Cálculo de probabilidades I

21 de diciembre de 2025

A quien corresponda.

Índice general

1 El modelo matemático de la probabilidad

1.1. Espacio muestral y sucesos

Los resultados posibles constituyen un conjunto finito que se denomina **espacio muestral** del fenómeno aleatorio en cuestos y se designa genéricamente por Ω.

Los subconjuntos del espacio muestral Ω se denominan **sucesos** o **acontecimientos**, a los que puede dar lugar el azar en el experimento considerado. Los subconjuntos con un único elemento, representan uno sólo de los resultados posibles del fenómeno, se llaman **sucesos simples**, mientras que los **sucesos compuestos** se identifican con los subconjuntos que tienen más de un elemento y son, por tanto, uniones de sucesos simples.

El espacio muestral Ω es el **suceso seguro**, y el subconjunto $\emptyset$ es el **suceso imposible**.

Si la intersección de dos sucesos A y B es vacía $A \cap B = \emptyset$, se dice que A y B son **incompatibles**.

El **suceso contrario** es $A^c = \Omega - A$.

1.2. El concepto de probabilidad

Sobre un espacio muestral finito Ω, una probabilidad es una aplicación

$$P : P(\Omega) \longrightarrow [0, 1]$$

que verifique:

1. **Aditividad**: Si $A, B \subset \Omega, A \cap B = \emptyset$, entonces $P(A \cup B) = P(A) + P(B)$.
2. **Normalización**: $P(\Omega) = 1$.

El espacio de probabilidad finito es el par (Ω, P).

1.3. Primeras propiedades de la probabilidad

- Si $A \subset B \subset \Omega$, entonces $P(A) \leq P(B)$.
- $P(B) = P(A) + P(B \cap A^c)$ * Si $A \subset B \subset \Omega$, entonces $P(B - A) = P(B) - P(A)$.
- Para cualquier $A \subset \Omega$, $P(A^c) = 1 - P(A)$.
- $P(\emptyset) = 0$

- $P(B) = P(B \cap A) + P(B - A)$
- $P(A \cup B) = P(A) + P(B) + P(A \cap B)$.
- Si $A_1, A_2, \ldots, A_n$ son sucesos disjuntos dos a dos, es decir que verifiquen $A_i \cap A_j = \emptyset$ para $i \neq j$ se cumple

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

- Si $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ es el espacio muestral de probabilidad finito (Ω, P) y $A = \{\omega_1, \omega_2, \ldots, \omega_k\}$ es un suceso se cumple

$$P(A) = P(\{\omega_1\}) + P(\{\omega_2\}) + \cdots + P(\{\omega_k\})$$

2 Asignación de probabilidades

2.1. Regla de Laplace

La probabilidad de un suceso A relativo a un fenómeno aleatorio es

$$P(A) = \frac{\text{número de casos favorables de A}}{\text{número de casos posibles}}$$

supuesto que todos los casos posibles son igualmente probables.

3 Las formulas de inclusión y exclusión

Probabilidad de una unión de sucesos

Si $A_1, A_2, \ldots, A_n$ son sucesos cualesquiera de un espacio de probabilidad, se cumple

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = S_1 - S_2 + S_3 - \cdots + (-1)^{n-2} S_{n-1} + (-1)^{n-1} S_n$$

donde

$$S_k = \sum_{1 \leq i_1 < i_2 < \cdots < i_k < n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k})$$

Probabilidad de que se realicen m sucesos

En un espacio de probabilidad, si $A_1, A_2, \ldots, A_n$ son sucesos cualesquiera, la probabilidad $P_{[m]}$ de que ocurran exactamente m de ellos, es

$$P_{[m]} = S_m - \binom{m+1}{m} S_{m+1} + \binom{m+2}{m} S_{m+2} - \cdots + (-1)^{n-m} \binom{n}{m} S_n$$

4 Extensiones del modelo matemático

4.1. Espacios muestrales numerables

Los conjuntos finitos son **discretos** y los infinitos numerables son **numerables**.

Si Ω es infinito numerable, el conjunto de sucesos, $P(\Omega)$, no es numerable, si no que su cardinal es aún mayor ($2^{\mathbb{N}}$).

4.2. Propiedades adicionales de la probabilidad

Sobre un espacio muestral numerable Ω, una probabilidad es una aplicación

$$P : P(\Omega) \to [0, 1]$$

que verifique:

1. Si A_n es una sucesión de suceso tales que $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

2. $P(\Omega) = 1$ Si se cumplen ambas condiciones, (Ω, P) constituye un espacio de probabilidad numerable.

