02MAIR - Matemáticas para la IA
Probabilidad y estadística

David Zorío Ventura

Universidad Internacional de Valencia
Máster en Inteligencia Artificial

- La probabilidad es una medida que asigna un valor de certidumbre a la ocurrencia de un determinado suceso.
- Habitualmente toma valores comprendidos entre 0 y 1, donde valores próximos a 0 indica poca probabilidad y los próximos a 1 bastante probabilidad.
- Por ejemplo, si lanzamos una moneda equilibrada, es razonable asignar $P(\text{cara}) = 0.5$ y $P(\text{cruz}) = 0.5$.
- Otro ejemplo: si tenemos una urna con 30 bolas blancas y 70 negras y extraemos una al azar, también es razonable asignar $P(\text{blanca}) = 0.3$ y $P(\text{negra}) = 0.7$.
- La probabilidad pretende asignar la proporción a la que tienden a darse los resultados tras muchos experimentos:
  - En el ejemplo de las monedas: 50 % caras y 50 % cruces.
  - En el ejemplo de la urna: 30 % blancas y 70 % negras.

Definición

El conjunto formado por todos los resultados posibles, Ω, recibe el nombre de **espacio muestral***. Cada subconjunto formado por esas posibles ocurrencias, $A \subseteq \Omega$, reciben el nombre de* **sucesos** *o* **eventos***.*

Ejemplos

1. *El espacio muestral asociado al lanzamiento de una moneda es* $\Omega = \{cara, cruz\}$.
2. *Extracción en una urna con bolas blancas y negras:* $\Omega = \{blanca, negra\}$.
3. *Lanzamiento de un dado:* $\Omega = \{1, 2, 3, 4, 5, 6\}$.
4. *Lanzamiento de dos monedas:* $\Omega = \{(cara, cara), (cara, cruz), (cruz, cara), (cruz, cruz)\}$.

Ejemplos

1. *Lanzamiento de una moneda equilibrada:* $\Omega = \{cara, cruz\}$. *Posibles sucesos:*
   - *No ocurre nada:* $A = \emptyset$. $P(A) = 0$.
   - *Sale cara:* $A = \{cara\}$. $P(A) = 0.5$.
   - *Sale cruz:* $A = \{cruz\}$. $P(A) = 0.5$.
   - *Sale alguna de las dos caras:* $A = \{cara, cruz\}$. $P(A) = 1$.
2. *Lanzamiento de un dado:* $\Omega = \{1, 2, 3, 4, 5, 6\}$. *Ejemplos de sucesos:*
   - *Sale un número par:* $A = \{2, 4, 6\}$. $P(A) = 0.5$.
   - *Sale un número impar:* $A = \{1, 3, 5\}$. $P(A) = 0.5$.
   - *Sale un número menor que* 5*:* $A = \{1, 2, 3, 4\}$. $P(A) = \frac{2}{3}$.
   - *Sale un número que empieza por "c":* $A = \{4, 5\}$. $P(A) = \frac{1}{3}$.

Ejemplos

3. *Lanzamiento de dos monedas:*
   $\Omega = \{(cara, cara), (cara, cruz), (cruz, cara), (cruz, cruz)\}$.
   Ejemplos de sucesos:
   - *Obtener el mismo resultado:* $A = \{(cara, cara), (cruz, cruz)\}$. $P(A) = 0.5$.
   - *No obtener dos caras:* $A = \{(cara, cruz), (cruz, cara), (cruz, cruz)\}$. $P(A) = 0.75$.
4. *Lanzamiento de dos dados:* $\Omega = \{(i, j) \mid 1 \leq i, j \leq 6\}$. *Ejemplos de sucesos:*
   - *La suma de los resultados es* 10*:* $A = \{(4, 6), (5, 5), (6, 4)\}$. $P(A) = \frac{1}{12}$.
   - *Ambos resultados son pares:* $A = \{(2, 2), (2, 4), (2, 6), (4, 2), (4, 4), (4, 6), (6, 2), (6, 4), (6, 6)\}$. $P(A) = \frac{1}{4}$.

Los sucesos pueden combinarse entre sí mediante las operaciones básicas entre conjuntos (unión, intersección, complementario, etc.).

Ejemplos

1. *Lanzamos un dado y consideramos el siguiente suceso: "sacar un número par o que sea mayor que 3". Llamamos A al suceso "sacar un número par" (por lo que $A = \{2, 4, 6\}$) y B al suceso "sacar un número mayor que 3" (luego $B = \{4, 5, 6\}$). Por tanto, el suceso que buscamos es*

$$A \cup B = \{2, 4, 6\} \cup \{4, 5, 6\} = \{2, 4, 5, 6\}.$$

2. *Bajo las condiciones anteriores, consideramos el suceso "sacar un número par mayor que 3". En este caso la operación que debemos realizar es*

$$A \cap B = \{2, 4, 6\} \cap \{4, 5, 6\} = \{4, 6\}.$$

Definición

Dos sucesos $A, B \subseteq \Omega$ son **mutuamente excluyentes** *si $A \cap B = \emptyset$.*

Ejemplo

Al lanzar un dado, $A = \{1, 3\}$, $B = \{2, 4, 6\}$. Se cumple $A \cap B = \emptyset$.

Definición

Una probabilidad P definida sobre el conjunto de sucesos asociado a un espacio muestral Ω es una función real cumpliendo:

1. $0 \leq P(A) \leq 1$ *para cualquier suceso $A \subseteq \Omega$.*
2. $P(\Omega) = 1$.
3. *Para cada par de sucesos $A, B \subseteq \Omega$* **mutuamente excluyentes** *(es decir, $A \cap B = \emptyset$), se tiene $P(A \cup B) = P(A) + P(B)$.*

De los axiomas asociados a una probabilidad se pueden deducir las siguientes propiedades:

Propiedades

Sea Ω un espacio muestral, P una probabilidad y $A, B \subseteq \Omega$ sucesos. Entonces se cumple:

1. $P(A^c) = 1 - P(A)$.
2. $P(\emptyset) = 0$.
3. *Si* $A \subseteq B$ *, entonces* $P(A) \leq P(B)$.
4. *Si* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
5. *En particular,* $P(A \cup B) \leq P(A) + P(B)$.

Ejercicio

Demuéstrense las propiedades anteriores utilizando los tres axiomas de la probabilidad.

Ejemplo

En un grupo de amigos, un 40 % de ellos estudia Ingeniería Informática, un 30 % Matemáticas y un 10 % ambas cosas a la vez. ¿Cuál es la probabilidad de que una persona elegida al azar estudie alguna de las dos carreras? Si $A = \{\textit{estudiar Ingeniería Informática}\}$ *y* $B = \{\textit{estudiar Matemáticas}\}$*, se nos pide*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.3 - 0.1 = 0.6.$$

La ocurrencia de determinados sucesos (o el conocimiento de determinada información) puede *condicionar* la ocurrencia de otros sucesos.

