

Estadística Matemática I

Diplomatura Estadística (2003/2004)

Contenidos

Artículos

Axiomatica de Kolmogorov	1
Axiomas de probabilidad	1
Sigma-álgebra	3
Variables Aleatorias	5
Dominio de definición	5
Función de densidad de probabilidad	7
Variable aleatoria	9
Esperanza matemática	14
Media aritmética	15
Mediana (estadística)	17
Moda (estadística)	21
Varianza	24
Asimetría estadística	27
Desviación estándar	29
Covarianza	33
Correlación	35
Regresión lineal	36
Función generadora de momentos	41
Distribuciones	43
Distribución de probabilidad	43
Distribución uniforme continua	46
Distribución gamma	51
Distribución exponencial	52
Distribución beta	54
Distribución normal	56
Distribuciones Multivariantes	75
Distribución binomial	75
Distribución multinomial	78
Distribución de Poisson	79
Distribución hipergeométrica	83
Distribución normal multivariante	84

Transformación de Vectores Aleatorios	92
Distribución χ^2	92
Distribución t de Student	95
Distribución F	98
Problema Central del Límite	100
Desigualdad de Chebyshev	100
Desigualdad de Márkov	101
Ley de los grandes números	102
Teorema del límite central	103
Referencias	
Fuentes y contribuyentes del artículo	106
Fuentes de imagen, Licencias y contribuyentes	108
Licencias de artículos	
Licencia	109

Axiomática de Kolmogorov

Axiomas de probabilidad

Los **axiomas de probabilidad** son las condiciones mínimas que deben verificarse para que una función definida sobre un conjunto de sucesos determine consistentemente sus probabilidades. Fueron formulados por Kolmogórov en 1933.

Axiomas de Kolmogórov

Dado un conjunto de sucesos elementales, Ω , sobre el que se ha definida una σ -álgebra (léase sigma-álgebra) σ de subconjuntos de Ω y una función P que asigna valores reales a los miembros de σ , a los que denominamos "sucesos", se dice que P es una probabilidad sobre (Ω, σ) si se cumplen los siguientes tres axiomas.

Primer axioma

La probabilidad de un suceso A es un número real mayor o igual que 0.

$$P(A) \geq 0$$

Segundo axioma

La probabilidad del total, Ω , es igual a 1, es decir,

$$P(\Omega) = 1$$

Tercer axioma

Si A_1, A_2, \dots son sucesos mutuamente excluyentes (incompatibles dos a dos, disjuntos o de intersección vacía dos a dos), entonces:

$$P(A_1 \cup A_2 \cup \dots) = \sum P(A_i).$$

Según este axioma se puede calcular la probabilidad de un suceso compuesto de varias alternativas mutuamente excluyentes sumando las probabilidades de sus componentes.

En términos más formales, una probabilidad es una medida sobre una σ -álgebra de subconjuntos del espacio muestral, siendo los subconjuntos miembros de la σ -álgebra los sucesos y definida de tal manera que la medida del total sea 1. Tal medida, gracias a su definición matemática, verifica igualmente los tres axiomas de Kolmogórov. A la terna formada por el espacio muestral, la σ -álgebra y la función de probabilidad se la denomina **Espacio probabilístico**, esto es, un "espacio de sucesos" (el espacio muestral) en el que se han definido los posibles sucesos a considerar (la σ -álgebra) y la probabilidad de cada suceso (la función de probabilidad).

Propiedades que se deducen de los axiomas

De los axiomas anteriores se deducen otras propiedades de la probabilidad:

1. $P(\emptyset) = 0$ donde el conjunto vacío (\emptyset) representa en probabilidad el **suceso imposible**
2. Para cualquier suceso $P(A) \leq 1$
3. $P(A^c) = 1 - P(A)$
4. Si $A \subseteq B$ entonces $P(A) \leq P(B)$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Ejemplos

Como ejemplo se puede tomar como espacio muestral a los posibles resultados al arrojar un dado corriente $\Omega = \{1, 2, 3, 4, 5, 6\}$, tomaremos como σ -álgebra todos los subconjuntos posibles de Ω (que en matemáticas se denota por $\mathcal{P}(\Omega)$) y como función de probabilidad

$$P(A) = \frac{\#A}{6}$$

donde $\#A$ representa el número de elementos del conjunto A .

Es fácil comprobar que esta función verifica los tres axiomas de Kolmogórov y, por tanto, constituye una probabilidad sobre este conjunto.

1. $P(A) = \frac{\#A}{6} \geq 0$, puesto que es el cociente de dos números positivos
2. $P(\Omega) = P(\{1, 2, 3, 4, 5, 6\}) = \frac{\#\{1, 2, 3, 4, 5, 6\}}{6} = \frac{6}{6} = 1$
3. Si $A = A_1 \cup A_2 \cup A_3 \cup \dots$ de tal manera que $A_i \cap A_j = \emptyset \quad \forall i \neq j$ entonces

$$\#A = \#A_1 + \#A_2 + \#A_3 + \dots$$
 con lo que $P(A) = \sum P(A_i)$

Véase también

- Axioma
- σ -álgebra o Sigma-álgebra
- Andréi Kolmogórov
- Teoremas de Probabilidad

Sigma-álgebra

En matemáticas, una **σ -álgebra** (léase *sigma-álgebra*) sobre un conjunto X es una familia Σ no vacía de subconjuntos de X , cerrada bajo complementos, uniones e intersecciones contables. Las σ -álgebras (también conocidas como "tribus") se usan principalmente para definir medidas en X . El concepto es muy importante en análisis matemático y en teoría de la probabilidad.

Definición formal

Formalmente, una familia de conjuntos de X (o, equivalentemente, un subconjunto del conjunto de las partes de X), a la que llamaremos Σ es una σ -álgebra sobre X si y solo si se cumplen las siguientes propiedades:

1. El conjunto vacío está en Σ , es decir, $\emptyset \in \Sigma$.
2. Si E está en Σ , también está su conjunto complementario $X \setminus E$.
3. Si E_1, E_2, E_3, \dots es una sucesión (contable) en Σ , entonces su unión (contable) también está en Σ . Entiéndase *contable* como *finito o numerable*.

En virtud de la segunda propiedad, la primera propiedad equivale a $X \in \Sigma$; de 2 y 3 se concluye que la σ -álgebra también es cerrada bajo intersecciones contables (gracias a las leyes de De Morgan).

Los elementos de la σ -álgebra se denominan *conjuntos Σ -medibles* (o simplemente *conjuntos medibles*, cuando no hay ambigüedad sobre Σ). Un par ordenado (X, Σ) , donde X es un conjunto y Σ una σ -álgebra sobre éste, se denomina **espacio medible**. Una función entre dos espacios medibles se denomina *medible* si la preimagen de todo conjunto medible es también medible; esto es, si (X, Σ) y (Y, Ω) son dos espacios medibles, una función $f: X \rightarrow Y$ es *medible* si para todo $E \in \Omega$, $f^{-1}(E) \in \Sigma$.

Una medida es una cierta clase de función medible de una σ -álgebra en el intervalo $[0, \infty]$.

Notación

Las σ -álgebras usualmente se denotan con letras manuscritas mayúsculas en lugar de Σ , con lo que (X, \mathcal{A}) se usa en lugar de (X, Σ) . Esto es útil para evitar que Σ se confunda con el operador de sumatoria.

Espacio medible

Como se ha dicho antes, un par (X, Σ) formado por un conjunto X y un σ -álgebra Σ se denomina espacio medible, o también espacio de Borel.^[1] Algunos autores, sin embargo, reservan el término *espacio de Borel* a aquellas σ -álgebras generadas por una topología.

Es importante no confundir este término con el muy parecido y muy relacionado *espacio de medida*. Un espacio de medida consiste en un espacio medible dotado de una medida.

Ejemplos

- Si X es cualquier conjunto, la familia $\{ \emptyset, X \}$ es una σ -álgebra sobre X , llamada *σ -álgebra trivial* por razones obvias.
 - El conjunto de partes de X .
 - La familia de subconjuntos de X que son contables o de conjunto complementario contable (esta familia es distinta del conjunto de las partes de X si y solo si X es incontable). Esta es la σ -álgebra generada por los conjuntos unitarios de X .
 - Si $\{ \Sigma_\alpha \}$ es una familia de σ -álgebras sobre X , la intersección de todos los conjuntos Σ_α es también una σ -álgebra sobre X .
-

Si U es una familia arbitraria de subconjuntos de X , existe una mínima σ -álgebra sobre X que contiene a U , llamada *σ -álgebra generada por U* . Ésta se denota por $\sigma(U)$, y se puede construir como sigue:

- Es claro que existe al menos una σ -álgebra sobre X que contiene a U ; a saber, el conjunto de partes de X .
- Sea Φ la familia (no vacía) de todas las σ -álgebras sobre X que contienen a U (esto es, una σ -álgebra Σ sobre X está en Φ si y solo si $U \subseteq \Sigma$).
- Defínase entonces $\sigma(U)$ como la intersección de todas las σ -álgebras en Φ . Por el párrafo anterior, $\sigma(U)$ es una σ -álgebra sobre X ; y por construcción, es la mínima que contiene a U .

Esto lleva a lo que tal vez sea el ejemplo más importante: el álgebra de Borel, o boreliana, sobre un espacio topológico es la σ -álgebra generada por el conjunto de conjuntos abiertos (o equivalentemente, el conjunto de conjuntos cerrados). En general, esta σ -álgebra no es el conjunto de partes, lo cual puede demostrarse usando el axioma de elección.

En el espacio euclídeo \mathbf{R}^n , cabe destacar otra σ -álgebra: la formada por los conjuntos Lebesgue-medibles. Ésta contiene más conjuntos que el álgebra de Borel en \mathbf{R}^n , y es la que se prefiere en teoría de integración.

Véase también

- Álgebra de conjuntos

Referencias

[1] <http://planetmath.org/encyclopedia/BorelSubsetSpace.html>

Variables Aleatorias

Dominio de definición

En matemáticas, el **dominio** (**conjunto de definición** o **conjunto de partida**) de una función $f: X \rightarrow Y$ es el conjunto de existencia de ella misma, es decir, los valores para los cuales la función está definida. Es el conjunto de todos los objetos que puede transformar, se denota Dom_f o bien D_f y está definido por:

$$D_f = \{x \in X \mid \exists y \in Y : f(x) = y\}$$

En \mathbb{R}^n se denomina dominio un conjunto conexo, abierto y cuyo interior no sea vacío.

Propiedades

Dadas dos funciones reales:

$$f: X_1 \rightarrow \mathbb{R} \quad \text{y} \quad g: X_2 \rightarrow \mathbb{R}$$

Se tienen las siguientes propiedades:

1. $D_{(f+g)} = X_1 \cap X_2$
2. $D_{(f-g)} = X_1 \cap X_2$
3. $D_{(f \cdot g)} = X_1 \cap X_2$
4. $D_{(f/g)} = \{x \in (X_1 \cap X_2) \mid g(x) \neq 0\}$

Ejemplos

Algunos dominios de funciones reales de variable real:

$f(x) = x^2$ El dominio de esta función es \mathbb{R}

$f(x) = \frac{1}{x}$ El dominio de esta función es $\mathbb{R} - \{0\}$ puesto que la función no está definida para $x = 0$.