En las construcciones de espacios de probabilidad no finitos, la regla de Laplace para atribuir probabilidades resulta inservible directamente.

Modelo geométrico o de Pascal

$$\Omega = \mathbb{N} \qquad P(\{n\}) = p(1-p)^{n-1} \text{ para cada } n \in \mathbb{N}$$

4.3. Modelos continuos

No es posible que un modelo continuo asigne probabilidad a todos los subconjuntos de $\mathbb{R}$. Se toman en cuenta los intervalos de $\mathbb{R}$.

La probabilidad de un suceso $A \subset \mathbb{R}$ no puede obtenerse sumando la probabilidad de los puntos contenidos en A.

Una función $f : \mathbb{R} \to \mathbb{R}$ (con un número finito de discontinuidades) se denomina **densidad de probabilidad** si verifica 1. $f(x) \geq 0$ para todo $x \in \mathbb{R}$ 2. $\int_{\mathbb{R}} f(x)dx = 1$ El modelo probabilístico asociado a la función de densidad f asigna probabilidad

$$P(x) = \int_I f(x)dx$$

a cualquier intervalo I de $\mathbb{R}$.

Se puede aproximar mediante

$$P([x, x + \triangle x]) \simeq f(x) \triangle x$$

Distribución normal

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

5 Probabilidad condicionada

$$P(A|B) = \frac{\text{número de casos favorables de } A \text{ y } B}{\text{número de casos en los que ocurre } B}$$

Si A y B son sucesos de un cierto de espacio de probabilidad y se cumple que $P(B) > 0$, la probabilidad de A condicionada a B es

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

$$P(A)P(B|A) = P(B)P(A|B)$$

5.1. Propiedades

1. Si $A_1, A_2, \ldots, A_n, \ldots$ son sucesos disjuntos se cumple

$$\begin{aligned} &P(A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots |B) \\ &\qquad = \frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \cdots \cup (A_n \cap B) \cdots)}{P(B)} \\ &\qquad = \frac{P((A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B) + \cdots)}{P(B)} \\ &\qquad = P(A_1|B) + P(A_2|B) + \cdots + P(A_n|B) + \cdots \end{aligned}$$

   de forma que $P(\cdot|B)$ es numeralmente aditiva.

2. Por otra parte

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = 1$$

   y más concretamente

$$P(B|B) = 1$$

Proposición 5.1 (Formula de las probabilidades totales) *Si $B_1, B_2, \ldots, B_n, \ldots$ es una familia de sucesos de probabilidades positivas, disjuntos dos a dos, tales que*

$$B_1 \cup B_2 \cup \cdots \cup B_n \cup \cdots = \Omega$$

se verifica

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_n)P(A|B_n) + \cdots$$

para cualquier suceso $A \subset \Omega$.

Proposición 5.2 (Formula de Bayes) *Si* $B_1, B_2, \ldots, B_n, \ldots$ *es una familia de sucesos de probabilidades positivas, disjuntos dos a dos, tales que*

$$B_1 \cup B_2 \cup \cdots \cup B_n \cup \cdots = \Omega$$

para cualquier suceso A *de probabilidad positiva, de acuerdo con la definición, se tiene*

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

lo cual, combinado con la fórmula de las probabilidades totales se puede expresar

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_n)P(A|B_n)}$$

5.2. El método recurrente

Una situación en la que obtener la probabilidad de un suceso dependa exclusivamente de lo ocurrido en la etapa anterior se llama situación **markoviana**.

6 Independencia de sucesos

6.1. Sucesos dependientes e independientes

Si un suceso A influye, favorablemente o desfavorablemente, en otro B, se dice que el segundo es **dependiente** del primero.

En un fenómeno aleatorio, el suceso A se dice **independiente** del suceso B (supuesto que B tiene probabilidad positiva) si

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

Si A y B son independientes, también lo son A y B^c.

$$P(A \cap B^c) = P(A)P(B^c)$$

6.2. Espacios producto

Dados dos espacios de probabilidad (Ω_1, P_1) y (Ω_2, P_2), el **espacio producto** (Ω, P) sería

$$\Omega = \Omega_1 \times \Omega_2 \qquad P = P(A \times B) = P_1(A)P_2(B)$$

Un **conjunto cilindro** de una sucesión de términos a_i de $\Omega^{(N)}$ es aquella en la que los primeros n términos deben cumplir una condición y el resto no esta sometido a ninguna limitación.