Ejemplo

Lanzamos un dado equilibrado y consideramos el suceso $A = \{obtener\ un\ 2\}$. *En ese caso, y sin saber nada más, sabemos que la probabilidad de que ocurra ese suceso es* $\frac{1}{6}$. *Sin embargo, esto puede cambiar si conocemos de antemano que ha ocurrido otro suceso* B.

- *Por ejemplo, si* $B = \{se\ ha\ obtenido\ un\ número\ par\}$, *entonces la probabilidad de* A *condicionada a* B *es* $\frac{1}{3}$.
- *Por otra parte, si* $B = \{se\ ha\ obtenido\ un\ número\ impar\}$, *entonces dicha probabilidad es* 0.

Definición

Sean A, B sucesos de un espacio muestral Ω y P una probabilidad asociada a éste. Definimos la probabilidad de A **condicionada** *a B como*

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

Ejemplo

En el ejemplo anterior del dado, y tomando $A = \{obtener\ un\ 2\}$:

- *Si $B = \{obtener\ un\ número\ par\}$, entonces*
  $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$
- *Si $B = \{obtener\ un\ número\ impar\}$, entonces*
  $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = \frac{0}{\frac{1}{2}} = 0.$$

Teorema (Teorema de Bayes)

Sean A, B sucesos de un espacio muestral Ω y P una probabilidad asociada. Entonces

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

Ejemplo

Consideramos el lanzamiento de un dado y los sucesos $A = \{1, 2, 4, 5\}$ y $B = \{2, 4, 6\}$. Entonces

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

Por otra parte, utilizando el teorema de Bayes:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{\frac{2}{3} \cdot \frac{1}{2}}{\frac{2}{3}} = \frac{1}{2}.$$

Definición

Diremos que dos sucesos $A, B \subseteq \Omega$ son independientes si $P(A|B) = P(A)$.

En general, de la definición de probabilidad condicionada, se deduce que para cualquier par de sucesos $A, B \subseteq \Omega$ (independientes o no) se cumple

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

Por tanto:

Teorema

Dos sucesos $A, B \subseteq \Omega$ son independientes si y sólo si se cumple $P(A \cap B) = P(A)P(B)$.

Ejemplo

Lanzamos una moneda y un dado, ambos equilibrados. ¿Cuál es la probabilidad de obtener una cara en la moneda y, simultáneamente, un 3 en el dado?

Si llamamos A al suceso "obtener una cara al lanzar la moneda" y B al suceso "obtener un 3 al lanzar el dado", entonces se nos pide $P(A \cap B)$.

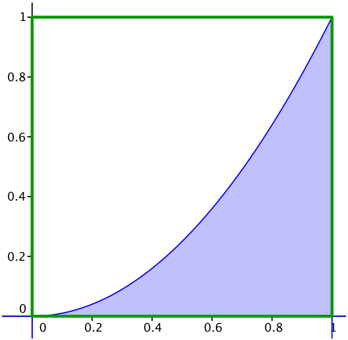
Dado que los sucesos A y B son independientes, se tiene:

$$P(A \cap B) = P(A)P(B) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}.$$

En ocasiones nos podemos encontrar con espacios muestrales infinitos.

Ejemplo

Sea $\Omega = [0,1] \times [0,1]$ *y la función* $f : [0,1] \to [0,1]$ *dada por* $f(x) = x^2$*. ¿Cuál es la probabilidad de que un punto elegido al azar en el cuadrado* $[0,1] \times [0,1]$ *esté por debajo de la gráfica de* f*?*



El área encerrada por la región es:

$$\int_0^1 x^2 \mathrm{d}x = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}.$$

Por tanto, la probabilidad es:

$$\frac{\text{Área región}}{\text{Área cuadrado}} = \frac{\frac{1}{3}}{1} = \frac{1}{3}.$$

Definición

Una **variable aleatoria** *(V.A.) X sobre un espacio muestral Ω es una función $\Omega \to \mathbb{R}$. En ese caso, es posible inducir una probabilidad asociada a esa variable aleatoria sobre $X(\Omega) \subseteq \mathbb{R}$. Dado un suceso $A \subseteq \mathbb{R}$, definimos*

$$P_X(A) = P\big(X^{-1}(A)\big) = P\big(\{\omega \in \Omega \mid X(\omega) \in A\}\big).$$

Ejemplo

Sea Ω el conjunto formado por la población mundial $X : \Omega \to \mathbb{R}$ la variable aleatoria que a una persona $\omega \in \Omega$ le asigna su año de nacimiento, $X(\omega) \in \mathbb{R}$. Entonces si $A = \{1984, 1985, 1986\}$:

$$X^{-1}(A) = \{\text{personas nacidas entre 1984 y 1986}\}$$
$$P_X(A) = \text{proporción de personas nacidas entre 1984 y 1986.}$$

Definición

Sea X una V.A. Si el conjunto $X(\Omega)$ es finito o numerable (indexable por números naturales), entonces diremos que X es una **variable aleatoria discreta***.*

En ese caso, definimos la **función de densidad** *o función de masa de probabilidad asociada como*

$$f_X(x) = P(X = x) = P_X(\{x\}) = P(\{\omega \in \Omega \mid X(\omega) = x\}).$$

Por otra parte, definimos la **función de distribución** *asociada como*

$$F_X(x) = P(X \leq x) = \sum_{t \in \Omega, X(t) \leq x} f_X(t).$$

Condición de normalización: $\sum_{t \in \Omega} f_X(t) = 1.$

Ejemplos

1. *Sea* $\Omega = \{uno, dos, tres, cuatro, cinco, seis\}$ *el conjunto de los posibles resultados al lanzar un dado y* $X : \Omega \to \mathbb{R}$ *la V.A. que le asigna su valor:* $X(uno) = 1$, $X(dos) = 2$, $X(tres) = 3$, $X(cuatro) = 4$, $X(cinco) = 5$ *y* $X(seis) = 6$. *Entonces*

$$X(\Omega) = \{1, 2, 3, 4, 5, 6\}.$$

$$f_X(4) = P(X = 4) = P(\{cuatro\}) = \frac{1}{6}.$$

Por otra parte,

$$F_X(4) = P(X \leq 4) = P(\{uno, dos, tres, cuatro\}) = \frac{2}{3}.$$

Ejemplos

2. *Consideramos el experimento aleatorio consistente en lanzar una moneda hasta obtener como resultado cara. El espacio muestral es por tanto*

   $$\Omega = \{cara, (cruz, cara), (cruz, cruz, cara), (cruz, cruz, cruz, cara), ...\}.$$

   Sea X la V.A. aleatoria consistente en contar el número de lanzamientos realizados antes de obtener cara. Entonces:

   $$\begin{aligned} X(cara) &= 0 \\ X\left((cruz, cara)\right) &= 1 \\ X\left((cruz, cruz, cara)\right) &= 2 \\ X\left((cruz, cruz, cruz, cara)\right) &= 3 \end{aligned}$$

   Por tanto, $X(\Omega) = \{0, 1, 2, 3, \ldots\}$.