$f(x) = \log(x)$ El dominio de esta función es $(0, +\infty)$ ya que los logaritmos están definidos sólo para números positivos.

$f(x) = \sqrt{x}$ El dominio de esta función es $[0, +\infty)$ porque la raíz de un número negativo no existe en el campo de los Reales.

Cálculo del dominio de una función

Para el cálculo certero del dominio de una función, debemos introducir el concepto de restricción en el campo real. Estas restricciones nos ayudarán a identificar la existencia del dominio de una función. Las más usadas son:

Raíz n -ésima de $f(x)$

No existe restricción si n es impar, pero si n es par, la función $f(x)$ necesariamente deberá ser mayor o igual que cero, ya que las raíces negativas no están definidas en el campo real. Por ejemplo:

$$y = \sqrt{7x - 21}$$

El índice de la raíz es par (2), por tanto

$7x - 21 \geq 0$ despejando tenemos que

$x \geq 3$ El dominio entonces será el conjunto de todos los reales en el intervalo $[3, +\infty)$

Logaritmo de f(x)

La restricción está al estudiar las propiedades de los logaritmos las cuales nos dicen que estos no están definidos para números negativos, por tanto toda función contenida dentro de un logaritmo es necesariamente mayor estricto de cero. Por ejemplo:

$\log(x^2 - 9)$ Por la propiedad anteriormente citada tenemos que para que esta función exista, necesariamente $x^2 - 9 > 0$ despejando obtendremos dos soluciones $x > 3$ y $x < -3$. La unión de ambas soluciones representa el dominio de la función, que está definida como el conjunto $(-\infty, -3) \cup (3, +\infty)$.

Fracciones

Véase también: *División por cero*

Otras propiedades de las matemáticas nos pueden ayudar a obtener el dominio de una función y excluir puntos donde esta no está definida, por ejemplo, una función que tenga forma de *fracción no estará definida cuando el denominador valga cero*, ya que esto es una indeterminación que daría una tendencia al infinito. Veamos

la función $y = \frac{7 - 3x}{10x - 2}$ no estará definida cuando $10x - 2 = 0$, despejando $x = 1/5$, es decir la variable x

debe tener un valor diferente para poder existir, ya que en ese punto no está definida, por tanto el dominio de esta función será el conjunto de todos los reales menos ese punto. Su notación será $R - \{1/5\}$, que se lee, el conjunto de todos los reales menos el punto $0,20$.

El grado de dificultad se incrementa cuando buscamos el dominio de una función con variable en el denominador contenida dentro de un radical de índice par o logaritmo, ya que esto nos traslada a resolver una desigualdad. No obstante, el método de polos y ceros nos permite resolver esta clase de inecuaciones con facilidad.

Ejemplo

Para evidenciar este caso veamos este problema. Hallar el dominio de la siguiente función:

$$f(x) = \log\left(\frac{5x + 1}{3x - 2}\right),$$

Para que esta función exista, necesariamente

$$\frac{5x + 1}{3x - 2} > 0,$$

Ya que no existe logaritmo de expresiones negativas. La solución de esta desigualdad, es explicada paso por paso en el artículo polos y ceros anteriormente citado, su solución constituirá el dominio de la función que en este caso será:

solución: $(-\infty, -1/5) \cup (2/3, +\infty)$

Véase también

- Codominio
- Imagen

Función de densidad de probabilidad

En teoría de la probabilidad, la **función de densidad de probabilidad**, **función de densidad**, o, simplemente, **densidad** de una variable aleatoria continua es una función, usualmente denominada $f(x)$ que describe la densidad de la probabilidad en cada punto del espacio de tal manera que la probabilidad de que la variable aleatoria tome un valor dentro de un determinado conjunto sea la integral de la función de densidad sobre dicho conjunto.

Definición

Una función de densidad de probabilidad (FDP) es una función matemática que caracteriza el comportamiento probable de una población. Es una función $f(x)$ que especifica la posibilidad relativa de que una variable aleatoria continua X tome un valor cercano a x , y se define como la probabilidad de que X tome un valor entre x y $x+dx$, dividido por dx cuando dx es un número infinitesimalmente pequeño. La mayoría de las funciones de densidad de probabilidad requieren uno o más parámetros para especificarlas totalmente.

La probabilidad de que una variable aleatoria continua X esté ubicada entre los valores a y b está dada por el intervalo de la FDP, $f(x)$, comprendido en el rango entre a y b . $\Pr(a \leq X \leq b) = \int_a^b f(x) dx$ La FDP es la derivada (cuando existe) de la función de distribución: $f(x) = \frac{dF(x)}{dx}$ En situaciones prácticas, la FDP utilizada se elige entre un número relativamente pequeño de FDP comunes, y la labor estadística principal consiste en estimar sus parámetros. Por lo tanto, a los efectos de los inventarios, es necesario saber qué FDP se ha utilizado e indicarlo en la documentación de evaluación de la incertidumbre.

La definición formal de la función de densidad requiere de conceptos de la teoría de la medida. Si una variable aleatoria X sigue una función de probabilidad X_*P su **densidad** con respecto a una medida de referencia μ es la derivada de Radon–Nikodym

$$f = \frac{dX_*P}{d\mu}.$$

Es decir, f es una función con la propiedad de que

$$P[X \in A] = \int_{X^{-1}A} dP = \int_A f d\mu$$

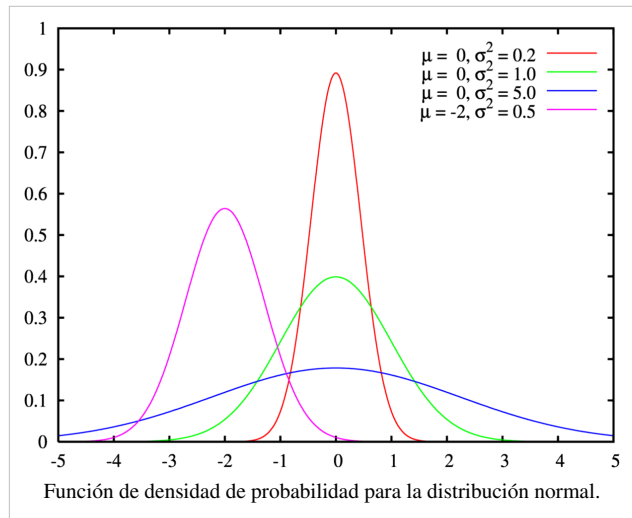
para cada conjunto medible A .

Hay que advertir que la función de densidad no es propiamente única: dos funciones distintas pueden representar la misma distribución de probabilidad si son distintas únicamente en un conjunto de medida nula. Además, que puede haber distribuciones de probabilidad que carezcan de función de densidad: sucede cuando, sin ser discretas, concentran su probabilidad en conjuntos de medida nula; así sucede con la distribución de Cantor cuando se toma la de Lebesgue como medida de referencia.

Cuando, como ocurre normalmente en las aplicaciones, X es una variable aleatoria real y μ es la medida de Lebesgue, la función de densidad es una función tal que

$$P[a \leq X \leq b] = \int_a^b f(x) dx.$$

De modo que si F es la función de distribución de X , entonces



$$F(x) = \int_{-\infty}^x f(u) \, du,$$

y

$$f(x) = \frac{d}{dx} F(x).$$

Intuitivamente, se puede pensar que $f(x) \, dx$ es la probabilidad de que X asuma valores en el intervalo *infinitesimal* $[x, x + dx]$.

Propiedades

De las propiedades de la función de distribución se siguen las siguientes propiedades de la fdp (a veces visto como pdf ^[1] del inglés):

- $f(x) \geq 0$ para toda x .
- El área total encerrada bajo la curva es igual a 1:

$$\int_{-\infty}^{\infty} f(x) \, dx = 1$$

- La probabilidad de que X tome un valor en el intervalo $[a, b]$ es el área bajo la curva de la función de densidad en ese intervalo o lo que es lo mismo, la integral definida en dicho intervalo. La gráfica $f(x)$ se conoce a veces como *curva de densidad*.

$$\Pr(a \leq X \leq b) = \int_a^b f(x) \, dx = F(b) - F(a)$$

Algunas FDP están declaradas en rangos de $-\infty$ a $+\infty$, como la de la distribución normal.

Enlaces externos

- [2] Simulación de la obtención de la probabilidad en un intervalo a partir de la función de densidad de una variable continua con R (lenguaje de programación)

Véase también

- Polinomio de Bernstein

Referencias

[1] http://en.wikipedia.org/wiki/Probability_density_function

[2] <http://cajael.com/mestadisticos/T3VAleatorias/node2.php>

Variable aleatoria

Una variable es aleatoria si su valor está determinado por el azar. En gran número de experimentos aleatorios es necesario, para su tratamiento matemático, cuantificar los resultados de modo que se asigne un número real a cada uno de los resultados posibles del experimento. De este modo se establece una relación funcional entre elementos del espacio muestral asociado al experimento y números reales.

En probabilidad y estadística, una **variable aleatoria** o **variable estocástica** es una variable cuyos valores se obtienen de mediciones en algún tipo de experimento aleatorio. Formalmente, una variable aleatoria es una función, que asigna eventos (p.e., los posibles resultados de tirar un dado dos veces: (1, 1), (1, 2), etc.) a números reales (p.e., su suma).

Los valores posibles de una variable aleatoria pueden representar los posibles resultados de un experimento aún no realizado, o los posibles valores de una cantidad cuyo valor actualmente existente es incierto (p.e., como resultado de medición incompleta o imprecisa). Intuitivamente, una variable aleatoria puede tomarse como una cantidad cuyo valor no es fijo pero puede tomar diferentes valores; una distribución de probabilidad se usa para describir la probabilidad de que se den los diferentes valores.

Las variables aleatorias suelen tomar valores reales, pero se pueden considerar valores aleatorios como valores lógicos, funciones... El término *elemento aleatorio* se utiliza para englobar todo ese tipo de conceptos relacionados. Un concepto relacionado es el de proceso estocástico, un conjunto de variables aleatorias ordenadas (habitualmente por orden o tiempo).

Una **variable aleatoria** (v.a.) X es una función real definida en el espacio muestral asociado a un experimento aleatorio, Ω .^{[1] [2]}

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

Se llama **rango** de una v.a. X y lo denotaremos R_X , al conjunto de los valores reales que ésta puede tomar, según la aplicación X . Dicho de otro modo, el rango de una v.a. es el recorrido de la función por la que ésta queda definida:

$$R_X = \{x \in \mathbb{R} \mid \exists \omega \in \Omega : X(\omega) = x\}$$

Definición de variable aleatoria

Concepto intuitivo

Una variable es aleatoria si su valor está determinado por el azar. En otras palabras se sabe qué valores puede tomar la variable pero no se tiene certeza de su ocurrencia, solo se sabe que puede ocurrir con cierta probabilidad. Por ejemplo, en una epidemia de cólera, se sabe que una persona cualquiera puede enfermar o no (suceso), pero no se sabe cual de los dos sucesos va a ocurrir. Solamente se puede decir que existe una probabilidad de que la persona enferme.