6.3. Independencia de varios sucesos

En un espacio de probabilidad, tres sucesos A_1, A_2, A_3 son independientes si se cumplen las condiciones

1. $P(A_1 \cap A_2) = P(A_1)P(A_2)$
2. $P(A_1 \cap A_3) = P(A_1)P(A_3)$
3. $P(A_2 \cap A_3) = P(A_2)P(A_3)$
4. $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$

Tres sucesos pueden ser independientes dos a dos, sin ser independientes.

En un espacio de probabilidad, los sucesos $\{A_i | i \in I\}$ se dicen independientes si

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

cualquiera que sea $k \in \mathbb{N}$ y cualesquiera que sea $i_1, i_2, \ldots, i_k \in I$.

6.4. La independencia condicional

Sea A, B y C tres sucesos en un espacio de probabilidad (Ω, P). Si se verifica

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

se dice que A y B son **independientes condicionalmente** a C.

Los sucesos A y B pueden ser condicionalmente independientes, tanto cuando ocurre C como cuando no ocurre, y no ser independientes.

Dos sucesos A y B pueden ser independientes, per no ser condicionalmente independientes a C ni a C^c.

7 Variables aleatorias

7.1. El concepto de variable aleatoria

En un espacio de probabilidad discreto (Ω, P), se denomina **variable aleatoria**, a cualquier función

$$X : \Omega \to \mathbb{R}$$

Si la variable aleatoria X estuviese definida en un espacio muestral no numerable no habría ninguna garantías que conjuntos del tipo $X^{-1}(a, b]$ fuesen sucesos.

7.2. Distribución de una variable aleatoria

La colección de probabilidades $= \{X \in B\}$, correspondientes a cada subconjunto B de $\mathbb{R}$, se denomina la **distribución de la variable aleatoria** X.

Una variable aleatoria X es **discreta** si toma a lo sumo un número numerable de valores

$$X(\Omega) = \{x_1, x_2, x_2, \ldots, x_n\}$$

La función que a cada x_k asigna el valores

$$p_k = P\{X = x_k\}$$

se denomina **función de probabilidad de** X.

Naturalmente debe ser

$$\sum_{k=1}^{\infty} p_k = 1$$

Una urna en la que se introducen tarjetas que llevan anotado uno de los números x_i. Si la proporción de tarjetas con el número x_i es p_i, el número obtenido al elegir una tarjeta al azar, es una variable aleatoria con la función de probabilidad requerida[1].

7.3. Variables aleatorias simultaneas

La igualdad entre variables aleatorias exige que ambas estén definidas en el mismo espacio de probabilidad; en cambio la igualdad de sus distribuciones puede ocurrir entre variables definidas en espacios distintos.

Dos variables aleatorias tendrán una **relación funcional** cuando una vez elegegida una de ellas, la segunda queda determinada sin ninguna aleatoriedad.

[1]El mecanismo tropezará con dificultades si alguna de las p_i es irracional, por la imposibilidad de que, con un número finito de tarjetas, la proporción de algún subconjunto de ellas sea irracional. Una alternativa es dividir $\Omega = [0, 1]$ en segmentos de longitud $p_1, p_2, \ldots, p_i, \ldots$ y escoger un punto con distribución uniforme en $[0, 1]$.

Si X_1, X_2 son variables aleatorias discretas definidas en el mismo espacio de probabilidad, la distribución conjunta de (X_1, X_2) es la función que asigna a cada $B \subset \mathbb{R}^2$ la probabilidad
$$P\{(X_1, X_2) \in B\}$$

En realidad, la distribución conjunto de (X_1, X_2) se caracteriza por la función de probabilidad conjunta que hace corresponder

$$P\{X_1 = x_1, X_2 = x_2\}$$

a cada par (x_1, x_2) de valores posibles de (X_1, X_2) (que siempre será un subconjunto de $X_1(\Omega) \times X_2(\Omega)$).

Las funciones de probabilidad $P\{X_1 = x_2\}$ y $P\{X_2 = x_2\}$ se denominan **funciones de probabilidad marginales** de X_2 y X_2 respectivamente. Claramente

$$P\{X_1 = x_1\} = \sum_{x_2 \in X_2(\Omega)} P\{X_1 = x_1, X_2 = x_2\}$$
$$P\{X_1 = x_1\} = \sum_{x_1 \in X_1(\Omega)} P\{X_1 = x_1, X_2 = x_2\}$$

La función de Probabilidad

$$P\{X_2 = x_2 | X_1 = x_1\} = \frac{P\{X_1 = x_1, X_2 = x_2\}}{P\{X_1 = x_1\}}$$

donde x_1 es fijo y x_2 variable, corresponde a la **distribución de X_2 condicionada por $X_1 = x_1$**. De igual forma $P\{X_1 = x_1 | X_2 = x_2\}$ (con x_2 fijo y x_1 variable) es la función de probabilidad de la distribución de X_1 condicionada por $X_2 = x_2$.