Definición

Sea X una V.A. asociada a un espacio muestral Ω. Diremos que X es una V.A. **continua** *si $X(\Omega)$ es continuo (por ejemplo, un intervalo o todos los reales).*

Definimos la **función de distribución** *asociada a X como*

$$F_X(x) = P(X \leq x).$$

Asimismo, diremos que X admite una **función de densidad** *si $\exists f_X$ no negativa e integrable tal que para cada $A \subseteq \mathbb{R}$*

$$P(X \in A) = \int_A f_X(t)\mathrm{d}t.$$

- Si X es una V.A. continua que admite una función de densidad, entonces podemos escribir

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t.$$

- La función de densidad debe satisfacer la condición de normalización:

$$\int_{-\infty}^{+\infty} f_X(t)\mathrm{d}t = P(X \in \mathbb{R}) = 1.$$

- De la existencia de una función de densidad se deduce que

$$P(X < x) = \lim_{t \to x^-} F_X(t) = F_X(x) = P(X \leq x).$$

- En particular, se deduce que en ese caso

$$P(X = x) = 0, \, \forall x \in \mathbb{R}.$$

Ejemplo

Se elige un número al azar dentro del intervalo real $[0, 1]$ *y se considera la V.A. aleatoria asociada a dicho experimento. Entonces la función de distribución asociada a* X *es:*

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & si\ x < 0, \\ x & si\ 0 \leq x \leq 1, \\ 1 & si\ x > 1. \end{cases}$$

Por tanto, la función de densidad asociada a X *es:*

$$f_X(x) = \begin{cases} 0 & si\ x < 0, \\ 1 & si\ 0 \leq x \leq 1, \\ 0 & si\ x > 1. \end{cases}$$

Definición

Sea X una V.A. asociada a un espacio muestral Ω, y con función de densidad $f_X(x)$. Definimos la **esperanza** *de X como*

$$E[X] = \begin{cases} \sum_{x \in X(\Omega)} x f_X(x) & \text{si } X \text{ es una V.A. discreta,} \\ \int_{\mathbb{R}} x f_X(x) \mathrm{d}x & \text{si } X \text{ es una V.A. continua.} \end{cases}$$

Propiedades

Sean X, Y variables aleatorias, $a \in \mathbb{R}$. Entonces:

1. $E[X+Y] = E[X] + E[Y]$.
2. $E[aX] = aE[X]$.
3. *Si X, Y son independientes, entonces $E[XY] = E[X]E[Y]$.*

Ejemplos

1. *El precio del número de una rifa es de 1 euro. Si ésta tiene un total de 1000 números y el premio tiene un valor de 500 euros, ¿cuál es el promedio de beneficio esperado?*

   Llamamos X a la V.A. que cuantifica el beneficio obtenido al participar en la rifa:

   $$\textit{Perder la rifa: } X = -1 \to P(X = -1) = \frac{999}{1000}.$$
   $$\textit{Ganar la rifa: } X = 499 \to P(X = 499) = \frac{1}{1000}.$$

   Por tanto:

$$E[X] = (-1)\cdot P(X = -1) + 499\cdot P(X = 499) = -\frac{999}{1000} + \frac{499}{1000} = -0.5.$$

Ejemplos

2. *Consideramos el intervalo* $[0, 1]$ *y seleccionamos un punto al azar y nos planteamos cuál es el valor sobre el que oscilará el promedio. Vimos que*

$$f_X(x) = \begin{cases} 0 & si\ x < 0, \\ 1 & si\ 0 \leq x \leq 1, \\ 0 & si\ x > 1. \end{cases}$$

Por tanto:

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) \mathrm{d}x = \int_0^1 x \mathrm{d}x = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}.$$

Definición

Sea X una V.A. y $g : \mathbb{R} \to \mathbb{R}$. Definimos

$$E[g(X)] = \begin{cases} \sum_{x \in X(\Omega)} g(x) f_X(x) & \text{si } X \text{ es una V.A. discreta,} \\ \int_{\mathbb{R}} g(x) f_X(x) \mathrm{d}x & \text{si } X \text{ es una V.A. continua.} \end{cases}$$

Definición

Sea X una V.A. aleatoria. Definimos la **varianza** *de X como:*

$$\mathrm{Var}(X) = E\left[(X - E[X])^2\right] = E\left[X^2\right] - E[X]^2.$$

Propiedades

1. $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$, $\forall a \in \mathbb{R}$.
2. $\mathrm{Var}(X + a) = \mathrm{Var}(X)$, $\forall a \in \mathbb{R}$.
3. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ *si X, Y son independientes.*

Ejemplos

1. *En el ejemplo de la rifa, se tiene*

$$E\left[X^2\right] = \sum_{x\in\{-1,499\}} x^2 f_X(x) = (-1)^2 \cdot P(X=-1) + 499^2 \cdot P(X=499)$$

$$= \frac{999}{1000} + \frac{249001}{1000} = 250.$$

Por tanto, $\mathrm{Var}(X) = E[X^2] - E[X]^2 = 250 - 0.5^2 = 249.75.$

2. *Selección al azar de un número en el intervalo* $[0,1]$*:*

$$E\left[X^2\right] = \int_{-\infty}^{+\infty} x^2 f_x(x)\mathrm{d}x = \int_0^1 x^2\mathrm{d}x = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}.$$

Luego $\mathrm{Var}(X) = E[X^2] - E[X]^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.$

- **Distribución de Bernoulli:**

Una V.A. X sigue la distribución de Bernoulli de parámetro $p \in [0,1]$, y lo denotaremos por $X \sim \mathrm{Be}(p)$, si $X \in \{0,1\}$, verificando

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

Podemos calcular la esperanza y varianza de X como sigue:

$$E[X]=p, \quad \mathrm{Var}(X)=p(1-p).$$

Ejemplo

Consideramos el experimento consistente en devolver éxito ($X=1$) si al lanzar un dado obtenemos un 5 y fracaso ($X=0$) en otro caso. Entonces $X \sim \mathrm{Be}(\frac{1}{6})$.