Definición formal

La definición formal de variable aleatoria requiere ciertos conocimientos profundos de matemática (en concreto de teoría de la medida). Es la siguiente:^{[3] [4]}

Dado un espacio de probabilidad (Ω, \mathcal{A}, P) y un espacio medible (S, Σ) , una aplicación $X : \Omega \rightarrow S$ es una **variable aleatoria** si es una aplicación \mathcal{A}, Σ -medible.

En la mayoría de los casos se toma como espacio medible de llegada el formado por los números reales junto con la σ -álgebra de Borel (el generado por la topología usual de \mathbb{R}), quedando pues la definición de esta manera:

Dado un espacio de probabilidad (Ω, \mathcal{A}, P) una variable aleatoria real es cualquier función $\mathcal{A}/\mathcal{B}(\mathbb{R})$ -medible donde $\mathcal{B}(\mathbb{R})$ es la σ -álgebra boreliana.

Ejemplo

Supongamos que se lanzan dos monedas al aire. El espacio muestral, esto es, el conjunto de resultados elementales posibles asociado al experimento, es

$$\Omega = \{cc, cx, xc, xx\},$$

donde (c representa "sale cara" y x, "sale cruz").

Podemos asignar entonces a cada suceso elemental del experimento el número de caras obtenidas. De este modo se definiría la variable aleatoria X como la función

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

dada por

$$cc \rightarrow 2$$

$$cx, xc \rightarrow 1$$

$$xx \rightarrow 0$$

El recorrido o rango de esta función, R_X , es el conjunto

$$R_X = \{0, 1, 2\}$$

Tipos de variables aleatorias

Para comprender de una manera más amplia y rigurosa los tipos de variables, es necesario conocer la definición de **conjunto discreto**. Un conjunto es discreto si está formado por un número finito de elementos, o si sus elementos se pueden enumerar en secuencia de modo que haya un primer elemento, un segundo elemento, un tercer elemento, y así sucesivamente.^[5]

- **Variable aleatoria discreta:** una v.a. es discreta si su recorrido es un conjunto discreto. La variable del ejemplo anterior es discreta. Sus probabilidades se recogen en la función de cuantía (véanse las distribuciones de variable discreta).
- **Variable aleatoria continua:** una v.a. es continua si su recorrido no es un conjunto numerable. Intuitivamente esto significa que el conjunto de posibles valores de la variable abarca todo un intervalo de números reales. Por ejemplo, la variable que asigna la estatura a una persona extraída de una determinada población es una variable continua ya que, teóricamente, todo valor entre, pongamos por caso, 0 y 2,50 m, es posible.^[6] (véanse las distribuciones de variable continua)
- **Variable aleatoria independiente:** Supongamos que "X" e "Y" son variables aleatorias discretas. Si los eventos $X = x / Y = y$ son variables aleatorias independientes. En tal caso: $P(X = x, Y = y) = P(X = x) P(Y = y)$.

De manera equivalente: $f(x,y) = f_1(x).f_2(y)$.

Inversamente, si para todo "x" e "y" la función de probabilidad conjunta $f(x,y)$ **no** puede expresarse sólo como el producto de una función de "x" por una función de "y" (denominadas funciones de probabilidad marginal de "X" e "Y"), entonces "X" e "Y" son **dependientes**.

Si "X" e "Y" son variables aleatorias continuas, decimos que son variables aleatorias independientes si los eventos $X \leq x$, e $Y \leq y$ y son eventos independientes para todo "x" e "y" .

De manera equivalente: $F(x,y) = F_1(x).F_2(y)$, donde $F_1(x)$ y $F_2(y)$ son las funciones de distribución (marginal) de "X" e "Y" respectivamente.

Inversamente, "X" e "Y" son variables aleatorias **dependientes** si para todo "x" e "y" su función de distribución conjunta $F(x,y)$ **no** puede expresarse como el producto de las funciones de distribución marginales de "X" e "Y".

Para variables aleatorias independientes continuas, también es cierto que la función de densidad conjunta $f(x,y)$ es el producto de las funciones densidad de probabilidad marginales de "X", $f_1(x)$, y de "Y", $f_2(y)$.

Distribución de probabilidad de una v.a.

La **distribución de probabilidad** de una v.a. X , también llamada **función de distribución** de X es la función $F_X(x)$, que asigna a cada evento definido sobre X una probabilidad dada por la siguiente expresión:

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

y de manera que se cumplan las siguientes tres condiciones:

1. $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$
2. Es continua por la derecha.
3. Es monótona no decreciente.

La distribución de probabilidad de una v.a. describe teóricamente la forma en que varían los resultados de un experimento aleatorio. Intuitivamente se trataría de una lista de los resultados posibles de un experimento con las probabilidades que se esperarían ver asociadas con cada resultado.

Función de densidad de una v.a. continua

La **función de densidad de probabilidad** (FDP) o, simplemente, función de densidad, representada comúnmente como $f(x)$, se utiliza con el propósito de conocer cómo se distribuyen las probabilidades de un suceso o evento, en relación al resultado del suceso.

La FDP es la derivada (ordinaria o en el sentido de las distribuciones) de la función de distribución de probabilidad $F(x)$, o de manera inversa, la función de distribución es la integral de la función de densidad:

$$F(x) = \int_{-\infty}^x f(t) dt$$

La función de densidad de una v.a. determina la concentración de probabilidad alrededor de los valores de una variable aleatoria continua.

Funciones de variables aleatorias

Sea una variable aleatoria X , sobre Ω y una función medible de Borel $g : \mathbb{R} \rightarrow \mathbb{R}$, entonces $Y = g(X)$ será también una variable aleatoria sobre Ω , dado que la composición de funciones medibles es también medible a no ser que g sea una función medible de Lebesgue. El mismo procedimiento que permite ir de un espacio de probabilidad (Ω, P) a (\mathbb{R}, dF_X) puede ser utilizado para obtener la distribución de Y . La función de probabilidad acumulada de Y es

$$F_Y(y) = P(g(X) \leq y).$$

Si la función g es invertible, es decir g^{-1} existe, y es monótona creciente, entonces la anterior relación puede ser extendida para obtener

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

y, trabajando de nuevo bajo las mismas hipótesis de invertibilidad de g y asumiendo además diferenciabilidad, podemos hallar la relación entre las funciones de densidad de probabilidad al diferenciar ambos términos respecto de y , obteniendo

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|.$$

Si g no es invertible pero cada y tiene un número finito de raíces, entonces la relación previa con la función de densidad de probabilidad puede generalizarse como

$$f_Y(y) = \sum_i f_X(g_i^{-1}(y)) \left| \frac{dg_i^{-1}(y)}{dy} \right|$$

donde $x_i = g_i^{-1}(y)$. Las fórmulas de densidad no requieren que g sea creciente.

Ejemplo

Sea X una variable aleatoria real continua y sea $Y = X^2$.

$$F_Y(y) = P(X^2 \leq y).$$

Si $y < 0$, entonces $P(X^2 = y) = 0$, por lo tanto

$$F_Y(y) = 0 \quad \text{si } y < 0.$$

Si $y = 0$, entonces

$$P(X^2 \leq y) = P(|X| \leq \sqrt{y}) = P(-\sqrt{y} \leq X \leq \sqrt{y}),$$

por lo tanto

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \quad \text{si } y \geq 0.$$

Parámetros de una v.a.

La función de densidad o la distribución de probabilidad de una v.a. contiene exhaustivamente toda la información sobre la variable. Sin embargo resulta conveniente resumir sus características principales con unos cuantos valores numéricos. Estos son, fundamentalmente la esperanza y la varianza.

Esperanza

La **esperanza matemática** (o simplemente **esperanza**) o **valor esperado** de una v.a. es la suma del producto de la probabilidad de cada suceso por el valor de dicho suceso.

Si todos los sucesos son de igual probabilidad la esperanza es la media aritmética.

Para una variable aleatoria discreta con valores posibles $x_1, x_2 \dots x_n$ y sus probabilidades representadas por la función de probabilidad $p(x_i)$ la esperanza se calcula como:

$$E[X] = \sum_{i=1}^n x_i p(x_i)$$

Para una variable aleatoria continua la esperanza se calcula mediante la integral de todos los valores y la función de densidad $f(x)$:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

$$\text{o } E[X] = \int_{\Omega} X dP$$

La esperanza también se suele simbolizar con $\mu = E[X]$

El concepto de esperanza se asocia comúnmente en los juegos de azar al de beneficio medio o beneficio esperado a largo plazo.

Varianza

La **varianza** es una medida de dispersión de una variable aleatoria X respecto a su esperanza $E[X]$. Se define como la esperanza de la transformación $(X - E[X])^2$:

$$\sigma = \sqrt{V(X)} \text{ o bien } \sigma^2 = V(X)$$

Véase también

- Distribución de probabilidad.
- Distribución binomial.
- Distribución normal.
- Esperanza matemática.
- Varianza.


Referencias

- [1] http://www.hrc.es/bioest/estadis_21.html Definición de variable aleatoria. Esta definición no es en absoluto rigurosa, ya que no define una variable aleatoria, sino cualquier función real. Es de remarcar que en la referencia no se dice en ningún momento que eso sea una definición. Sin embargo, en la mayoría de las aplicaciones prácticas, es suficiente.
- [2] La definición rigurosa de variable aleatoria exige dotar a \mathbb{R} de estructura de espacio medible e imponer a X la condición de ser función medible (véase la definición formal de variable aleatoria, en este mismo artículo).
- [3] <http://planetmath.org/encyclopedia/DiscreteRandomVariable.html>
- [4] <http://mathworld.wolfram.com/RandomVariable.html>
- [5] Véase conjunto finito para una definición más rigurosa
- [6] En experimentos reales la continuidad de una variable es rarísima, ya que la escasa precisión de los instrumentos de medida obliga a un conjunto discreto de valores posibles.

Bibliografía

- Peña Sánchez de Rivera, Daniel (2008). *Fundamentos de Estadística* (1ª edición). Alianza Editorial. pp. 688. ISBN 9788420683805.
- Ropero Moriones, Eva (2009). *Manual de estadística empresarial* (1ª edición). Delta Publicaciones. pp. 200. ISBN 9788492453214.

Enlaces externos

-  Wikimedia Commons alberga contenido multimedia sobre **Variable aleatoria** Commons.

Esperanza matemática

En estadística la **esperanza matemática** (también llamada **esperanza**, **valor esperado**, **media poblacional** o **media**) de una variable aleatoria X , es el número $E(X)$ que formaliza la idea de *valor medio* de un fenómeno aleatorio.