7.4. Variables aleatorias independientes

Dos variables aleatorias X_1 y X_2, definidas en el mismo espacio probabilidad, se denominas **independientes** si se verifica

$$P\{X_1 = x_1, X_2 = x_2\} = P\{X_1 = x_1\}P\{X_2 = x_2\}$$

cualquiera que sean x_1 y x_2 entre los valores posibles de las variables.

Las variables aleatorias discretas $X_1, X_2, \ldots, X_r$, definidas en el mismo espacio de probabilidad, se denominan **independientes** si

$$P\{X_1 = x_1, X_2 = x_2, \ldots, X_r = x_r\} = P\{X_1 = x_1\}P\{X_2 = x_2\} \cdots P\{X_r = x_r\}$$

cualquiera que sean $x_1, x_2, \ldots, x_r$ dentro de los valores posibles de $X_1, X_2, \ldots, X_r$ respectivamente.

Lema 7.1 *Si $X_1, X_2, \ldots, X_r$ son variables aleatorias discretas e independientes y $C_1, C_2, \ldots, C_r$ sus subconjuntos cualesquiera de sus conjuntos de valores posibles, se verifica*

$$P\{X_1 \in C_1, X_2 \in C_2, \ldots, X_r \in C_r\} = P\{X_1 \in C_1\}P\{X_2 \in C_2\} \cdots P\{X_r \in C_r\}$$

En consecuencia, los sucesos $\{X_1 \in C_1\}, \{X_2 \in C_2\}, \ldots, \{X_r \in C_r\}$ son independientes.

8 Esperanza matemática

Si X es una variable aleatoria definida en un espacio de probabilidad discreto (Ω, P), el **valor esperado**[1], **esperanza matemática** o **media** de X es

$$E[X] = \sum_{\omega \in \Omega} X(\omega) P(\omega)$$

en el supuesto que

$$\sum_{\omega | X(\omega) > 0} X(\omega) P(\omega) < \infty \text{ o } \sum_{\omega | X(\omega) < 0} -X(\omega) P(\omega) < \infty$$

Si las dos series anteriores divergen, se dice que $E[X]$ no existe mientras que, si la primera diverge y la segunda converge, se toma $E[X] = +\infty$ y, si es al revés, $E[X] = -\infty$.

Si el conjunto de valores posibles de una variable aleatoria X es

$$X(\Omega) = \{x_1, x_2, \ldots, x_n\}$$

su **valor esperado**, si existe, se expresa

$$E[X] = \sum_k x_k P\{X = x_k\} = \sum_k x_k p_k$$

donde p_k es la función de probabilidad de X.

8.1. Propiedades de la esperanza matemática

1. Cualquier constante $c \in \mathbb{R}$ puede considerarse una variable aleatoria, basta definir $X(\omega) = c$ para todo $\omega \in \Omega$. Su distribución se denomina **distribución causal en** c.

$$E[c] = c$$

2. **Linealidad**: Sean X_1 y X_2 variables aleatorias discretas en el mismo espacio de probabilidad, cuyas medias son finitas. Si $c_1, c_2 \in \mathbb{R}$

$$E[c_1 X_1 + c_2 X_2] = c_1 E[X_1] + c_2 E[X_2]$$

3. Sea X una variable aleatoria no negativa, es decir $X(\omega) \geq 0$ para todo $\omega \in \Omega$. Entonces

$$E[X] \geq 0$$
$$X \geq Y \Rightarrow E[X] \geq E[Y]$$

[1] Antiguamente, se conocía como **esperanza moral**

4. Si I_A es la función indicatriz de un suceso A, se tiene

$$E[I_A] = P(A)$$

5. si X es una variable aleatoria que toma solo valores enteros no negativos, se cumple

$$E[X] = \sum_{m=1}^{\infty} P\{X \geq m\}$$

o lo que es lo mismo

$$E[X] = \sum_{m=0}^{\infty} P\{X > m\}$$

Por consiguiente, para una variable aleatoria X con valores enteros

$$E[|X|] = \sum_{m=1}^{\infty} P\{|X| \geq m\}$$

Las probabilidades $P\{|X| \geq m\}$ se denominan **colas de la distribución**, porque dan la probabilidad de que X esté fuera de la región central $(-m, m)$. Si $X = f(X)$ es una variable aleatoria función no lineal de otra, suele ser

$$E[f(X)] \neq f(E[X])$$

6. Si X, Y son variables aleatorias independientes, con esperanza finita, se verifica

$$E[XY] = E[X]E[Y]$$

8.2. Esperanza condicionada y métodos recurrentes

El que haya ocurrido un suceso B, de probabilidad $P(B) > 0$, da lugar la **esperanza matemática condicionada por** B

$$E[X|B] = \sum_{\omega\in\Omega} X(\omega)P\{\omega|B\} = \sum_{k} x_k P\{X = x_k|B\}$$

