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

1 Tema 1: Sucesos y probabilidades

1.1 Formalizando

Definición 1.1 (Espacio muestral). En un experimento aleatorio, el espacio muestral es el conjunto (no vacío) de sus posibles resultados y se denota por 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) $(\Omega = \{\omega_1, \dots, \omega_N\} \implies |\Omega| = N) \wedge \big(\mathcal{F} = P(\Omega) \to |\mathcal{F}| = 2^N\big)$.
   Dato: $p_1, \dots, p_N \geq 0 \implies \sum_{j=1}^{N} p_j = 1$. Asignamos $\forall j \in \mathbb{N}_N : p_j := P(\{\omega_j\})$.
   Definimos $\forall A \subset \Omega : P(A) = \sum_{\omega \in A} P(\omega)$.

3. Lanzas n veces una moneda. Dato: $p \in (0,1)$.
$$\implies \Omega = \{111\cdots 1, \dots, 000\cdots 0\} \wedge |\Omega| = 2^N \wedge \text{escogemos } \mathcal{F} = P(\Omega)$$
$$\implies \forall \omega \in \Omega : P(\omega) = p^{\#\text{unos de } \omega}(1-p)^{\#\text{ceros de } \omega}$$
   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: $\sum_{j=1}^{\infty} p(1-p)^{j-1} = p \sum_{k=0}^{\infty} (1-p)^k = p \frac{1}{1-(1-p)} = 1$

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.

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

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

$\blacksquare$

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 $\iff \forall J \subset \mathbb{N}_N : P\left(\bigcap_{j \in I} A_j\right) = \prod_{j \in J} P(A_j)$

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

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.

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

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

■

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?

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

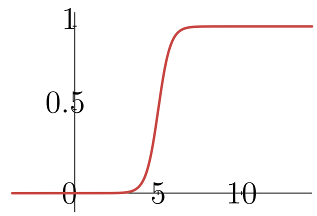
$$\implies P(5b5n|U_1) = \binom{10}{5}0.8^5 0.2^5 = P(5b5n|U_2) \implies P(U_1|5b5n) = \frac{1}{2}$$

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

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

Si dibujamos la gráfica de la función

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

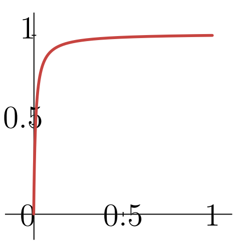


podemos ver que el cambio es muy brusco. Es decir, una pequeña diferencia en la evidencia puede cambiar mucho la probabilidad de que se haya usado una urna u otra.

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

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

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



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

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

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

Ejemplo 1.6 (Sobre independencia). $(\Omega, \mathcal{F}, P) \wedge A_1, \dots, A_n$ sucesos independientes tal que $\forall j \in \mathbb{N}_n : P(A_j) = \frac{1}{n}$. ¿Qué sabemos sobre $P\left(\bigcup_{j=1}^n A_j\right)$?

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

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

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

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*

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

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

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.

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

■

2 Tema 2: Variables aleatorias discretas

Definición 2.1. Sea $(\Omega, \mathcal{F}, P)$ un espacio de probabilidades, $X\colon \Omega \longrightarrow \mathbb{R}$ 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

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

Teorema 2.1. *Sea $S = \{x_1, x_2, \dots\}$ 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 $\{1, \cdots, N\}, N \geq 2$ ($X \sim \mathrm{UNIF}\,(N)$).

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

   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.

3. X sigue una distribución **binomial** de parámetros $n \geq 1 \wedge p \in (0,1)$ $(X \sim \mathrm{BIN}\,(n,p))$
$$\iff S = \{0,1,\dots,n\} \wedge \forall j \in \{0,1,\dots,n\} : P(x=j) = \binom{n}{j} p^j (1-p)^{n-j}$$
Sirve para modelizar el número de caras que salen al lanzar n veces una moneda de probabilidad p.
$$n! \sim n^n e^{-n}\sqrt{2\pi n} \implies \binom{n}{n/2} \sim \frac{n^n e^{-n}\sqrt{2\pi n}}{(n/2)^{(n/2)} e^{-(n/2)}\sqrt{2\pi(n/2)}(n/2)^{(n/2)} e^{-(n/2)}\sqrt{2\pi(n/2)}}$$
$$\implies \frac{n^n\sqrt{n}}{(n/2)^{(n/2)}\sqrt{2\pi(n/2)}(n/2)^{(n/2)}\sqrt{(n/2)}} = \frac{n^n\sqrt{n}}{(n/2)^n\sqrt{2\pi}(n/2)} = \frac{n^n\sqrt{2}}{(n/2)^n\sqrt{\pi n}} = 2^n\sqrt{\frac{2}{\pi n}}$$

4. La variable X sigue una distribución **geométrica** de parámetro $p \in (0,1)$ $(X \sim \mathrm{GEOM}\,(p))$.
$$\iff S = \{1,2,\dots\} \wedge \forall j \geq 1 : P(x=j) = p(1-p)^{j-1}$$
Sirve para modelizar el número de lanzamientos hasta que sale un resultado C en cuestión con $P(X=C)=p$.