Cuando la variable aleatoria es discreta, la esperanza es igual a la suma de la probabilidad de cada posible suceso aleatorio multiplicado por el valor de dicho suceso. Por lo tanto, representa la cantidad media que se "espera" como resultado de un experimento aleatorio cuando la probabilidad de cada suceso se mantiene constante y el experimento se repite un elevado número de veces. Cabe decir que el valor que toma la esperanza matemática en algunos casos puede no ser "esperado" en el sentido más general de la palabra - el valor de la esperanza puede ser improbable o incluso imposible.

Por ejemplo, el valor esperado cuando tiramos un dado equilibrado de 6 caras es 3,5. Podemos hacer el cálculo

$$\begin{aligned} E(X) &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3,5 \end{aligned}$$

y cabe destacar que 3,5 no es un valor posible al rodar el dado. En este caso, en el que todos los sucesos son de igual probabilidad, la esperanza es igual a la media aritmética.

Una aplicación común de la esperanza matemática es en las apuestas o los juegos de azar. Por ejemplo, la ruleta americana tiene 38 casillas equiprobables. La ganancia para acertar una apuesta a un solo número paga de 35 a 1 (es decir, cobramos 35 veces lo que hemos apostado y recuperamos la apuesta, así que recibimos 36 veces lo que hemos apostado). Por tanto, considerando los 38 posibles resultados, la esperanza matemática del beneficio para apostar a un solo número es:

$$\left(-1 \cdot \frac{37}{38}\right) + \left(35 \cdot \frac{1}{38}\right),$$

que es -0,0526 aproximadamente. Por lo tanto uno esperaría, en media, perder unos 5 céntimos por cada euro que apuesta, y el valor esperado para apostar 1 euro son 0.9474 euros. En el mundo de las apuestas, un juego donde el beneficio esperado es cero (no ganamos ni perdemos) se llama un "juego justo".

Nota: El primer paréntesis es la "esperanza" de perder tu apuesta de \$1, por eso es negativo el valor. El segundo paréntesis es la esperanza matemática de ganar los \$35. La esperanza matemática del beneficio es el valor esperado a ganar menos el valor esperado a perder.

Definición

Para una variable aleatoria discreta con valores posibles $x_1, x_2 \dots x_n$ y sus probabilidades representadas por la función de probabilidad $p(x_i)$ la esperanza se calcula como:

$$E[X] = x_1 p(X = x_1) + \dots + x_n p(X = x_n) = E[X] = \sum_{i=1}^n x_i p(x_i)$$

Para una variable aleatoria continua la esperanza se calcula mediante la integral de todos los valores y la función de densidad $f(x)$:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

La definición general de esperanza se basa, como toda la teoría de la probabilidad, en el marco de la teoría de la medida y se define como la siguiente integral:

$$E[X] = \int_{\Omega} X \, dP$$

La esperanza también se suele simbolizar con $\mu = E[X]$

Las esperanzas $E[X^k]$ para $k = 0, 1, 2, \dots$ se llaman **momentos** de orden k . Más importantes son los momentos centrados $E[(X - E[X])^k]$.

No todas las variables aleatorias tienen un valor esperado. Por ejemplo, la distribución de Cauchy no lo tiene.

Propiedades

La esperanza es un operador lineal, ya que:

$$E(X + c) = E(X) + c$$

$$E(X + Y) = E(X) + E(Y)$$

$$E(aX) = aE(X)$$


Combinando estas propiedades, podemos ver que -

$$E(aX + b) = aE(X) + b$$

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

donde X e Y son variables aleatorias y a y b y c son tres constantes cualesquiera.

Véase también

- Varianza
-  Portal:Matemática. Contenido relacionado con **Matemática**.

Media aritmética

En matemáticas y estadística, la **media aritmética** (también llamada **promedio** o simplemente **media**) de un conjunto finito de números es igual a la suma de todos sus valores dividida entre el número de sumandos. Cuando el conjunto es una muestra aleatoria recibe el nombre de **media muestral** siendo uno de los principales estadísticos muestrales.

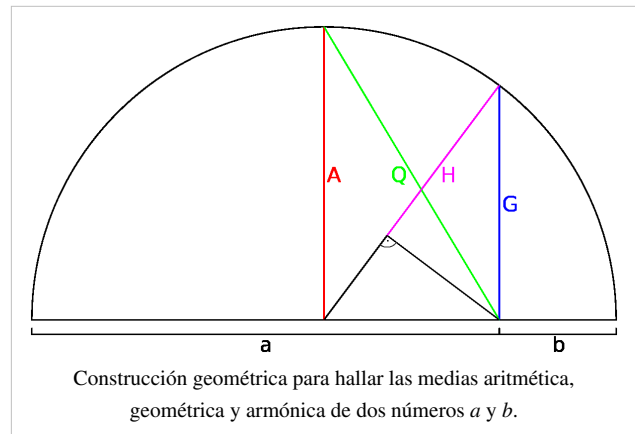
Expresada de forma más intuitiva, podemos decir que la media (aritmética) es la cantidad total de la variable distribuida a partes iguales entre cada observación.

Por ejemplo, si en una habitación hay tres personas, la

media de dinero que tienen en sus bolsillos sería el resultado de tomar todo el dinero de los tres y dividirlo a partes iguales entre cada uno de ellos. Es decir, la media es una forma de resumir la información de una distribución (dinero en el bolsillo) suponiendo que cada observación (persona) tuviera la misma cantidad de la variable.

También la media aritmética puede ser denominada como centro de gravedad de una distribución, el cual no está necesariamente en la mitad.

Una de las limitaciones de la media aritmética es que se trata de una medida muy sensible a los valores extremos; valores muy altos tienden a aumentarla mientras que valores muy bajos tienden a reducirla, lo que implica que puede dejar de ser representativa de la población.



Definición

Dados los n números $\{a_1, a_2, \dots, a_n\}$, la **media aritmética** se define simplemente como:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n a_i = \frac{a_1 + a_2 + \dots + a_n}{n}$$

Por ejemplo, la media aritmética de 8, 5 y -1 es igual a:

$$\bar{x} = \frac{8 + 5 + (-1)}{3} = 4$$

Se utiliza la letra X con una barra horizontal sobre el símbolo para representar la media de una muestra (\bar{X}), mientras que la letra μ (mu) se usa para la media aritmética de una población, es decir, el valor esperado de una variable.

En otras palabras, es la suma de n valores de la variable y luego dividido por n : donde n es el número de sumandos, o en el caso de estadísticas el número de datos.

Propiedades

- La media aritmética de un conjunto de números positivos siempre es igual o superior a la media geométrica:

$$\sqrt[n]{x_1 x_2 \dots x_n} \leq \frac{x_1 + \dots + x_n}{n}$$


- La media aritmética está comprendida entre el valor máximo y el valor mínimo del conjunto de datos:

$$\min\{x_1, x_2, \dots, x_n\} \leq \frac{x_1 + \dots + x_n}{n} \leq \max\{x_1, x_2, \dots, x_n\}$$

Véase también

- Medidas de tendencia central
- Curtosis
- Desviación estándar
- Esperanza matemática o Valor esperado
- Estadística descriptiva
- Ley de promedios
- Media, que es una medida de *tendencia central*.
- Media esférica
- Medianas
- Moda (estadística)
- Parámetro estadístico

Enlaces externos

-  Portal:Matemática. Contenido relacionado con **Matemática**.
- Weisstein, Eric W., «Media aritmética^[1]» (en inglés), *MathWorld*, Wolfram Research.
- Comparación entre la media aritmética y geométrica de dos números^[2] (en inglés)
- Diccionario Estadístico - Divestadística^[3] (en castellano)
- Simulación de la media con R-Project^[4]

Referencias

- [1] <http://mathworld.wolfram.com/ArithmeticMean.html>
- [2] <http://www.sengpielaudio.com/calculator-geommean.htm>
- [3] http://www.divestadistica.es/es/diccionario_estadistico.html#M
- [4] <http://cajael.com/mestadisticos/T1EDescriptiva/node4.php>

Mediana (estadística)

En el ámbito de la estadística, la **mediana**, representa el valor de la variable de posición central en un conjunto de datos ordenados. De acuerdo con esta definición el conjunto de datos menores o iguales que la mediana representarán el 50% de los datos, y los que sean mayores que la mediana representarán el otro 50% del total de datos de la muestra. La mediana coincide con el percentil 50, con el segundo cuartil y con el quinto decil.

Cálculo

Existen dos métodos para el cálculo de la mediana:

1. Considerando los datos en forma individual, sin agruparlos.
2. Utilizando los datos agrupados en intervalos de clase.

A continuación veamos cada una de ellas.

Datos sin agrupar

Sean $x_1, x_2, x_3, \dots, x_n$ los datos de una muestra *ordenada en orden creciente* y designando la mediana como M_e , distinguimos dos casos:

a) Si n es impar, la mediana es el valor que ocupa la posición $(n + 1)/2$ una vez que los datos han sido ordenados (en orden creciente o decreciente), porque éste es el valor central. Es decir: $M_e = x_{(n+1)/2}$.

Por ejemplo, si tenemos 5 datos, que ordenados son: $x_1 = 3, x_2 = 6, x_3 = 7, x_4 = 8, x_5 = 9$ => El valor central es el tercero: $x_{(5+1)/2} = x_3 = 7$. Este valor, que es la mediana de ese conjunto de datos, deja dos datos por debajo (x_1, x_2) y otros dos por encima de él (x_4, x_5).

b) Si n es par, la mediana es la media aritmética de las dos observaciones centrales. Cuando n es par, los dos datos que están en el centro de la muestra ocupan las posiciones $n/2$ y $n/2 + 1$. Es decir:

$$M_e = (x_{n/2} + (x_{n/2+1}))/2.$$

Por ejemplo, si tenemos 6 datos, que ordenados son: $x_1 = 3, x_2 = 6, x_3 = 7, x_4 = 8, x_5 = 9, x_6 = 10$

=> Hay dos valores que están por debajo del $x_{\frac{n}{2}} = x_3 = 7$ y otros dos que quedan por encima del siguiente dato

$x_{\frac{n}{2}+1} = x_4 = 8$. Por tanto, la mediana de este grupo de datos es la media aritmética de estos dos datos:

$$M_e = \frac{x_3 + x_4}{2} = \frac{7 + 8}{2} = 7,5.$$

Datos agrupados

Al tratar con datos agrupados, si $\frac{n}{2}$ coincide con el valor de una frecuencia acumulada, el valor de la mediana coincidirá con la abscisa correspondiente. Si no coincide con el valor de ninguna abscisa, se calcula a través de semejanza de triángulos en el histograma o polígono de frecuencias acumuladas, utilizando la siguiente equivalencia:

$$\frac{N_i - N_{i-1}}{a_i - a_{i-1}} = \frac{\frac{n}{2} - N_{i-1}}{p} \Rightarrow p = \frac{\frac{n}{2} - N_{i-1}}{N_i - N_{i-1}} (a_i - a_{i-1})$$

Dónde N_i y N_{i-1} son las frecuencias absolutas acumuladas tales que $N_{i-1} < \frac{n}{2} < N_i$, a_{i-1} y a_i son los extremos, inferior y superior, del intervalo donde se alcanza la mediana y $M_e = a_{i-1}$ es la abscisa a calcular, la moda. Se observa que $a_i - a_{i-1}$ es la amplitud de los intervalos seleccionados para el diagrama.