9 Análisis descriptivo de las distribuciones de probabilidad

9.1. Momentos de una distribución

Respecto al origen

El **momento de orden r respecto del origen** de una variable aleatoria X, o de su distribución de probabilidad, es la esperanza matemática de X^r:

$$\alpha_r = E[X^r] = \sum_{\omega \in \Omega} X^r(\omega) P\{\Omega\} = \sum_k x_k^r p_k$$

que, cuando existe, se suele designar por α_r.

Salvo cuando X es positiva, no es conveniente considerar más que momentos de orden entero.

Cuando el momento de orden r es finito, también son finitos los momentos de orden inferior a r.

Respecto a la media o momento central

El **momento de orden $r > 1$ respecto a la media o momento central de orden** r de una variable aleatoria X, o de su distribución, es la esperanza matemática de $(X - E[X])^r$. Por tanto, si se le designa por μ_r, es

$$\mu_r = E[(X - E[X])^r] = \sum_k (x_k - \alpha_1)^r p_k$$

Los momentos centrales respecto al origen se relacionan

$$\mu_r = \sum_{i=0}^{r} (-1)^i \binom{r}{i} \alpha_1^i \alpha_{r-1}$$

siendo $\alpha_0 = 1$.

Por ejemplo,

$$\begin{aligned}
\mu_2 &= \alpha_2 - \alpha_1^2 \\
\mu_3 &= \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 \\
\mu_4 &= \alpha_4 - 4\alpha_1\alpha_3 + 6\alpha_1^2\alpha_2 - 3\alpha_1^4
\end{aligned}$$

El momento central de segundo orden, μ_2 es el equivalente al momento de inercia y se le conoce como **varianza de la distribución**, y se le designa como $V(X)$ o $\sigma^2(X)$.

$$V(X) = \sigma^2 = E[(X - E[X])^2] = \sum_k (x_k - \alpha_1)^2 p_k = E[X^2] - E[X]^2$$

Como $\sigma^2 \geq 0$ siempre, se tiene que

$$E[X^2] \geq E[X]^2$$

La **desviación típica de la distribución**, σ, es la raíz cuadrada de la varianza (σ^2) y se interpreta como la dispersión de la distribución alrededor de la media.

A veces, se utiliza **coeficiente de varianza de la variable** X

$$\frac{\sigma(X)}{E[X]}$$

Si la dispersión se calcula a partir de un punto genérico a, mediante $E[(X-a)^2]$. Esta se minimiza si calcula alrededor de la media, $a = E[X]$.

Proposición 9.1 (Desigualdad de Tchebychev) *Cualquiera que sea la distribución de probabilidad de una variable aleatoria X, se verifica*

$$P\{|X - E[X]| > k\sigma\} \leq \frac{1}{k^2}$$

para cualquier $k > 0$, o lo que es lo mismo

$$P\{|X - E[X]| > c\} \leq \frac{\sigma^2}{c^2}$$

para cualquier $c > 0$.

Proposición 9.2 (Desigualdad de Markov)

$$P\{f(X) > c\} \leq \frac{E[f(X)]}{c}$$

El momento central de tercer orden

$$\mu_3 = \sum_k (x_k - \alpha_1)^3 p_k$$

mide la asimetría de la distribución alrededor de su media. También se usa el **coeficiente de asimetría de la distribución**.

$$\gamma_3 = \frac{\mu_3}{\sigma_3}$$

El **coeficiente de apuntalamiento** o **curtosis de la distribución**

$$\gamma_4 = \frac{\mu_4}{\sigma_4} - 3$$

9.2. Momentos de una distribución conjunta

Respecto al origen, los momentos de la distribución conjunta con el valor esperado del producto de dos potencias de las variables.

$$\sigma_{r,s} = E[X_1^r X_2^s]$$

Respecto al centro de gravedad $(E[X_1], E[X_2])$ de la distribución, debe restarse a cada variable su media.

$$\mu_{r,s} = E[(X_1 - E[X_1])^r (X_2 - E[X_2])^s]$$

El **orden** de uno de estos momentos es el número $r + s$.

