int_{-\infty}^{\infty} ue^{-\frac{u^2}{2}} \cdot \mathbb{1}_{\{0<u<z+\varepsilon\}}(u)\,\mathrm{d}u & z < \varepsilon \\ \dfrac{1}{2\varepsilon} \displaystyle\int_{-\infty}^{\infty} ue^{-\frac{u^2}{2}} \cdot \mathbb{1}_{\{z-\varepsilon<u<z+\varepsilon\}}(u)\,\mathrm{d}u & z \ge \varepsilon \end{cases} = \begin{cases} \dfrac{1}{2\varepsilon} \displaystyle\int_{0}^{z+\varepsilon} ue^{-\frac{u^2}{2}}\,\mathrm{d}u & z < \varepsilon \\ \dfrac{1}{2\varepsilon} \displaystyle\int_{z-\varepsilon}^{z+\varepsilon} ue^{-\frac{u^2}{2}}\,\mathrm{d}u & z \ge \varepsilon \end{cases}$$

$$\implies \forall z \ge -\varepsilon : f_Z(z) = \begin{cases} \dfrac{1}{2\varepsilon} \left[-e^{-\frac{u^2}{2}}\right]_0^{z+\varepsilon} & z < \varepsilon \\ \dfrac{1}{2\varepsilon} \left[-e^{-\frac{u^2}{2}}\right]_{z-\varepsilon}^{z+\varepsilon} & z \ge \varepsilon \end{cases} = \begin{cases} \dfrac{1}{2\varepsilon} \left(1 - e^{-\frac{(z+\varepsilon)^2}{2}}\right) & z < \varepsilon \\ \dfrac{1}{2\varepsilon} \left(e^{-\frac{(z-\varepsilon)^2}{2}} - e^{-\frac{(z+\varepsilon)^2}{2}}\right) & z \ge \varepsilon \end{cases}$$



5.5 Hoja 5

9. Sean $(Z_n)_{n\in\mathbb{N}}$ una sucesión de v.a. y quiero probar que $Z_n \xrightarrow[n\to\infty]{\text{cuad}} \implies Z_n \xrightarrow[n\to\infty]{P} Z$.

Demostración. Por la desigualdad de Markov:

$$P\left(|Z_n - Z| > \varepsilon\right) = P\left(|Z_n - Z|^2 > \varepsilon^2\right) \leq \frac{E\left(|Z_n - Z|^2\right)}{\varepsilon^2} \xrightarrow{n\to\infty} 0$$

■

Ahora queremos demostrar que $Z_n \xrightarrow[n\to\infty]{P} Z \implies Z_n \xrightarrow[n\to\infty]{d} Z$.

Demostración. Sea $t \in \mathbb{R}$ un punto de continuidad de F_Z:

$$\begin{aligned}\forall\varepsilon > 0 : P\left(Z_n \leq t\right) &= P\left(Z_n \leq t \wedge Z \leq t+\varepsilon\right) + P\left(Z_n \leq t \wedge Z > t+\varepsilon\right)\\ &\leq P\left(Z \leq t+\varepsilon\right) + P\left(|Z_n - Z| > \varepsilon\right)\end{aligned}$$

Análogamente:

$$\begin{aligned}\forall\varepsilon > 0 : P\left(Z \leq t-\varepsilon\right) &= P\left(Z \leq t-\varepsilon \wedge Z_n \leq t\right) + P\left(Z \leq t-\varepsilon \wedge Z_n > t\right)\\ &\leq P\left(Z_n \leq t\right) + P\left(|Z_n - Z| > \varepsilon\right)\end{aligned}$$

Para todo $t \in \mathbb{R}$ punto de continuidad de F_Z y para todo $\varepsilon > 0$:

$$P\left(Z \leq t-\varepsilon\right) + P\left(|Z_n - Z| > \varepsilon\right) \leq P\left(Z_n \leq t\right) \leq P\left(Z \leq t+\varepsilon\right) + P\left(|Z_n - Z| > \varepsilon\right)$$

$$\implies F_Z(t-\varepsilon) = P\left(Z \leq t-\varepsilon\right) \leq \lim_{n\to\infty} P\left(Z_n \leq t\right) \leq P\left(Z \leq t+\varepsilon\right) = F_Z(t+\varepsilon)$$

$$\implies \lim_{n\to\infty} F_{Z_n}(t) = F_Z(t) \implies Z_n \xrightarrow[n\to\infty]{d} Z \text{ porque } F_Z \text{ es continua en } t$$

■

11. Sean $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$ dos sucesiones de v.a. tales que $X_n \xrightarrow[n\to\infty]{\text{cuad}} X \wedge Y_n \xrightarrow[n\to\infty]{\text{cuad}} Y$.

Queremos ver que $X_n + Y_n \xrightarrow[n\to\infty]{\text{cuad}} X + Y$.

Demostración.

$$E\left(|X_n + Y_n - X - Y|^2\right) = E\left(|X_n - X|^2\right) + E\left(|Y_n - Y|^2\right) + 2E\left((X_n - X)(Y_n - Y)\right)$$

Ahora, por Cauchy-Schwarz:

$$2E\left((X_n - X)(Y_n - Y)\right) \leq \sqrt{E\left(|X_n - X|^2\right) E\left(|Y_n - Y|^2\right)}$$

■

Ahora queremos demostrar que $X_n \xrightarrow[n\to\infty]{P} X \wedge Y_n \xrightarrow[n\to\infty]{P} Y \implies X_n + Y_n \xrightarrow[n\to\infty]{P} X + Y$.

Demostración.

$$\begin{aligned}\forall\varepsilon > 0 : P\left(|X_n + Y_n - X - Y| > \varepsilon\right) &= P\left(|X_n - X + Y_n - Y| > \varepsilon\right)\\ &\leq P\left(\{|X_n - X| > \varepsilon/2\} \cup \{|Y_n - Y| > \varepsilon/2\}\right)\\ &\leq P\left(|X_n - X| > \varepsilon/2\right) + P\left(|Y_n - Y| > \varepsilon/2\right)\end{aligned}$$

Como $X_n \xrightarrow[n\to\infty]{P} X \wedge Y_n \xrightarrow[n\to\infty]{P} Y$: ■

17. Sea $(X_n)_{n\in\mathbb{N}}$ una sucesión de v.a. independientes con $E(X_i) = m_i \wedge V(X_i) = \sigma_i^2 < R \in \mathbb{R}$. Definimos $M_n = \frac{1}{n}\sum_{i=1}^n M_i$ y queremos ver que

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - M_n\right| < \varepsilon\right) \xrightarrow{n\to\infty} 1$$

Por la desigualdad de Chebyshev:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - M_n\right| > \varepsilon\right) \leq \frac{V\left(\frac{1}{n}\sum_{i=1}^n X_i)\right)}{\varepsilon^2} = \frac{\frac{1}{n^2}\sum_{j=1}^n \sigma_j^2}{\varepsilon^2} \leq \frac{R}{n^2}\frac{n}{\varepsilon^2} = \frac{R}{n\varepsilon^2} \xrightarrow{n\to\infty} 0$$