Ejemplos para datos sin agrupar

Ejemplo 1: Cantidad (N) impar de datos

xi	fi	Ni
1	2	2
2	2	4
3	4	8
4	5	13
5	8	21 > 19.5
6	9	30
7	3	33
8	4	37
9	2	39

Las calificaciones en la asignatura de Matemáticas de 39 alumnos de una clase viene dada por la siguiente tabla:

Calificaciones	1	2	3	4	5	6	7	8	9
Número de alumnos	2	2	4	5	8	9	3	4	2

Primero se hallan las frecuencias absolutas acumuladas N_i . Así, aplicando la fórmula asociada a la mediana para n impar, se obtiene $X(39 + 1)/2 = X20$.

- $N_{i-1} < n/2 < N_i = N_{19} < 19.5 < N_{20}$

Por tanto la mediana será el valor de la variable que ocupe el vigésimo lugar. En este ejemplo, 21 (frecuencia absoluta acumulada para $X_i = 5$) > 19.5 con lo que $M_e = 5$ puntos, la mitad de la clase ha obtenido un 5 o menos, y la otra mitad un 5 o más.

Ejemplo 2 : Cantidad (N) par de datos

Las calificaciones en la asignatura de Matemáticas de 38 alumnos de una clase viene dada por la siguiente tabla (debajo):

Calificaciones	1	2	3	4	5	6	7	8	9
Número de alumnos	2	2	4	5	6	9	4	4	2

xi	fi	Ni+w
1	2	2
2	2	4
3	4	8
4	5	13
5	6	19 = 19
6	9	28
7	4	32
8	4	36
9	2	38

Primero se hallan las frecuencias absolutas acumuladas N_i . Así, aplicando la fórmula asociada a la mediana para n par, se obtiene $X(38/2) = X19$.

- $N_{i-1} < n/2 < N_i = N_{18} < 19 < N_{19}$

Con lo cual la mediana será la media aritmética de los valores de la variable que ocupen el decimonoveno y el vigésimo lugar. En el ejemplo el lugar decimonoveno lo ocupa el 5 y el vigésimo el 6 con lo que $Me = (5+6)/2 = 5,5$ puntos, la mitad de la clase ha obtenido un 5,5 o menos y la otra mitad un 5,5 o más.

Ejemplo para datos agrupados

Entre 1.70 y 1.80 hay 3 estudiantes.

Entre 1.60 y 1.70 hay 5 estudiantes.

Entre 1.50 y 1.60 hay 2 estudiantes.

$$Mediana = 1.60 + \left(\frac{(10/2) - 3}{5} \right) 0.1 = 1.64$$

Método de cálculo general

x_i	f_i	N_i
$[x_{i1}-x_{i2}]$	f_1	N_1
.	.	.
.	.	.
.	.	$N_{(i-2)}$
$[x_{(i-1)1}-x_{(i-1)2}]$	$f_{(i-1)}$	$f_{(i-1)}-N_{(i-2)}=N_{(i-1)}$
$[x_{i1}-x_{i2}]$	f_i	$f_i-N_{i-1}=N_i$
$[x_{(i+1)1}-x_{(i+1)2}]$	$f_{(i+1)}$	$f_{(i+1)}-N_i=N_{(i+1)}$
.	.	.
.	.	.
.	.	.
$[x_{M1}-x_{M2}]$	f_M	$f_M-N_{(M-1)}=N_M$

Consideramos:

- x_{i1} valor mínimo < Entonces:

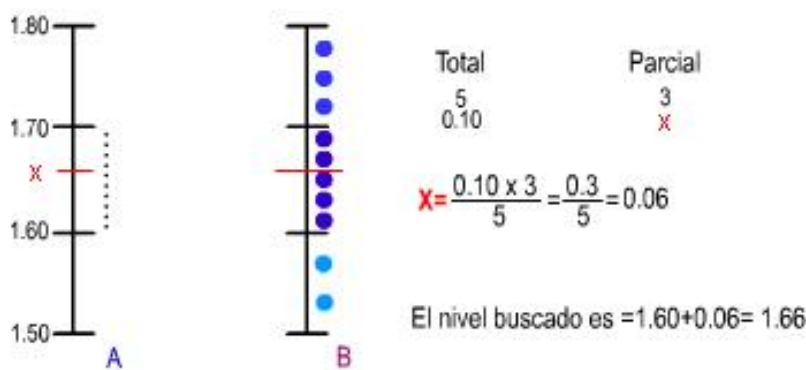
$$Mediana = x_{i1} + \left(\frac{(N_M/2) - N_{i-1}}{f_i} \right) \cdot (x_{i2} - x_{i1})$$

Método proyectivo

Con base en el método proyectivo, se puede obtener la mediana para datos agrupados de la siguiente forma:

1. Tomar el número total de frecuencias y dividirlo entre dos.
2. Restar a ese número el total de frecuencias de las clases anteriores a la clase mediana.
3. Usar el número obtenido para hacer un cambio del doble superior de escala entre las frecuencias de la clase mediana y sus rangos para obtener la distancia parcial
4. Sumamos la distancia parcial obtenida a el límite inferior de la clase.

Usando el ejemplo anterior:



1. El número total de frecuencias es de; $(3+5+2)/2 = 10/2 = 5$
2. El total de frecuencias anteriores es 2; $(5 - 2) = 3$
3. Hacemos el cambio de escalas:

$$3 : 5 :: x : 0.10$$

Resolviendo:

$$x = \frac{(0.10)(3)}{5} = 0.06 \text{ la mediana es la suma de todos los datos dividido entre el número de datos}$$

4. Se suma la distancia parcial al límite inferior:

$$\text{Mediana} = 0.06 + 1.60 = 1.66$$

Véase también

- Desviación estándar
- Frecuencia
- Moda (estadística)
- Media aritmética
- Parámetro estadístico
- Promedio
- Valor esperado

Moda (estadística)

En estadística, la **moda** es el valor con una mayor frecuencia en una distribución de datos.

$$Md = x_i, \text{ Si } n_i = \max \{ f_j, j \in \{1, 2, \dots, k\} \}$$

Hablaremos de una distribución bimodal de los datos adquiridos en una columna cuando encontremos dos modas, es decir, dos datos que tengan la misma frecuencia absoluta máxima. Una distribución trimodal de los datos es en la que encontramos tres modas. Si todas las variables tienen la misma frecuencia diremos que no hay moda.

El **intervalo modal** es el de mayor frecuencia absoluta. Cuando tratamos con datos agrupados antes de definir la moda, se ha de definir el intervalo modal.

La moda, cuando los datos están agrupados, es un punto que divide al intervalo modal en dos partes de la forma p y

$$\frac{p}{c-p} = \frac{n_i - n_{i-1}}{n_i - n_{i+1}}$$

c-p, siendo c la amplitud del intervalo, que verifiquen que:

Siendo la frecuencia absoluta del intervalo modal las frecuencias absolutas de los intervalos anterior y posterior,
 $y n_{i-1} y n_{i+1}$

respectivamente, al intervalo modal.

Moda de datos agrupados

Para obtener la moda en datos agrupados se usa la siguiente fórmula:

$$\text{Moda} = L_{i-1} + \left(\frac{D_1}{D_1 + D_2} \right) i$$

Donde:

L_{i-1} = Límite inferior de la clase modal.

D_1 = es el delta de frecuencia absoluta modal y la frecuencia absoluta premodal.

D_2 = es el delta de frecuencia absoluta modal y la frecuencia absoluta postmodal.

i = intervalo.

Ejemplo

Encontrar la estatura modal de un grupo que se encuentra distribuido de la siguiente forma:

- Entre 1 y 1.10 hay 1 estudiante
- Entre 1.10 y 1.15 hay 1,5 estudiantes
- Entre 1.20 y 1.25 hay 2 estudiantes
- Entre 1.30 y 1.35 hay 2,3 estudiantes.
- Entre 1.45 y 1.55 hay 3 estudiantes.
- Entre 1.50 y 1.60 hay 4 estudiantes.
- Entre 1.60 y 1.70 hay 10 estudiantes.
- Entre 1.70 y 1.80 hay 8 estudiantes.

Clase modal = 1.60 y 1.70 (es la que tiene frecuencia absoluta más alta, 10)

$L_{i-1} = 1.60$ $D_1 = 6$ $D_2 = 2$ $i = 0.10$

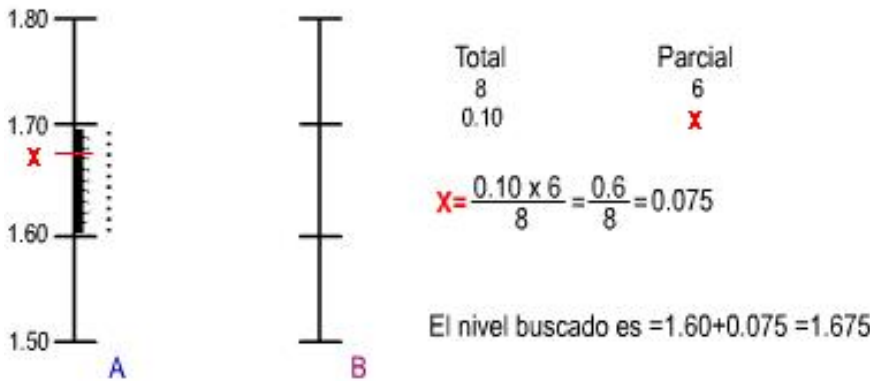
$Moda = 1.60 + (6/8) * 0.1 = 1.675$

Método proyectivo

Con base en el Método Proyectivo se obtiene la moda de la siguiente manera usando el ejemplo anterior:

- 1.- Se Identifica la clase modal, que es la clase que tiene más frecuencias.
- 2.- Se identifica las diferencias con las clases vecinas.
- 3.- Se hace un cambio de escala

En el Ejemplo anterior:



1.- Clase con más frecuencias:

1.60 a 1.70 (con 10 frecuencias)

2.- Diferencias con las clases vecinas: **2** (clase superior) y **6** (clase inferior) que se obtiene de restar (10-8) y (10-4)

3.-Cambio de escala:

Distancia parcial en la escala uno es a la distancia total de la misma escala como el valor buscado es a la distancia total de la escala dos.

$$6 : 8 :: x : 0.10$$

Resolviendo:

$$x = \frac{(0.10)(6)}{8} = 0.075$$

Se suma 0.075 (la distancia parcial) a 1.60 (el límite inferior), obteniéndose la moda.

$$Moda = 0.075 + 1.60 = 1.675$$

Propiedades

Sus principales propiedades son:

- Cálculo sencillo.
- Interpretación muy clara.
- Al depender sólo de las frecuencias, puede calcularse para variables cualitativas. Es por ello el parámetro más utilizado cuando al resumir una población no es posible realizar otros cálculos, por ejemplo, cuando se enumeran en medios periodísticos las características más frecuentes de determinado sector social. Esto se conoce informalmente como "retrato robot".^[1]

Inconvenientes

- Su valor es independiente de la mayor parte de los datos, lo que la hace muy sensible a variaciones muestrales. Por otra parte, en variables agrupadas en intervalos, su valor depende excesivamente del número de intervalos y de su amplitud.
- Usa muy pocas observaciones, de tal modo que grandes variaciones en los datos fuera de la moda, no afectan en modo alguno a su valor.
- No siempre se sitúa hacia el centro de la distribución.
- Puede haber más de una moda en el caso en que dos o más valores de la variable presenten la misma frecuencia (distribuciones bimodales o multimodales).