El momento más utilizado es el valor esperado del producto de las dos variables o **covarianza** entre ellas

$$\alpha_{1,1} = E[X_1 X_2]$$

y, más aún, la **covarianza** entre ellas

$$\mathrm{Cov}(X_1, X_2) = \mu_{1,1} = E[(X_1 - E[X_1])(X_2 - E[X_2])] = E[X_1 X_2] - E[X_1]E[X_2]$$

Dos variables aleatorias X_1 y X_2 independientes, tienen covarianza nula. Pero dos variables con covarianza nula, no son independientes.

Otra muy usada es el **coeficiente de correlación**

$$\rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma(X_1)\sigma(X_2)}$$

Las variables cuyo coeficiente de correlación es nulo se denominan **incorreladas**.

La **matriz de covarianzas** de X_1 y X_2 es la matriz

$$\Sigma(X_1, X_2) = \begin{pmatrix} \sigma^2(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_1, X_2) & \sigma^2(X_2) \end{pmatrix}$$

Es una matriz simétrica semidefinida positiva y con determinante positivo, por lo que

$$\mathrm{Cov}(X_1, X_2)^2 \leq \sigma^2(X_1)\sigma^2(X_2) \qquad -1 \leq \rho(X_1, X_2) \leq 1$$

Si $\rho(X_1, X_2) = \pm 1$ es equivalente a que cada una de las variables sea función lineal de la otra.

La mejor previsión de X_2, mediante una función lineal de X_1, es la **recta de regresión de X_2 sobre X_1**

$$x_2 = a^*(x_1 - E[X_1]) + E[X_2] \qquad a^* = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma^2(X_1)}$$

Siendo la **varianza residual de** X_2, después de hacer la regresión sobre X_1.

$$\sigma^2(X_2)(1 - \rho^2(X_1, X_2))$$

Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes, se verifica

$$\sigma^2(X_1 + X_2 + \cdots + X_n) = \sigma^2(X_1) + \sigma^2(X_2) + \cdots + \sigma^2(X_n)$$

9.3. Otros indicadores de posición y dispersión

Indicadores de posición

Moda La **moda** de una distribución que asigna probabilidades $p_1, p_2, \ldots$ a los puntos $x_1, x_2, \ldots$ es el valor x_k para el cual p_k es máximo.

Mediana La **mediana** de una distribución es aquel valor M que deja probabilidad $\frac{1}{2}$ tanto por debajo como por encima de él.

$$P\{X \leq M\} \geq \frac{1}{2} \qquad P\{X \geq M\} \geq \frac{1}{2}$$

Al igual que la media es el valor a que minimiza $E[(X-a)^2]$, la mediana es el valor que minimiza.

$$E[|X - a|]$$

Indicadores de dispersión

Promedio de las desviaciones absolutas a a Consiste en calcular $E[|X - a|]$ tomando a como $E[X]$ o $M(X)$.

Desviación probable Si $M(X)$ es única se puede calcular la **mediana de las desviaciones absolutas a la mediana**

$$M(|X - M(X|)$$

Tiene la virtud que si su valor es D, entonces el intervalo $[M(X)-D, M(X)+D]$ tiene probabilidad superior a $\frac{1}{2}$. Y si el intervalo es $(M(X) - D, M(X) + D)$ tiene probabilidad inferior a $\frac{1}{2}$.

Cuantil de orden p Dado cualquier número $p \in [0, 1]$, el menor valor x_k de la variable tal que $P\{X \leq x_k\} \geq p$ se denomina el **cuantil de orden** p de la distribución.

$$c_p = \text{mín}\{x_k | F(x_k) \geq p\}$$

$c_{\frac{1}{2}}$ es la mediana. $c_{\frac{1}{4}}$ y $c_{\frac{3}{4}}$ el primer y tercer cuartil. Y el **intervalo intercuartílico** $[c_{\frac{1}{4}}, c_{\frac{3}{4}}]$ tiene probabilidad igual o superior a $\frac{1}{2}$

9.4. Función generatriz

Si X es una variable aleatoria con valores enteros no negativos, con $P\{X = n\} = p_n$ para cada n, se llama función generatriz de X a la función

$$G(z) = E[z^X] = \sum_{n=0}^{\infty} z^n p_n$$

Si X_1 y X_2 son dos variables aleatorias independientes, con función generatrices $G_1(z)$ y $G_2(z)$, la función generatriz de la suma $X_1 + X_2$ es el producto $G_1(z)G_2(z)$.

10 Pruebas repetidas

Un **esquema de pruebas repetidas e independientes** es un experimento aleatorio, se repite una vez tras otra, sin que el resultado de cada realización pueda incluir en el resultado de los demás.