- **Distribución binomial:**

  Si $X = X_1 + \cdots + X_n$, con $X_i \sim \mathrm{Be}(p)$ independientes entre sí, es decir, si X es una V.A. aleatoria que cuenta el número de éxitos tras realizar n pruebas Bernoulli de parámetro p independientes, diremos que X sigue una distribución binomial de parámetros n y p, que denotaremos por $X \sim \mathrm{B}(n, p)$. Se cumple

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{donde } \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

  Por otra parte,

$$E[X] = np, \quad \mathrm{Var}(X) = np(1-p).$$

Ejemplo

Lanzamos un dado 10 *veces y queremos calcular cuál es la probabilidad de obtener tres veces un cinco. Si* X *es la V.A. que cuenta el número de veces que se obtiene un cinco tras lanzar el dado* 10 *veces, entonces* $X \sim \mathrm{B}(10, \frac{1}{6})$*, por lo que* $n = 10$*,* $p = \frac{1}{6}$ *y* $k = 3$*, con lo cual se tiene:*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{10!}{3! \cdot 7!} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^7 = \frac{390625}{2519424}.$$

- **Distribución de Poisson:**

  Supongamos que realizamos un número de pruebas Bernoulli muy alto n, con una probabilidad p_n cada vez más pequeña, de forma que
  $$\lim_{n\to\infty} np_n = \lambda.$$
  Si $X_n \sim \mathrm{B}(n, p_n)$, entonces se puede demostrar que
  $$\lim_{n\to\infty} P(X_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$
  Si X es una variable de este tipo, diremos que X sigue una distribución de Poisson de parámetro λ, que denotaremos por $X \sim \mathrm{Po}(\lambda)$. Se cumple
  $$E[X] = \lambda, \quad \mathrm{Var}(X) = \lambda.$$

Ejemplo

La probabilidad media de tener un accidente de tráfico al coger el coche en una determinada región es de un $0.01\,\%$*. Si una persona coge el coche una media total de* 20000 *veces a lo largo de su vida, ¿cuál es la probabilidad de que sufra algún accidente de tráfico en algún momento de su vida?*

Sea X *la V.A. aleatoria que contabiliza el número de accidentes de tráfico en esas circunstancias. Como* $n = 20000$ *y* $p = 0.0001$*, tenemos* $\lambda = np = 2$*, y podemos decir que* $X \approx \mathrm{Po}(2)$*. Por tanto, se nos pide*

$$P(X > 0) = 1 - P(X = 0) = 1 - \frac{e^{-\lambda}\lambda^0}{0!} = 1 - e^{-2} \approx 0.8647.$$

- **Distribución uniforme:**

  Si X es una V.A. continua que toma valores uniformemente en el intervalo $[a, b]$, entonces diremos que X sigue una distribución uniforme, y lo denotaremos por $X \sim \mathrm{Unif}\big([a, b]\big)$.

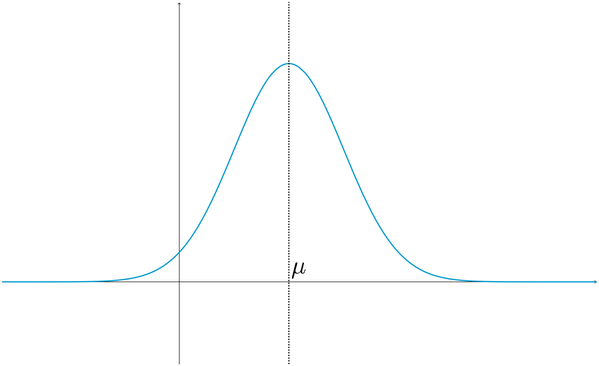
Ejercicio

Si $X \sim \mathrm{Unif}\big([a, b]\big)$, obtener razonadamente F_X, f_x, $E[X]$ y $\mathrm{Var}(X)$.

- **Distribución normal:**

  Una V.A. X sigue una distribución normal de **media** μ y **desviación típica** σ, denotado por $X \sim N(\mu, \sigma)$, si se cumple

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$



Se cumple $E[X] = \mu$ y $\mathrm{Var}(X) = \sigma^2$.

Teorema

Si $X \sim N(\mu, \sigma)$, entonces $\dfrac{X-\mu}{\sigma} \sim N(0,1)$. Usualmente esta última distribución recibe el nombre de **normal tipificada**.

Ejemplo

La altura (en centímetros) de una determinada población sigue una distribución normal de media 175 *y desviación típica* 10*. ¿Cuál es la probabilidad de que una persona elegida al azar mida más de* 185 *cm?*

Sea X *la V.A. asociada, luego* $X \sim N(175, 10)$*. Si definimos* $Z = \frac{X-175}{10} \sim N(0,1)$*, se tiene*

$$\begin{aligned} P\left(X > 185\right) &= P\left(\frac{X-175}{10} > \frac{185-175}{10}\right) \\ &= P\left(Z > 1\right) = 1 - P\left(Z \leq 1\right) \approx 1 - 0.8413 = 0.1587. \end{aligned}$$

Comandos utilizados:

```
from scipy.stats import norm
1-norm.cdf(1)
```

Teorema (Teorema central del límite)

Sea $\{X_i\}_{i=1}^{\infty}$ *una sucesión de variables aleatorias idénticamente distribuidas (en adelante, i.i.d.), con* $E[X_i] = \mu$ *y* $\mathrm{Var}(X_i) = \sigma^2$ $\forall i \in \mathbb{N}$. *Denotamos para cada* $n \in \mathbb{N}$

$$S_n = \sum_{i=1}^{n} X_i \text{ y definimos } Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Entonces se cumple

$$\lim_{n\to\infty} P(Z_n \leq t) = P(Z \leq t),$$

donde $Z \sim N(0,1)$.

En particular, $\overline{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$.

Ejemplo

¿Cuál es la probabilidad de que al lanzar una moneda 400 veces, salga cara en menos de 210 ocasiones?

Sea $X_i \sim \mathrm{Be}(0.5)$ *la V.A. que vale* 1 *si se obtiene cara en el lanzamiento i-ésimo y* 0 *en otro caso, por lo que* $E[X_i] = 0.5$ *y* $\mathrm{Var}[X_i] = 0.5^2 = 0.25$*. Si definimos* $S_{400} = \sum_{i=1}^{400} X_i$*, entonces se nos pide*

$$
\begin{aligned}
P(S_{400} < 210) &= P(S_{400} \leq 209) \\
&= P\left(\frac{S_{400} - 400 \cdot 0.5}{\sqrt{400} \cdot \sqrt{0.25}} \leq \frac{209 - 400 \cdot 0.5}{\sqrt{400} \cdot \sqrt{0.25}}\right) \\
&\approx P(Z < 0.9) \approx 0.8159,
\end{aligned}
$$

donde $Z \sim N(0, 1)$*.*

Teorema (Ley de los grandes números)

Sea $\{X_i\}_{i=1}^{\infty}$ *una sucesión de variables aleatorias independientes cumpliendo* $E[X_i] = \mu$ *y* $\mathrm{Var}(X_i) = \sigma^2\ \forall i \in \mathbb{N}$. *Si denotamos para cada* $n \in \mathbb{N}$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

entonces

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| < \varepsilon) = 1, \quad \forall \varepsilon > 0.$$