Referencias

- [1] Santos, María José (abril 2009). «Retrato robot del alcalde metropolitano (<http://www.correoandalucia.com/noticia.asp?idnoticia=4424170096095100100092424170>)». *El Correo de Andalucía*. Consultado el 07-04-2009.

Véase también

- Frecuencia
- Media (intervalo)
- Mediana (estadística)
- Parámetro estadístico
- Valor esperado

Varianza

En teoría de probabilidad, la **varianza** (que suele representarse como σ^2) de una variable aleatoria es una medida de su dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media.

Está medida en unidades distintas de las de la variable. Por ejemplo, si la variable mide una distancia en metros, la varianza se expresa en metros al cuadrado. La desviación estándar, la raíz cuadrada de la varianza, es una medida de dispersión alternativa expresada en las mismas unidades. La varianza tiene como valor mínimo 0.

Hay que tener en cuenta que la varianza puede verse muy influida por los valores atípicos y no se aconseja su uso cuando las distribuciones de las variables aleatorias tienen colas pesadas. En tales casos se recomienda el uso de otras medidas de dispersión más robustas.

El término *varianza* fue acuñado por Ronald Fisher en un artículo de 1918 titulado *The Correlation Between Relatives on the Supposition of Mendelian Inheritance*.

Definición

Dada una variable aleatoria X con media $\mu = E(X)$, se define su **varianza**, $\text{Var}(X)$ (también representada como σ_X^2 o, simplemente σ^2), como

$$\text{Var}(X) = E[(X - \mu)^2].$$

Desarrollando la definición anterior, se obtiene la siguiente definición alternativa (y equivalente):

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= E[(X^2 - 2X\mu + \mu^2)] \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2. \end{aligned}$$

Si una distribución no tiene esperanza, como ocurre con la de Cauchy, tampoco tiene varianza. Existen otras distribuciones que, aun teniendo esperanza, carecen de varianza. Un ejemplo de ellas es la de Pareto cuando su índice k satisface $1 < k \leq 2$.

Caso continuo

Si la variable aleatoria X es continua con función de densidad $f(x)$, entonces

$$\text{Var}(X) = \int (x - \mu)^2 f(x) dx,$$

donde

$$\mu = \int x f(x) dx,$$

y las integrales están definidas sobre el rango de X .

Caso discreto

Si la variable aleatoria X es discreta con pesos $x_1 \mapsto p_1, \dots, x_n \mapsto p_n$, entonces

$$\text{Var}(X) = \sum_{i=1}^n p_i \cdot (x_i - \mu)^2$$

donde

$$\mu = \sum_{i=1}^n p_i \cdot x_i.$$

Ejemplos

Distribución exponencial

La distribución exponencial de parámetro λ es una distribución continua con soporte en el intervalo $[0, \infty)$ y función de densidad

$$f(x) = \lambda e^{-\lambda x} 1_{[0, \infty)}(x),$$

Tiene media $\mu = \lambda^{-1}$. Por lo tanto, su varianza es:

$$\int_0^{\infty} f(x)(x - \mu)^2 dx = \int_0^{\infty} \lambda e^{-\lambda x} (x - \lambda^{-1})^2 dx = \lambda^{-2}.$$

Es decir, $\sigma^2 = \mu^2$.

Dado perfecto

Un dado de seis caras puede representarse como una variable aleatoria discreta que toma, valores del 1 al 6 con probabilidad igual a $\frac{1}{6}$. El valor esperado es $(1+2+3+4+5+6)/6 = 3.5$. Por lo tanto, su varianza es:

$$\sum_{i=1}^6 \frac{1}{6} (i - 3.5)^2 = \frac{1}{6} ((-2.5)^2 + (-1.5)^2 + (-0.5)^2 + 0.5^2 + 1.5^2 + 2.5^2) = \frac{1}{6} \cdot 17.50 = \frac{35}{12} \approx 2.92.$$

Propiedades de la varianza

Algunas propiedades de la varianza son:

- $V(X) \geq 0$
- $V(aX + b) = a^2 V(X)$ siendo a y b números reales cualesquiera. De esta propiedad se deduce que la varianza de una constante es cero, es decir, $V(b) = 0$
- $V(X + Y) = V(X) + V(Y) + 2\text{Cov}(X, Y)$, donde $\text{Cov}(X, Y)$ es la covarianza de X e Y .
- $V(X - Y) = V(X) + V(Y) - 2\text{Cov}(X, Y)$, donde $\text{Cov}(X, Y)$ es la covarianza de X e Y .

Varianza muestral

En muchas situaciones es preciso estimar la varianza de una población a partir de una muestra. Si se toma una muestra con reemplazamiento (y_1, \dots, y_n) de n valores de ella, de entre todos los estimadores posibles de la varianza de la población de partida, existen dos de uso corriente:

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \left(\frac{1}{n} \sum_{i=1}^n y_i^2 \right) - \bar{y}^2$$

y

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \sum_{i=1}^n y_i^2 - \frac{n}{n-1} \bar{y}^2 = \frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2}{n-1}$$

Cuando los datos están agrupados:

$$s^2 = \frac{\sum_{i=1}^n f_i (y_i - \bar{y})^2}{n - 1} = \frac{\sum_{i=1}^n f_i y_i^2 - n\bar{y}^2}{n - 1}$$

A los dos (cuando está dividido por n y cuando lo está por $n-1$) se los denomina **varianza muestral**. Difieren ligeramente y, para valores grandes de n , la diferencia es irrelevante. El primero traslada directamente la varianza de la muestra al de la población y el segundo es un estimador insesgado de la varianza de la población. De hecho,

$$\begin{aligned} E[s^2] &= E \left[\frac{1}{n-1} \sum_{i=1}^n Y_i^2 - \frac{n}{n-1} \bar{Y}^2 \right] \\ &= \frac{1}{n-1} \left(\sum E[Y_i^2] - nE[\bar{Y}^2] \right) \\ &= \frac{1}{n-1} \left(nE[Y_1^2] - nE[\bar{Y}^2] \right) \\ &= \frac{n}{n-1} \left(\text{Var}(Y_1) + E[Y_1]^2 - \text{Var}(\bar{Y}) - E[\bar{Y}]^2 \right) \\ &= \frac{n}{n-1} \left(\text{Var}(Y_1) + \mu^2 - \frac{1}{n} \text{Var}(Y_1) - \mu^2 \right) \\ &= \frac{n}{n-1} \left(\frac{n-1}{n} \text{Var}(Y_1) \right) \\ &= \text{Var}(Y_1) \\ &= \sigma^2 \end{aligned}$$

mientras que

$$E[s_n^2] = \frac{n-1}{n} \sigma^2$$

Propiedades de la varianza muestral

Como consecuencia de la igualdad $E(s^2) = \sigma^2$, s^2 es un estadístico insesgado de σ^2 . Además, si se cumplen las condiciones necesarias para la ley de los grandes números, s^2 es un estimador consistente de σ^2 .

Más aún, cuando las muestras siguen una distribución normal, por el teorema de Cochran, s^2 tiene la distribución chi-cuadrado:

$$(n-1) \frac{s^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Enlaces externos

- [1] Simulación de la varianza de una variable discreta con R (lenguaje de programación)

Véase también

- Desviación típica o desviación estándar
- Esperanza matemática o valor esperado
- Covarianza
- Análisis de varianza

Referencias

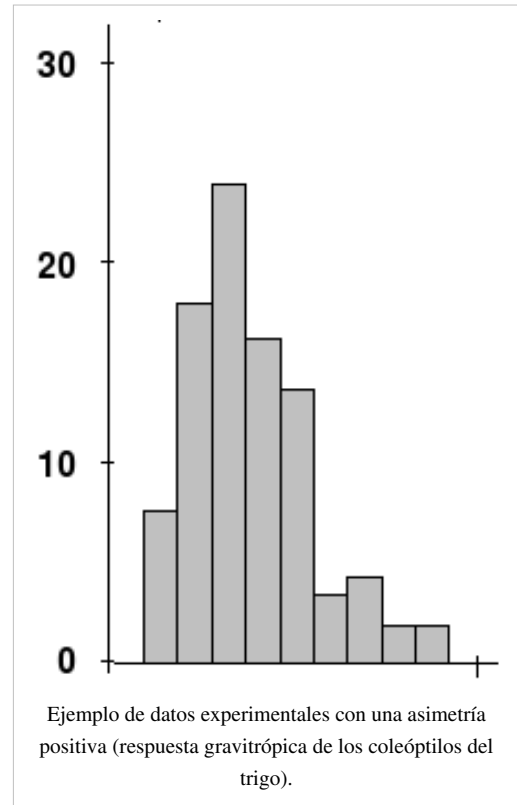
[1] <http://cajael.com/mestadisticos/T1EDescriptiva/node6.php>

Asimetría estadística

Definición

Las medidas de **asimetría** son indicadores que permiten establecer el *grado de simetría* (o asimetría) que presenta una distribución de probabilidad de una variable aleatoria sin tener que hacer su representación gráfica.

Como eje de simetría consideramos una recta paralela al eje de ordenadas que pasa por la media de la distribución. Si una distribución es simétrica, existe el mismo número de valores a la derecha que a la izquierda de la media, por tanto, el mismo número de desviaciones con signo positivo que con signo negativo. Decimos que hay asimetría positiva (o a la derecha) si la "cola" a la derecha de la media es más larga que la de la izquierda, es decir, si hay valores más separados de la media a la derecha. Diremos que hay asimetría negativa (o a la izquierda) si la "cola" a la izquierda de la media es más larga que la de la derecha, es decir, si hay valores más separados de la media a la izquierda.



Medidas de asimetría

Coefficiente de asimetría de Fisher

En teoría de la probabilidad y estadística, la medida de asimetría más utilizada parte del uso del tercer momento estándar. La razón de esto es que nos interesa mantener el signo de las desviaciones con respecto a la media, para obtener si son mayores las que ocurren a la derecha de la media que las de la izquierda. Sin embargo, no es buena idea tomar el momento estándar con respecto a la media de orden 1 (¡Ya que una simple suma de todas las desviaciones siempre es cero!). Por ello, lo más sencillo es tomar las desviaciones al cubo.

El **coeficiente de asimetría de Fisher**, representado por γ_1 , se define como:

$$\gamma_1 = \frac{\mu_3}{\sigma^3},$$

donde μ_3 es el tercer momento en torno a la media y σ es la desviación estándar.