Una **sucesión de variables aleatorias de Bernouilli** es una sucesión de variables aleatorias independientes I_i, cada una de las cuales vale 1 si se presenta el suceso A en la prueba i-ésima y 0 en caso contrario.

10.1. La distribución binomial y la aproximación de Poisson

Si en cada realización de un experimento aleatorio el suceso A tiene probabilidad p de ocurrir, el repetir n veces el experimento, el número de veces que se presenta el suceso A es una variable aleatoria, X_n con función de probabilidad.

$$P\{X_n = k\} = \binom{n}{k} p^k q^{n-k} \qquad \text{para } k = 0, 1, 2, \ldots, n$$

donde $q = 1 - p$. Se trata de una **distribución binomial** $B(n, p)$.

Sus momentos más importantes son:

$$E[X_n] = np \qquad \sigma^2(X_n) = np(1-p)$$

Teorema 10.1 (Teorema de Poisson) *Si n tiende a infinito y p tienda a cero, de tal manera que el producto np converge a una constante λ, se verifica*

$$\text{lím} \binom{n}{k} p^k q^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$$

para cada valor fijo de k.

10.2. La aproximación normal a la distribución binomial

Teorema 10.2 (Teorema de Moivre-Laplace) *Cuando n tiende a infinito, si k tiende hacia infinito de manera que*

$$x_k = \frac{k - np}{\sqrt{npq}}$$

se mantiene acotado en valor absoluto por alguna constante A, se cumple

$$\binom{n}{k} p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x_k^2}{2}}$$

en el sentido de que el cociente entre ambos miembros converge a 1. Más exactamente, se tiene

$$\binom{n}{k} p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x_k^2}{2}} (1+\rho_n)$$

donde $\frac{\sqrt{n}}{|\rho_n|}$ *esta acotado, a partir de un* n *en adelante, por una constante que solo depende de* A *(y de* p*).*

11 Apéndice I. Combinatoria

11.1. Principios generales

Contar es hallar el **cardinal** de un conjunto. La **combinatoria** es el arte de contar conjuntos sin hacer enumeraciones.

Si en una competición se inscriben n jugadores se juegan $n-1$ partidos.

El **procedimiento constructivo** consiste en recorrer mentalmente los pasos a seguir para formar todos los elementos del conjunto, anotando las alternativas que pueden elegirse en cada uno.

Si los conjuntos $A_1, A_2, \ldots, A_n$ tienen $n_1, n_2, \ldots, n_n$ elementos respectivamente, el producto cartesiano $A_1 \times A_2 \times \cdots \times A_n$ tiene $n_1 n_2 \cdots n_n$ elementos.

No tiene ninguna relevancia donde se producen las restricciones, solo cuales son estas.

11.2. Patrones más usuales

Ordenaciones

n objetos distintos pueden ordenarse en fila de $n!$ maneras distintas.

Una colección de n objetos, clasificados en k grupos de objetos idénticos entre sí, el primero con n_1 objetos, el segundo con n_2,... se pueden ordenar en fila de

$$\frac{n!}{n_1! n_2! \cdots n_k!}$$

maneras distintas, si no se consideran distintas las ordenaciones en las cuales dos objetos iguales han permutado su posición.

Subconjuntos ordenados

Hay

$$n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!} = (n)_r$$

subconjuntos ordenados posibles de r elementos, que pueden extraerse de un conjunto de n elementos. También conocidos como **variaciones sin repetición de** r **elementos tomados entre** n.

El número de **variaciones con repetición** de r elementos tomados entre n es n^r.

Subconjuntos

El número de subconjuntos distintos, con r elementos, que pueden extraerse de un conjunto de n elementos es

$$\binom{n}{r} = \frac{n!}{r!(n-r!)}$$

o **combinaciones** de r elementos tomados entre n.

La diferencia entre variaciones y subconjuntos es si el orden de aparición de los elementos es importante.

Repartos

Si hay que repartir r objetos iguales en n grupos, existen

$$\binom{n+r-1}{r}$$

repartos posibles.