Definición

Sean $X_1, \ldots, X_n$ variables aleatorias i.i.d. Una **muestra** *consiste en un conjunto de n valores, $x_1, \ldots, x_n$, obtenidos a partir de dichas variables aleatorias.*

Definición

Dada $\{x_i\}_{i=1}^n$ una muestra, definimos la **media muestral** *asociada a ésta como $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^n x_i$.*

Definición

Dada $\{x_i\}_{i=1}^n$ una muestra, definimos la **varianza muestral** *asociada a ésta como $s^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n (x_i - \overline{x})^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n x_i^2 - \dfrac{n}{n-1}\overline{x}^2$. La* **desviación típica muestral** *se define como $s = +\sqrt{s^2}$.*

Ejemplo

El peso en kg de 7 personas elegidas al azar ha sido el siguiente:

$$x_1 = 74,\ x_2 = 83,\ x_3 = 54,\ x_4 = 62,\ x_5 = 91,\ x_6 = 80,\ x_7 = 57.$$

La media muestral es por tanto

$$\overline{x} = \frac{1}{7}\sum_{i=1}^{7} x_i = \frac{74 + 83 + 54 + 62 + 91 + 80 + 57}{7} \approx 71.57,$$

luego la varianza muestral resulta

$$s^2 = \frac{1}{6}\sum_{i=1}^{7}(x_i - \overline{x})^2 \approx \frac{(74 - 71.57)^2 + \cdots + (57 - 71.57)^2}{6} \approx 199.62.$$

Por tanto, $s = \sqrt{s^2} \approx \sqrt{199.62} \approx 14.13.$

Sea $\alpha \in [0,1]$ un valor de probabilidad y $Z \sim N(0,1)$. Denotamos por z_α al (único) valor real que cumple $P(Z \leq z_\alpha) = \alpha$.

Ejemplo

Sea $Z \sim N(0,1)$. Calculemos $z_{0.9}$. Debemos obtener por tanto el valor $z_{0.9}$ tal que $P(Z \leq z_{0.9}) = 0.9$. Para ello, invocamos el siguiente código en Python

```python
from scipy import stats
stats.norm.ppf(0.9)
```

con el que obtenemos $z_{0.9} \approx 1.2816$.

Sea $\{Z_i\}_{i=1}^n$ una colección de n V.A. independientes tales que $Z_i \sim N(0,1)$. Diremos que la V.A. $X_n^2 = \sum_{i=1}^n Z_i^2$ sigue una **distribución χ^2 de Pearson** con n grados de libertad, que denotaremos por $X_n^2 \sim \chi_n^2$.

Dado $\alpha \in [0,1]$ un valor de probabilidad, denotamos por $x_{n,\alpha}$ al (único) valor real que cumple $P(X_n^2 \leq x_{n,\alpha}) = \alpha$.

Ejemplo

Sea $X_{11}^2 \sim \chi_{11}^2$. Calculemos $x_{11,0.4}$. Es decir, se nos pide obtener el valor $x_{11,0.4}$ tal que $P(X_{11}^2 \leq x_{11,0.4}) = 0.4$. Invocando el siguiente código Python

```python
from scipy import stats
stats.chi.ppf(0.4, 11)
```

obtenemos $x_{11,0.4} \approx 3.0393$.

Sean X, Y V.A. independientes tales que $X \sim N(0,1)$ y $Y \sim \chi^2_n$ para un cierto $n \in \mathbb{N}$. Diremos que la V.A.

$$T_n = \frac{\sqrt{n}X}{\sqrt{Y}}$$

sigue una **distribución t de Student** con n grados de libertad, que denotaremos por $T_n \sim t_n$.
Dado $\alpha \in [0,1]$ un valor de probabilidad, denotamos por $t_{n,\alpha}$ al (único) valor real que cumple $P(T_n \leq t_{n,\alpha}) = \alpha$.

Ejemplo

Sea $T_{23} \sim t_{23}$. Calculemos $t_{23,0.95}$. Debemos obtener por tanto el valor $t_{23,0.95}$ tal que $P(T_{23} \leq t_{23,0.95}) = 0.95$. Código en Python:

```python
from scipy import stats
stats.t.ppf(0.95, 23)
```

Por tanto, $t_{23,0.95} \approx 1.7139$.

Sea $X \sim \chi_m$ e $Y \sim \chi_n$. Diremos que la V.A.

$$F_{m,n} = \frac{X/m}{Y/n}$$

sigue una **distribución F de Snedecor** con (m, n) grados de libertad, y lo denotaremos por $F_{m,n} \sim \mathcal{F}_{m,n}$.
Dado $\alpha \in [0, 1]$ un valor de probabilidad, denotamos por $f_{m,n,\alpha}$ al (único) valor real que cumple $P(F_{m,n} \leq f_{m,n,\alpha}) = \alpha$.

Ejemplo

Sea $F_{17,24} \sim \mathcal{F}_{17,24}$. Calculemos $f_{17,24,0.8}$. Debemos obtener por tanto el valor $f_{17,24,0.8}$ tal que $P(F_{17,24} \leq f_{17,24,0.8}) = 0.8$. Código en Python:

```python
from scipy import stats
stats.f.ppf(0.8, 17, 24)
```

Por tanto, $f_{17,24,0.8} \approx 1.4441$.

- En la práctica, no se conocen los parámetros $N(\mu, \sigma)$ de donde procede la muestra objeto de estudio.
- La media muestral $\overline{x}$ y la desviación típica muestral s son estimadores *puntuales* de dichos parámetros.
- Pero podemos hablar también de una franja de valores entre los cuales se encuentran μ y σ con un determinado nivel de *confianza* o probabilidad, $\alpha \in (0, 1)$.
- Dichos conjuntos de valores reciben el nombre de **intervalos de confianza**, con un nivel de confianza α, o equivalentemente con un nivel de *significación* $\varepsilon = 1 - \alpha$ (margen de error).

Supongamos que $X \sim N(\mu, \sigma)$, con μ y σ desconocidas y $\{x_i\}_{i=1}^{n}$ es una muestra obtenida a partir de X.

- **Intervalo de confianza para la media:**

  En este caso, el intervalo de confianza con un nivel de confianza α (nivel de significación $\varepsilon = 1 - \alpha$) para μ es

$$\left[\overline{x} - t_{n-1,1-\frac{\varepsilon}{2}} \frac{s}{\sqrt{n}}, \overline{x} + t_{n-1,1-\frac{1}{2}} \frac{s}{\sqrt{n}}\right].$$

- **Intervalo de confianza para la varianza:**

  En este caso, el intervalo de confianza con un nivel de confianza α (nivel de significación $\varepsilon = 1 - \alpha$) para σ^2 es

$$\left[\frac{(n-1)s^2}{x_{n-1,1-\frac{\varepsilon}{2}}}, \frac{(n-1)s^2}{x_{n-1,\frac{\varepsilon}{2}}}\right].$$

Ejemplo

Los siguientes datos representan la estatura de 7 niños de 12 años elegidos al azar:

$$143.2, 140.3, 141.4, 139.9, 142.7, 144.1, 142.4.$$

Calcúlese un intervalo de confianza para la media y otro para la varianza con un nivel de confianza de un 90 %.