Si $\gamma_1 = 0$, la distribución es simétrica.

Si $\gamma_1 > 0$, la distribución es asimétrica positiva o a la derecha.

Si $\gamma_1 < 0$, la distribución es asimétrica negativa o a la izquierda.

Coefficiente de asimetría de Pearson

Sólo se puede utilizar en distribuciones uniformes, unimodales y moderadamente asimétricas. Se basa en que en distribuciones simétricas la media de la distribución es igual a la moda.

$$A_p = \frac{\mu - moda}{\sigma},$$

Si la distribución es simétrica, $\mu = moda$ y $A_p = 0$. Si la distribución es asimétrica positiva la media se sitúa por encima de la moda y, por tanto, $A_p > 0$.

Coefficiente de asimetría de Bowley

Está basado en la posición de los cuartiles y la mediana, y utiliza la siguiente expresión:

$$A_B = \frac{Q_{3/4} + Q_{1/4} - 2Me}{Q_{3/4} - Q_{1/4}}$$

En una distribución simétrica el tercer cuartil estará a la misma distancia de la mediana que el primer cuartil. Por tanto $A_B = 0$.

Si la distribución es positiva o a la derecha, $A_B > 0$.

Utilidad

La asimetría resulta útil en muchos campos. Muchos modelos simplistas asumen una distribución normal, esto es, simétrica en torno a la media. La distribución normal tiene una asimetría cero. Pero en realidad, los valores no son nunca perfectamente simétricos y la asimetría de la distribución proporciona una idea sobre si las desviaciones de la media son positivas o negativas. Una asimetría positiva implica que hay más valores distintos a la derecha de la media.

Las medidas de asimetría, sobre todo el coeficiente de asimetría de Fisher, junto con las medidas de apuntamiento o curtosis se utilizan para contrastar si se puede aceptar que una distribución estadística sigue la distribución normal. Esto es necesario para realizar numerosos contrastes estadísticos en la teoría de inferencia estadística.

Referencias

- 'Introducción a la Estadística Económica y Empresarial. Teoría y Práctica.' de Fco. Javier Martín-Pliego López, Editorial Thomson, 2007 (Madrid).
- 'Manual de Estadística Empresarial con ejercicios resueltos' de Eva Roperó, María Eleftheriou, Luana Gava y Eva Romero. Editorial Delta Publicaciones. 2008 (Madrid).

Enlaces externos

- An Asymmetry Coefficient for Multivariate Distributions (<http://petitjeanmichel.free.fr/itoweb.petitjean.skewness.html>) by Michel Petitjean

Desviación estándar

La **desviación estándar** o **desviación típica** (σ) es una medida de centralización o dispersión para variables de razón (ratio o cociente) y de intervalo, de gran utilidad en la estadística descriptiva.

Se define como la raíz cuadrada de la varianza. Junto con este valor, la desviación típica es una medida (cuadrática) que informa de la media de distancias que tienen los datos respecto de su media aritmética, expresada en las mismas unidades que la variable.

Para conocer con detalle un conjunto de datos, no basta con conocer las medidas de tendencia central, sino que necesitamos conocer también la desviación que presentan los datos en su distribución respecto de la media aritmética de dicha distribución, con objeto de tener una visión de los mismos más acorde con la realidad al momento de describirlos e interpretarlos para la toma de decisiones.

Formulación

La varianza representa la media aritmética de las desviaciones con respecto a la media que son elevadas al cuadrado.

Si atendemos a la colección completa de datos (la población en su totalidad) obtenemos la varianza poblacional; y si por el contrario prestamos atención sólo a una muestra de la población, obtenemos en su lugar la varianza muestral. Las expresiones de estas medidas son las que aparecen a continuación.

Expresión de la varianza muestral:

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

Segunda forma de calcular la varianza muestral:

$$S_X^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2$$

demonstración

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{\sum_{i=1}^n (X_i^2 + \bar{X}^2 - 2X_i\bar{X})}{n} = \frac{\sum_{i=1}^n X_i^2}{n} + \frac{\sum_{i=1}^n \bar{X}^2}{n} - \frac{\sum_{i=1}^n 2X_i\bar{X}}{n} = \frac{\sum_{i=1}^n X_i^2}{n} + \frac{\bar{X}^2 \sum_{i=1}^n 1}{n} - \frac{2\bar{X} \sum_{i=1}^n X_i}{n}$$

podemos observar que como

$$\frac{\sum_{i=1}^n 1}{n} = 1 \quad (\text{sumamos } n \text{ veces } 1 \text{ y luego dividimos por } n)$$

y como

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

obtenemos

$$\frac{\sum_{i=1}^n X_i^2}{n} + \bar{X}^2 - 2\bar{X}\bar{X} = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2$$

Expresión de la cuasivarianza muestral (estimador insesgado de la varianza poblacional):

$$S_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

Expresión de la varianza poblacional:

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

donde μ es el valor medio de $\{ X_i \}$

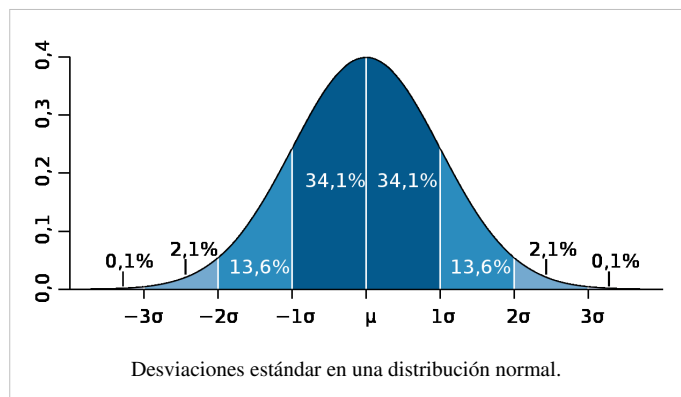
Expresión de la desviación estándar poblacional:

$$\sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$

El término *desviación estándar* fue incorporado a la estadística por Karl Pearson en 1894.

Por la formulación de la varianza podemos pasar a obtener la desviación estándar, tomando la raíz cuadrada positiva de la varianza. Así, si efectuamos la raíz de la varianza muestral, obtenemos la desviación típica muestral; y si por el contrario, efectuamos la raíz sobre la varianza poblacional, obtendremos la desviación típica poblacional.

Expresión de la desviación estándar muestral:



$$\sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

También puede ser tomada como

$$S = \sqrt{\frac{a - s^2/n}{n - 1}}$$

con a como $\sum_{i=1}^n x_i^2$ y s como $\sum_{i=1}^n x_i$ además se puede tener una mejor tendencia de medida al desarrollar las formulas indicadas pero se tiene que tener en cuenta la media, mediana y moda

Interpretación y aplicación

La desviación estándar es una medida del grado de dispersión de los datos con respecto al valor promedio. Dicho de otra manera, la desviación estándar es simplemente el "promedio" o variación esperada con respecto a la media aritmética.

Por ejemplo, las tres muestras (0, 0, 14, 14), (0, 6, 8, 14) y (6, 6, 8, 8) cada una tiene una media de 7. Sus desviaciones estándar muestrales son 8,08, 5,77 y 1,15, respectivamente. La tercera muestra tiene una desviación mucho menor que las otras dos porque sus valores están más cerca de 7.

La desviación estándar puede ser interpretada como una medida de incertidumbre. La desviación estándar de un grupo repetido de medidas nos da la precisión de éstas. Cuando se va a determinar si un grupo de medidas está de acuerdo con el modelo teórico, la desviación estándar de esas medidas es de vital importancia: si la media de las medidas está demasiado alejada de la predicción (con la distancia medida en desviaciones estándar), entonces consideramos que las medidas contradicen la teoría. Esto es coherente, ya que las mediciones caen fuera del rango de

valores en el cual sería razonable esperar que ocurrieran si el modelo teórico fuera correcto. La desviación estándar es uno de tres parámetros de ubicación central; muestra la agrupación de los datos alrededor de un valor central (la media o promedio).

Desglose

La **desviación estándar** (DS/DE), también llamada **desviación típica**, es una medida de dispersión usada en estadística que nos dice cuánto tienden a alejarse los valores concretos del promedio en una distribución. De hecho, específicamente, la desviación estándar es "el promedio de la distancia de cada punto respecto del promedio". Se suele representar por una **S** o con la letra sigma, σ .

La desviación estándar de un conjunto de datos es una medida de cuánto se desvían los datos de su media. Esta medida es más estable que el recorrido y toma en consideración el valor de cada dato.

Es posible calcular la desviación estándar de una variable aleatoria continua como la raíz cuadrada de la integral

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

donde

$$\mu = \int_{-\infty}^{\infty} x f(x) dx$$

- La DS es la raíz cuadrada de la varianza de la distribución

$$\sigma^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Así la varianza es la media de los cuadrados de las diferencias entre cada valor de la variable y la media aritmética de la distribución.

Aunque esta fórmula es correcta, en la práctica interesa realizar inferencias poblacionales, por lo que en el denominador en vez de n , se usa $n-1$ (Corrección de Bessel)

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

También hay otra función más sencilla de realizar y con menos riesgo de tener equivocaciones :

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n - 1}$$

Ejemplo

Aquí se muestra cómo calcular la desviación estándar de un conjunto de datos. Los datos representan la edad de los miembros de un grupo de niños. { 4, 1, 11, 13, 2, 7 }

1. Calcular el promedio o media aritmética \bar{x} .

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

En este caso, $N = 6$ porque hay seis datos:

$$x_1 = 4$$

$$x_2 = 1$$

$$x_3 = 11$$

$$x_4 = 13$$

$$x_5 = 2$$

$$x_6 = 7$$

i=número de datos para sacar desviación estándar

$$\bar{x} = \frac{1}{6} \sum_{i=1}^6 x_i \quad \text{Sustituyendo } N \text{ por } 6$$

$$\bar{x} = \frac{1}{6} (x_1 + x_2 + x_3 + x_4 + x_5 + x_6)$$

$$\bar{x} = \frac{1}{6} (4 + 1 + 11 + 13 + 2 + 7)$$

$$\bar{x} = 6,33 \quad \text{Este es el promedio.}$$

2. Calcular la desviación estándar σ

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

$$\sigma = \sqrt{\frac{1}{5} \sum_{i=1}^6 (x_i - \bar{x})^2} \quad \text{Sustituyendo } N-1 \text{ por } 5; (6-1)$$

$$\sigma = \sqrt{\frac{1}{5} \sum_{i=1}^6 (x_i - 6,33)^2} \quad \text{Sustituyendo } \bar{x} \text{ por } 6,33$$

$$\sigma = \sqrt{\frac{1}{5} [(4 - 6,33)^2 + (1 - 6,33)^2 + (11 - 6,33)^2 + (13 - 6,33)^2 + (2 - 6,33)^2 + (7 - 6,33)^2]}$$

$$\sigma = \sqrt{\frac{1}{5} [(-2,33)^2 + (-5,33)^2 + 4,67^2 + 6,67^2 + (-4,33)^2 + 0,67^2]}$$

$$\sigma = \sqrt{\frac{1}{5} (5,43 + 28,4 + 21,8 + 44,5 + 18,7 + 0,449)}$$

$$\sigma = \sqrt{\frac{119,28}{5}}$$


$$\sigma = \sqrt{23,86}$$

$$\sigma = 4,88 \quad \text{Éste es el valor de la desviación estándar.}$$

Enlaces externos

- [1] Simulación de la desviación típica de una variable discreta con R (lenguaje de programación)

Véase también

- Parámetro estadístico
-  Wikimedia Commons alberga contenido multimedia sobre **Desviación estándar** Commons.