11.3. Identidades combinatorias

Una **identidad combinatoria** es una igualdad $f(n) = g(n)$, válida para cada n natural, en la cual $f(n)$ y $g(n)$ son cantidades relacionadas con el cardinal de algún conjunto.

$$\binom{n}{k} = \binom{n}{n-k}$$

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

$$\sum_{i=0}^{n} \binom{n}{i}^2 = \binom{2n}{n}$$

$$\sum_{k=0}^{m-1} \binom{n}{k}\binom{n-1}{m-k-1} = \binom{n+m-1}{n}$$

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

$$\binom{n}{k} = \sum_{j=0}^{k} \binom{n-j-1}{k-j} = \sum_{j=1}^{n-k+1} \binom{n-j}{k-1}$$

Si $x \in \mathbb{R}$ y $r > 0$

$$\binom{x}{r} = \frac{x(x-1)(x-2)\cdots(x-r+1)}{r!}$$

y si $r < 0$

$$\binom{x}{0} = 1 \qquad \binom{x}{r} = 0$$

Identidad inmediata

$$\binom{-x}{r} = (-1)^r \binom{x+r-1}{r}$$

Desarrollo de Taylor Si $|t| < 1$ y para cualquier $x \in \mathbb{R}$

$$(1+t)^x = 1 + \binom{x}{1} t + \binom{x}{2} t^2 + \cdots + \binom{x}{r} t^r + \cdots$$

y si $|b| < |a|$ y cualquier $x \in \mathbb{R}$

$$(a+b)^x = a^x + \binom{x}{1} b a^{x-1} + \binom{x}{2} b^2 a^{x-2} + \cdots + \binom{x}{r} b^r a^{a-r} + \cdots$$

$$(1-t)^n \sum_{r=0}^{\infty} \binom{n+r-1}{r} t^r = 1$$

$$(-1)^n \binom{-\frac{1}{2}}{n} = \binom{2n}{n} \frac{1}{2^{2n}}$$

$$\sum_{n=0}^{\infty} \frac{2n}{n} \frac{t^n}{2^{2n}} = (1-t)^{\frac{-1}{2}}$$

y si n es un entero positivo

$$\binom{x+y}{n} = \binom{x}{0}\binom{y}{n} + \binom{x}{1}\binom{y}{n-1} + \binom{x}{2}\binom{y}{n-2} + \cdots + \binom{x}{n}\binom{y}{0}$$

11.4. Formula de Stirling

$$n! = \sqrt{2n} n^{n+\frac{1}{2}} e^{-n}$$

con más precisión

$$1 < \frac{n!}{\sqrt{2n} n^{n+\frac{1}{2}} e^{-n}} < e^{\frac{1}{8n}}$$

Desarrollo de series

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \qquad x \in (-\infty, \infty) \tag{11.1}$$

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}, \qquad x \in (-\infty, \infty) \tag{11.2}$$

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}, \qquad x \in (-\infty, \infty) \tag{11.3}$$

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n, \qquad x \in (-1, 1) \tag{11.4}$$

$$\frac{1}{1+x} = \sum_{n=0}^{\infty} (-1)^n x^n, \qquad x \in (-1, 1) \tag{11.5}$$

$$\log(1-x) = \sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1}, \qquad x \in (-1, 1) \tag{11.6}$$

$$\log(1+x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{n+1}}{n+1}, \qquad x \in (-1, 1) \tag{11.7}$$

$$\log(\frac{1-x}{1+x}) = 2\sum_{n=0}^{\infty} \frac{x^{2n+1}}{2n+1}, \qquad x \in (-1, 1) \tag{11.8}$$

$$\frac{1}{1+x*2} = \sum_{n=0}^{\infty} (-1)^n x^{2n}, \qquad x \in (-1, 1) \tag{11.9}$$

$$\arctan(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}, \qquad x \in (-1, 1) \tag{11.10}$$

$$(1+x)^\alpha = 1 + \sum_{n=1}^{\infty} \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!} x^n, \qquad x \in (-1, 1) \tag{11.11}$$

$$\frac{1}{\sqrt{1+x}} = 1 + \sum_{n=1}^{\infty} (-1)^n \frac{(2n-1)!!}{(2n)!!} x^n, \qquad x \in (-1, 1) \tag{11.12}$$

$$\frac{1}{\sqrt{1-x^2}} = 1 + \sum_{n=1}^{\infty} \frac{(2n-1)!!}{(2n)!!} x^{2n}, \qquad x \in (-1, 1) \tag{11.13}$$

$$\frac{1}{\sqrt{1+x^2}} = 1 + \sum_{n=1}^{\infty} (-1)^n \frac{(2n-1)!!}{(2n)!!} x^{2n}, \qquad x \in (-1, 1) \tag{11.14}$$

$$\arcsin x = x + \sum_{n=1}^{\infty} \frac{(2n-1)!!}{(2n)!!} \frac{x^{2n+1}}{2n+1}, \qquad x \in (-1, 1) \tag{11.15}$$

$$argshx = x + \sum_{n=1}^{\infty} (-1)^n \frac{(2n-1)!!}{(2n)!!} \frac{x^{2n+1}}{2n+1}, \qquad x \in (-1, 1) \tag{11.16}$$