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

5. La variable X sigue una distribución de **Poisson** con parámetro $\lambda > 0$ $(X \sim \mathrm{POISSON}\,(\lambda))$
$$\iff S = \{0,1,\dots\} \wedge \forall j \geq 0 : P(x=j) = e^{-\lambda}\frac{\lambda^j}{j!}$$
Se usa para modelizar la frecuencia de eventos determinados durante un intervalo de tiempo fijado a partir de la frecuencia media de aparición de dichos eventos.

Proposición 2.1. *Sea $X \sim \mathrm{BIN}\,(n,p)$ una v.a.d.*
$$\implies \textit{cuando } n \textit{ es grande, } \mathrm{BIN}\,(n,p) \sim \mathrm{POISSON}\,(np)$$

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

• Dada cualquier serie convergente $\sum_{n=1}^{\infty} a_n = s$, se puede definir la variable aleatoria X tal que $S = \{1, 2, \dots\} \wedge P(x = k) = \frac{a_k}{s}$

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.

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

Ejemplo 2.4 (Cálculo de la esperanza).

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

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 $\forall k \geq 0 : P(X = k) = \frac{6}{\pi^2} \frac{1}{k^2}$

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

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

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

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

Teorema 2.2. *Sea X una v.a.d. que toma los valores x_j con probabilidades p_j para $j \geq 1$.*

Sea $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ 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 $i_{\text{s}} \geq 1 : g(x_i) = y$ 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. $X \sim \text{UNIF}\left(\{-1, 0, 1\}\right) \wedge Y = X^2 \implies E(X) = 0 \wedge E(Y) = {}^2\!/_3$
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 $Y = g(X) = e^X$

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

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 $\mu := E(X)$

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

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

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

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

   $$\implies V(X) = \frac{1}{N}\sum_{j=1}^{N} j^2 = \frac{1}{N} \cdot \frac{N(N+1)(2N+1)}{6} = \frac{(N+1)(N-1)}{12}$$
3. $X \sim \text{BIN}\,(n,p) \implies E(X) = np \wedge \boxed{V(X) = np(1-p)}$

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

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

   Demostración.

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

   $\blacksquare$

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.* $: E(X), V(X) < \infty.$

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

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

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

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

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

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

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 $\lambda := \dfrac{\alpha}{\sigma(X)}$

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

■

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.

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

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

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

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

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

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

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

2.3 Varias variables aleatorias

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

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

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

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

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

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

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

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

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

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

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

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

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

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

■

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

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

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

Observación 2.6. Si X e Y son independientes $\implies E(X \cdot Y) = E(X) \cdot E(Y)$.
Sin embargo, la implicación recíproca no es cierta. **(Motivo de excomunión)**

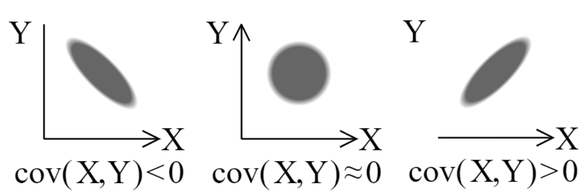
Definición 2.10 (Covarianza). En $(\Omega, \mathcal{F}, P)$, sean X e Y v.a.d., $\text{cov}(X,Y)$ es la covarianza de X e Y

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

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

De forma más general:

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

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

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

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

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

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

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.

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

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.

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

■

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*

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

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

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

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

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

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

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

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

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

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 $Z_n := \frac{1}{n}S_n = \frac{1}{n}(X_1 + \cdots + X_n) \implies E(Z_n) = \mu \wedge V(Z_n) = \frac{\sigma^2}{n} \xrightarrow{n\to\infty} 0$

$$\implies Z_n \text{ no es aleatoria si } n \to \infty \text{ (ley de los grandes números)}$$

05/03/2024

2.4 Funciones generatrices de probabilidad

2.4.1 Series de potencias

Sea $(a_n)_{n=0}^{\infty}$ una sucesión y $f(x) := \displaystyle\sum_{n=0}^{\infty} a_n x^n$ 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. $X \sim \text{GEOM}\,(p) \implies G_X(s) = \sum_{j=1}^{\infty} (1-p)^{j-1} ps^j = \dfrac{ps}{1-(1-p)s}$

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

   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 $(p_n)_{n=0}^{\infty}$
   $$G_X(s) = \sum_{n=0}^{\infty} p_n s^n \implies G'_X(s) = \sum_{n=1}^{\infty} np_n s^{n-1} \implies G'_X(1) = \sum_{n=1}^{\infty} np_n = E(X)$$
   Si seguimos derivando, obetenemos
   $$\implies G''_X(s) = \sum_{n=2}^{\infty} n(n-1)p_n s^{n-2} = \sum_{n=2}^{\infty} n^2 p_n s^{n-2} - \sum_{n=2}^{\infty} np_n s^{n-2}$$
   $$\implies G''_X(1) = \sum_{n=2}^{\infty} n^2 p_n - \sum_{n=2}^{\infty} np_n = E(X^2) - E(X)$$
   $$\implies V(X) = E(X^2) - E(X)^2 = G''_X(1) + G'_X(1)\,(1 - G'_X(1))$$

   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:

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

■

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

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

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

06/03/2024

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

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

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

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

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

3 Ejercicios

3.1 Hoja 1

3.2 Hoja 2

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

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

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

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

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

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

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

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

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

■

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

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

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

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

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:

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

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