$n = 7$, $\alpha = 0.9$, $\varepsilon = 1 - \alpha = 1 - 0.9 = 0.1$, $\frac{\varepsilon}{2} = 0.05$, $1 - \frac{\varepsilon}{2} = 0.95$.

$$\overline{x} = 142,\ s^2 = 2.36,\ s \approx 1.5362,$$

$$t_{6,0.95} \approx 1.9432,\ x_{6,0.95} \approx 12.5916,\ x_{6,0.05} \approx 1.6354.$$

Ejemplo

- *Intervalo de confianza para μ:*

$$\left[\overline{x} - t_{6,0.95}\frac{s}{\sqrt{n}}, \overline{x} + t_{6,0.95}\frac{s}{\sqrt{n}}\right]$$

$$\approx \left[142 - 1.9432 \cdot \frac{1.5362}{\sqrt{7}}, 142 + 1.9432 \cdot \frac{1.5362}{\sqrt{7}}\right] \approx [140.87, 143.12].$$

- *Intervalo de confianza para σ^2:*

$$\left[\frac{(n-1)s^2}{x_{6,0.95}}, \frac{(n-1)s^2}{x_{6,0.95}}\right] \approx \left[\frac{6 \cdot 2.36}{12.5916}, \frac{6 \cdot 2.36}{1.6354}\right] \approx [1.12, 8.66].$$

Sean $X \sim N(\mu_X, \sigma_X)$ e $Y \sim N(\mu_Y, \sigma_Y)$ y $\{x_i\}_{i=1}^m$, $\{y_i\}_{i=1}^n$ muestras obtenidas a partir de X e Y, respectivamente.

- **Intervalo de confianza para la diferencia de las medias:** En este caso, el intervalo de confianza con un nivel de confianza α (nivel de significación $\varepsilon = 1-\alpha$) para $\mu_X - \mu_Y$, asumiendo m y n suficientemente grandes, es

$$\left[(\overline{x}-\overline{y}) - z_{1-\frac{\varepsilon}{2}}\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}},\ (\overline{x}-\overline{y}) + z_{1-\frac{\varepsilon}{2}}\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}}\right].$$

- **Intervalo de confianza para el cociente entre varianzas:** En este caso, el intervalo de confianza con un nivel de confianza α (nivel de significación $\varepsilon = 1-\alpha$) para σ_X^2/σ_Y^2 es

$$\left[f_{m-1,n-1,\frac{\varepsilon}{2}}\frac{s_x^2}{s_y^2},\ f_{m-1,n-1,1-\frac{\varepsilon}{2}}\frac{s_x^2}{s_y^2}\right].$$

Ejemplo

Un equipo de fútbol encajó una media de 1.72 *goles por partido tras disputar* 105 *encuentros con un portero A de titular, con una desviación típica de* 0.78*. mientras que encajó una media de* 1.43 *goles por partido tras disputar* 74 *encuentros con un portero B de titular, con una desviación típica de* 0.82*. Obténganse los intervalos de confianza para la diferencia de medias y el cociente entre varianzas con un nivel de confianza de un* 95 %*.*

Se tiene $\alpha = 0.95$*, luego* $\varepsilon = 0.05$*,* $\frac{\varepsilon}{2} = 0.025$ *y* $1 - \frac{\varepsilon}{2} = 0.975$*. Además,* $z_{0.975} \approx 1.96$ *y*

$$m = 105,\ \overline{x} = 1.72,\ s_x = 0.78,\ s_x^2 = 0.6084, n = 74,\ \overline{y} = 1.43,$$

$$s_y = 0.82,\ s_y^2 = 0.6724,\ f_{104,73,0.025} \approx 0.6583,\ f_{104,73,0.975} \approx 1.5437.$$

Ejemplo

- *Intervalo de confianza para $\mu_X - \mu_Y$:*

$$\left[(\overline{x}-\overline{y}) - z_{0.975}\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}},\ (\overline{x}-\overline{y}) + z_{0.975}\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}}\right]$$
$$\approx \left[(1.72-1.43) \pm 1.96\cdot\sqrt{\frac{0.6084}{105}+\frac{0.6724}{74}}\right] \approx [0.05, 0.53].$$

- *Intervalo de confianza para σ_X^2/σ_Y^2:*

$$\left[f_{104,73,0.025}\frac{s_x^2}{s_y^2},\ f_{104,73,0.975}\frac{s_x^2}{s_y^2}\right] \approx \left[0.6583\cdot\frac{0.6084}{0.6724},\ 1.5437\cdot\frac{0.6084}{0.6724}\right]$$
$$\approx [0.6, 1.4].$$

- Un **contraste de hipótesis** consiste en tomar una decisión acerca de un parámetro θ (desconocido) de una distribución a partir de una muestra obtenida de la misma, $\{x_i\}_{i=1}^n$.
- Se contrasta una hipótesis nula, $H_0 : \theta = \theta_0$, frente a una hipótesis alternativa, que puede ser de varios tipos:

$$H_a : \begin{cases} \theta \neq \theta_0 \\ \theta > \theta_0 \\ \theta < \theta_0 \end{cases}$$

- Al pronunciarnos a favor de H_0 o H_a, podemos tomar varias decisiones incorrectas:
  - Rechazar H_0 cuando ésta es cierta (error del primer tipo).
  - Aceptar H_0 cuando ésta es falsa (error del segundo tipo).
- Llamamos ε a la probabilidad de cometer un error del primer tipo, que llamaremos **nivel de significación**.

- **Contraste de hipótesis para la media:**

$H_0 : \mu = \mu_0$. Calculamos

$$K = \frac{\sqrt{n}(\overline{x} - \mu_0)}{s}.$$

Tabla de decisión:

H_a	Aceptar H_0 si	Rechazar H_0 si
$\mu \neq \mu_0$	$\lvert K \rvert \leq t_{n-1,1-\frac{\varepsilon}{2}}$	$\lvert K \rvert > t_{n-1,1-\frac{\varepsilon}{2}}$
$\mu > \mu_0$	$K \leq t_{n-1,1-\varepsilon}$	$K > t_{n-1,1-\varepsilon}$
$\mu < \mu_0$	$K \geq t_{n-1,\varepsilon}$	$K < t_{n-1,\varepsilon}$

Ejemplo

Se ha medido temperatura media anual de una región a lo largo de 20 años, obteniéndose un promedio de 14.3° C, con una desviación típica de 1.8° C. ¿Se puede afirmar con un nivel de confianza de un 95 %, que la temperatura media de la región es de 15° C?

$n = 20$, $\overline{x} = 14.3$, $s = 1.8$.

- $H_0 : \mu = 15$.
- $H_a : \mu \neq 15$.

$\varepsilon = 0.05$, $t_{19,0.975} \approx 2.093$.

$$K = \frac{\sqrt{n}(\overline{x} - \mu_0)}{s} = \frac{\sqrt{20} \cdot (14.3 - 15)}{1.8} \approx -1.7392.$$

Como $|K| = 1.7392 \leq 2.093 = t_{19,0.975}$, *aceptamos* H_0.

- **Contraste de hipótesis para la varianza:**

$H_0 : \sigma^2 = \sigma_0^2$. Calculamos

$$K = \frac{(n-1)s^2}{\sigma_0^2}.$$

Tabla de decisión:

H_a	Aceptar H_0 si	Rechazar H_0 si
$\sigma^2 \neq \sigma_0^2$	$x_{n-1,\frac{\varepsilon}{2}} \leq K \leq x_{n-1,1-\frac{\varepsilon}{2}}$	$K < x_{n-1,\frac{\varepsilon}{2}} \vee K > x_{n-1,1-\frac{\varepsilon}{2}}$
$\sigma^2 > \sigma_0^2$	$K \leq x_{n-1,1-\varepsilon}$	$K > x_{n-1,1-\varepsilon}$
$\sigma^2 < \sigma_0^2$	$K \geq x_{n-1,\varepsilon}$	$K < x_{n-1,\varepsilon}$

Ejemplo

Una máquina que fabrica una pieza determinada se considera defectuosa si la desviación típica de la longitud de las piezas que fabrica supera los 0.5 *cm. Si se ha obtenido una muestra de* 23 *piezas con* $\overline{x} = 45$ *cm y* $s = 0.8$ *cm. Se puede afirmar, con un nivel de confianza de un* 99 %*, que la máquina es defectuosa?*

- $H_0 : \sigma^2 = 0.5^2 = 0.25.$
- $H_a : \sigma^2 > 0.5^2 = 0.25.$

$\varepsilon = 0.01$, $x_{22,0.99} \approx 40.2894.$

$$K = \frac{(n-1)s^2}{\sigma_0^2} = \frac{22 \cdot 0.8^2}{0.5^2} = 56.32.$$

Como $K = 56.32 > 40.2894 \approx x_{22,0.99}$*, rechazamos* H_0.

- **Contraste de dos medias:**

$H_0 : \mu_X = \mu_Y$. Supongamos m, n suficientemente grandes. Calculamos

$$K = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}}.$$

Tabla de decisión:

H_a	Aceptar H_0 si	Rechazar H_0 si
$\mu_X \neq \mu_Y$	$\lvert K \rvert \leq z_{1-\frac{\varepsilon}{2}}$	$\lvert K \rvert > z_{1-\frac{\varepsilon}{2}}$
$\mu_X > \mu_Y$	$K \leq z_{1-\varepsilon}$	$K > z_{1-\varepsilon}$
$\mu_X < \mu_Y$	$K \geq z_{\varepsilon}$	$K < z_{\varepsilon}$

Ejemplo

Un equipo de fútbol encajó una media de 1.72 *goles por partido tras disputar* 105 *encuentros con un portero A de titular, con una desviación típica de* 0.78*. mientras que encajó una media de* 1.43 *goles por partido tras disputar* 74 *encuentros con un portero B de titular, con una desviación típica de* 0.82*. ¿Se puede concluir, con un nivel de confianza de un* 95 %*, que el portero B es mejor que el portero A?*

- $H_0 : \mu_A = \mu_B$
- $H_a : \mu_A > \mu_B$

$\varepsilon = 0.95$, $z_{0.95} \approx 1.6449$.

$$K = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}} = \frac{1.72 - 1.43}{\sqrt{\frac{0.78^2}{105} + \frac{0.82^2}{74}}} \approx 2.3773.$$

$K \approx 2.3773 > 1.6449 \approx z_{0.95}$, *por lo que rechazamos* H_0.

- **Contraste de dos varianzas:**

  $H_0 : \sigma_X^2 = \sigma_Y^2$. Calculamos

$$K = \frac{s_x^2}{s_y^2}.$$

  Tabla de decisión:

H_a	Aceptar H_0 si	Rechazar H_0 si
$\sigma_X^2 \neq \sigma_Y^2$	$f_{m-1,n-1,\frac{\varepsilon}{2}} \leq K \leq f_{m-1,n-1,1-\frac{\varepsilon}{2}}$	$K < f_{m-1,n-1,\frac{\varepsilon}{2}} \vee K > f_{m-1,n-1,1-\frac{\varepsilon}{2}}$
$\sigma_X^2 > \sigma_Y^2$	$K \leq f_{m-1,n-1,1-\varepsilon}$	$K > f_{m-1,n-1,1-\varepsilon}$
$\sigma_X^2 < \sigma_Y^2$	$K \geq f_{m-1,n-1,\varepsilon}$	$K < f_{m-1,n-1,\varepsilon}$

Ejemplo

Comparemos las varianzas asociadas a los goles encajados de los porteros del ejemplo anterior.

- *Portero A:* $n = 105$, $s_x = 0.78$.
- *Portero B:* $n = 74$, $s_y = 0.82$.

Nos interesa saber si existe alguna diferencia significativa entre las varianzas con un nivel de confianza de un $99\,\%$.

- $H_0 : \sigma_A^2 = \sigma_B^2$.
- $H_a : \sigma_A^2 \neq \sigma_B^2$.

$\varepsilon = 0.01$, $f_{104,73,0.01} \approx 0.6088$, $f_{104,73,0.99} \approx 1.6766$.

$$\frac{s_x^2}{s_y^2} = \frac{0.78^2}{0.82^2} \approx 0.9048.$$

$f_{104,73,0.01} \leq K \leq f_{104,73,0.99}$, *luego aceptamos* H_0.

- **Contraste del parámetro de una binomial:**

  Sea A un suceso tal que $p(A) = p$, con p desconocida. Realizamos N experimentos, con N suficientemente grande y contamos el número de ocurrencias de A, n. Queremos contrastar $H_0 : p = p_0$. Calculamos

$$K = \frac{\frac{n}{N} - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} = \frac{n - Np_0}{\sqrt{Np_0(1-p_0)}}.$$

  Tabla de decisión:

H_a	Aceptar H_0 si	Rechazar H_0 si
$p \neq p_0$	$\lvert K \rvert \leq z_{1-\frac{\varepsilon}{2}}$	$\lvert K \rvert > z_{1-\frac{\varepsilon}{2}}$
$p > p_0$	$K \leq z_{1-\varepsilon}$	$K > z_{1-\varepsilon}$
$p < p_0$	$K \geq z_{\varepsilon}$	$K < z_{\varepsilon}$

Ejemplo

Efectuamos el lanzamiento de una moneda 1000 *veces, de las cuales en* 563 *ocasiones sale cara. ¿Se puede afirmar, con un nivel de significación de un* 99 %*, que la moneda está trucada?*

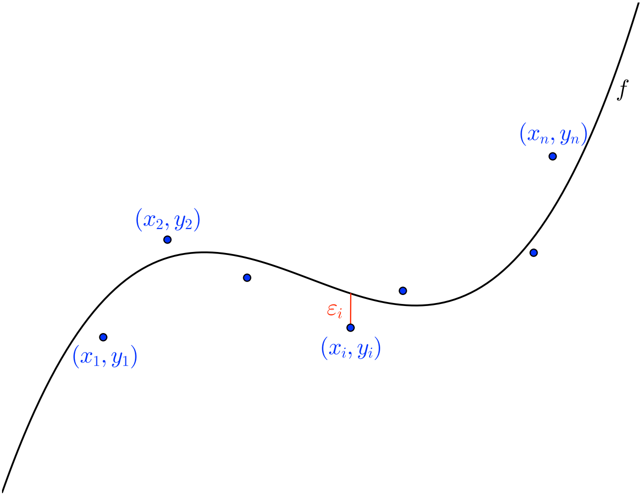
- $H_0 : p = 0.5$.
- $H_a : p \neq 0.5$.

$N = 1000$, $n = 563$ $\varepsilon = 0.01$, $z_{0.995} \approx 2.5758$.

$$K = \frac{n - Np_0}{\sqrt{Np_0(1-p_0)}} = \frac{563 - 1000 \cdot 0.5}{\sqrt{1000 \cdot 0.5 \cdot (1-0.5)}} \approx 3.9845.$$

$|K| \approx 3.9845 > 2.5758 \approx z_{0.995}$, *por lo que rechazamos* H_0.

En ocasiones, nos puede interesar ajustar los datos de algún fenómeno procedente de una V.A. Y dependiente de otra V.A. X, mediante un tipo de concreto de funciones.



- Dado un conjunto de datos $\{(x_i, y_i)\}_{i=1}^n$, buscamos una recta de la forma $f(x) = a + bx$ que "mejor" se ajuste los datos.
- Un criterio consiste en el método de los **mínimos cuadrados**, que consiste en minimizar la suma de los cuadrados de los errores del ajuste, $\varepsilon_i = y_i - f(x_i)$, $1 \leq i \leq n$.
- En otras palabras, encontrar los valores de a y b que consiguen minimizar esta expresión (la recta resultante recibe el nombre de **recta de regresión**):

$$\mathcal{E}(a, b) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \big(y_i - f(x_i)\big)^2 = \sum_{i=1}^{n} \big(y_i - a - bx_i\big)^2.$$

- Debemos por tanto encontrar el mínimo de la función cuadrática $\mathcal{E}(a, b)$, por lo que imponemos

$$\frac{\partial \mathcal{E}}{\partial a}(a, b) = 0, \quad \frac{\partial \mathcal{E}}{\partial b}(a, b) = 0.$$

- Derivando en cada caso, obtenemos

$$\frac{\partial \mathcal{E}}{\partial a}(a,b) = \sum_{i=1}^{n} 2(y_i - a - bx_i),$$

$$\frac{\partial \mathcal{E}}{\partial b}(a,b) = \sum_{i=1}^{n} -2x_i(y_i - a - bx_i).$$

- Imponiendo $\frac{\partial \mathcal{E}}{\partial a}(a,b) = 0$ y $\frac{\partial \mathcal{E}}{\partial b}(a,b) = 0$, y resolviendo el sistema de ecuaciones para a y b, obtenemos

$$b = \frac{s_{xy}}{s_x^2}, \quad a = \overline{y} - b\overline{x},$$

donde

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2, \quad s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}).$$

Ejemplo

Sea la siguiente muestra formada por 5 *puntos de* $\mathbb{R}^2$*:*

$$\{(1,0.5),(1.5,1.3),(2.5,2.1),(3,2),(4,2.7)\}.$$

Se pide hallar la recta de regresion asociada a dichos puntos.

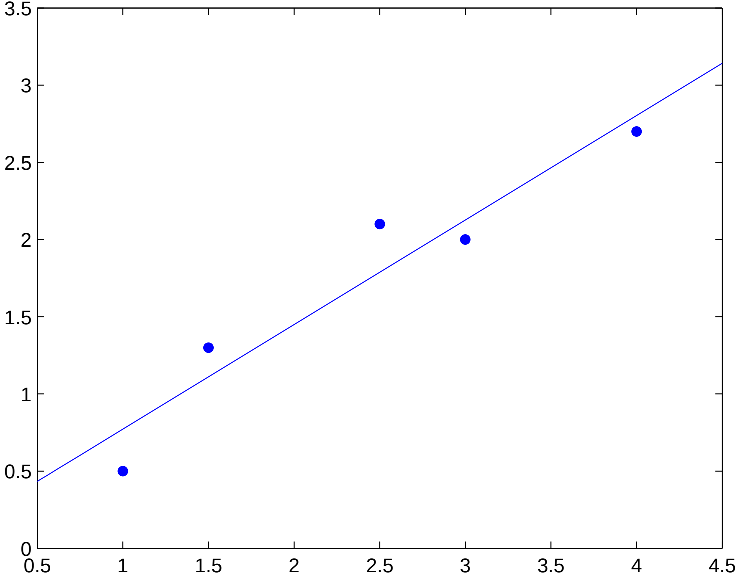
$$\overline{x} = 2.4,\ \overline{y} = 1.72,\ s_x^2 = 1.14,$$

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy}$$

$$= \frac{1\cdot 0.5 + 1.5\cdot 1.3 + 2.5\cdot 2.1 + 3\cdot 2 + 4\cdot 2.7}{5} - 2.4\cdot 1.72 = 0.772.$$

$$b = \frac{s_{xy}}{s_x^2} = \frac{0.772}{1.14} \approx 0.6772, \quad a = \overline{y} - b\overline{x} \approx 1.72 - 0.6772\cdot 2.4 \approx 0.0947.$$

Luego la recta de regresión es $y \approx 0.0947 + 0.6772x$.


3.5
3
2.5
2
1.5
1
0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5

Definición

Dada una muestra $\{(x_i, y_i)\}_{i=1}^{n}$, *se define el* **coeficiente de correlación** *como*

$$r = \frac{s_{xy}}{s_x s_y}.$$

Propiedades

1. $|r| \leq 1$.
2. *Si* $|r| = 1$, *entonces el ajuste es perfecto, es decir, la recta de regresión pasa por todos los puntos de la muestra.*

Ejemplo

Coeficiente de correlación de la recta de regresión anterior:

$$r = \frac{s_{xy}}{s_x s_y} = \frac{0.772}{\sqrt{1.14} \cdot \sqrt{0.5696}} \approx 0.958.$$