Referencias

- [1] <http://cajael.com/mestadisticos/T1EDescriptiva/node7.php>

Covarianza

En estadística la **covarianza** es una medida de dispersión conjunta de dos variables estadísticas.

Definición

La covarianza S_{XY} (a veces también denotada $Cov(X, Y)$) de dos variables aleatorias X e Y es:

$$S_{XY} = E([X - E(X)][Y - E(Y)]),$$

donde $E(\cdot)$ es el operador esperanza. Para distribuciones discretas la fórmula anterior se concreta en

$$S_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_{i=1}^n x_i * y_i - \bar{x}\bar{y}.$$

Cuando las variables aleatorias X e Y son n -dimensionales, es decir, $X = (X_1, \dots, X_n)^t$ e $Y = (Y_1, \dots, Y_n)^t$, su matriz de covarianzas Σ_{XY} es:

$$\Sigma_{XY} = E([X - E(X)][Y - E(Y)]^t)$$

Interpretación de la covarianza

- Si $S_{xy} > 0$ hay dependencia directa (positiva), es decir, a grandes valores de x corresponden grandes valores de y .
- Si $S_{xy} = 0$ Una covarianza 0 se interpreta como la no existencia de una relación lineal entre las dos variables estudiadas.
- Si $S_{xy} < 0$ hay dependencia inversa o negativa, es decir, a grandes valores de x corresponden pequeños valores de y .

Propiedades

Si X, Y, W , y V son variables aleatorias y a, b, c, d son constantes ("constante" en este contexto significa no aleatorio), se cumple que:

- $Cov(X, a) = 0$
- $Cov(X, X) = Var(X)$, la varianza de X
- $Cov(X, Y) = Cov(Y, X)$
- $Cov(aX, bY) = ab Cov(X, Y)$
- $Cov(X + a, Y + b) = Cov(X, Y)$
- $Cov(aX + bY, cW + dV) = ac Cov(X, W) + ad Cov(X, V) + bc Cov(Y, W) + bd Cov(Y, V)$
- $Cov(X, Y) = E(XY) - E(X)E(Y)$, fórmula que suele emplearse en la práctica para calcular la covarianza.

Estas propiedades se deducen de manera casi directa de la definición de la covarianza. En otras palabras la covarianza trata de explicar que tan relacionadas se encuentran dos variables entre sí, que tanto se mueve una cuando la otra se mueve otro tanto. Ejemplo, si la variable X se mueve 1, supongamos que la variable Y se mueve 2, entonces podemos decir que la variable Y se mueve positivamente el doble de lo que se movería la variable X .

No correlación e independencia

Si X e Y son independientes, entonces su covarianza es cero. Esto ocurre por la propiedad de independencia,

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

Lo opuesto, sin embargo, generalmente no es cierto: algunos pares de variables aleatorias tienen covarianza cero pese a que no son independientes. Bajo algunas hipótesis adicionales, la covarianza de valor cero implica independencia, como por ejemplo en el caso de la distribución normal multivariante.

Relación con el producto escalar

La mayoría de las propiedades de la covarianza se deducen de las del producto escalar:

1. Bilinealidad: para las constantes a y b , y las variables aleatorias X , Y , y U , $\text{Cov}(aX + bY, U) = a \text{Cov}(X, U) + b \text{Cov}(Y, U)$
2. Simetría: $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
3. Es un operador *positivo definido*: $\text{Var}(X) = \text{Cov}(X, X) \geq 0$; además, si $\text{Cov}(X, X) = 0$ entonces X es una variable aleatoria constante.

De hecho, la covarianza es un producto interior sobre el espacio cociente de las variables aleatorias de momentos finitos iguales salvo constante.

Enlaces Externos

- Simulación de la covarianza de una variable bidimensional continua [1] y discreta [2] con R (lenguaje de programación)

Véase también

- Regresión lineal
- Correlación
- ANOVA
- Varianza

Referencias

[1] <http://cajael.com/mestadisticos/T4DProbabilidad/node7.php>

[2] <http://cajael.com/mestadisticos/T4DProbabilidad/node6.php>

Correlación

En probabilidad y estadística, la **correlación** indica la fuerza y la dirección de una relación lineal entre dos variables aleatorias. Se considera que dos variables cuantitativas están correlacionadas cuando los valores de una de ellas varían sistemáticamente con respecto a los valores homónimos de la otra: si tenemos dos variables (A y B) existe correlación si al aumentar los valores de A lo hacen también los de B y viceversa. La correlación entre dos variables no implica, por sí misma, ninguna relación de causalidad (Véase Cum hoc ergo propter hoc).

Fuerza, sentido y forma de la correlación

La relación entre dos super variables cuantitativas queda representada mediante la línea de mejor ajuste, trazada a partir de la nube de puntos. Los principales componentes elementales de una línea de ajuste y, por lo tanto, de una correlación, son la fuerza, el sentido y la forma:

- La **fuerza** extrema según el caso, mide el grado en que la línea representa a la nube de puntos: si la nube es estrecha y alargada, se representa por una línea recta, lo que indica que la relación es *fuerte*; si la nube de puntos tiene una tendencia elíptica o circular, la relación es *débil*.
- El **sentido** mide la variación de los valores de B con respecto a A: si al crecer los valores de A lo hacen los de B, la relación es *positiva*; si al crecer los valores de A disminuyen los de B, la relación es *negativa*.
- La **forma** establece el tipo de línea que define el mejor ajuste: la línea rectal, la curva monotónica o la curva no monotónica.

Coefficientes de correlación

Existen diversos coeficientes que miden el grado de correlación, adaptados a la naturaleza de los datos. El más conocido es el coeficiente de correlación de Pearson (introducido en realidad por Francis Galton), que se obtiene dividiendo la covarianza de dos variables por el producto de sus desviaciones estándar. Otros coeficientes son:

- Coeficiente de correlación de Spearman
- Correlación canónica
- [Coeficiente de Correlación Intraclase]

Interpretación geométrica

Ambas series de valores $X(x_1, \dots, x_n)$ e $Y(y_1, \dots, y_n)$ pueden ser consideradas como vectores en un espacio a n dimensiones. Reemplacémoslos por vectores centrados:

$$X(x_1 - \bar{x}, \dots, x_n - \bar{x}) \text{ e } Y(y_1 - \bar{y}, \dots, y_n - \bar{y}).$$

El coseno del ángulo α entre estos vectores es dada por la fórmula siguiente:

$$\cos(\alpha) = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

Pues $\cos(\alpha)$ es el coeficiente de correlación de Pearson.

¡El coeficiente de correlación es el coseno entre ambos vectores centrados!

Si $r = 1$, el ángulo $\alpha = 0^\circ$, ambos vectores son colineales (paralelos).

Si $r = 0$, el ángulo $\alpha = 90^\circ$, ambos vectores son ortogonales.

Si $r = -1$, el ángulo $\alpha = 180^\circ$, ambos vectores son colineales de dirección opuesto.

Más generalmente: $\alpha = \arccos(r)$.

Por supuesto, del punto de vista geométrica, no hablamos de *correlación lineal*: el coeficiente de correlación tiene siempre un sentido, cualquiera que sea su valor entre -1 y 1. Nos informa de modo preciso, no tanto sobre el grado de dependencia entre las variables, que sobre su distancia angular en la hipersfera a n dimensiones.

La Iconografía de las correlaciones es un método de análisis multidimensional que reposa en esta idea.

la correlacion lineal se da cuando en una nube de puntos estos se encuentran o se distribuyen alrededor de una recta.

Enlaces externos

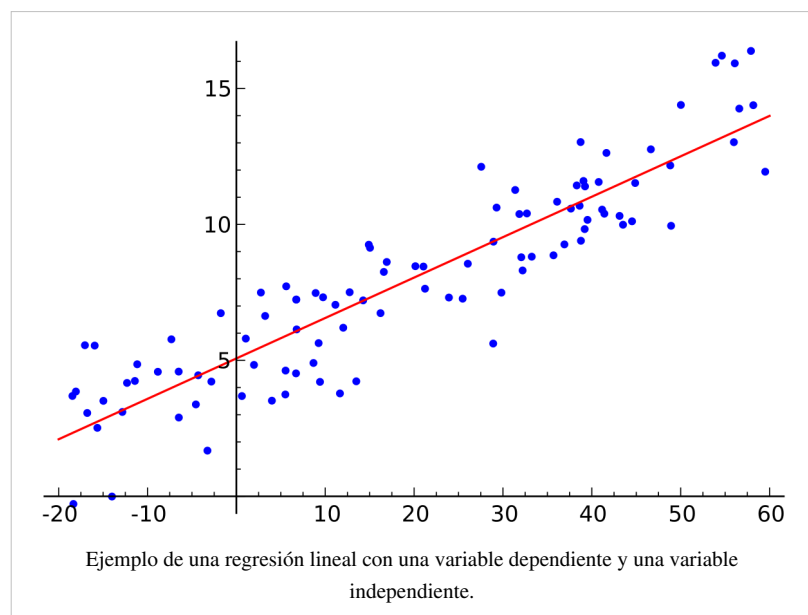
- Diccionario Estadístico - Divestadística ^[1] (en castellano)
- [2] Simulación de la correlación entre dos variables discretas con R (lenguaje de programación)

Referencias

- [1] http://www.divestadistica.es/es/diccionario_estadistico.html#C
 [2] <http://cajael.com/mestadisticos/T1EDescriptiva/node20.php>

Regresión lineal

En estadística la **regresión lineal** o **ajuste lineal** es un método matemático que modeliza la relación entre una variable dependiente Y , las variables independientes X_i y un término aleatorio ε . Este modelo puede ser expresado como:



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

donde β_0 es la intersección o término "constante", las β_i ($i > 0$) son los parámetros respectivos a cada variable independiente, y p es el número de parámetros independientes a tener en cuenta en la regresión. La regresión lineal puede ser contrastada con la regresión no lineal.

Historia

La primera forma de regresiones lineales documentada fue el método de los mínimos cuadrados, el cual fue publicado por Legendre en 1805,^[ref] Gauss publicó un trabajo en donde desarrollaba de manera más profunda el método de los mínimos cuadrados,^[1] y en dónde se incluía una versión del teorema de Gauss-Márkov.

Etimología

El término *regresión* se utilizó por primera vez en el estudio de variables antropométricas: al comparar la estatura de padres e hijos, resultó que los hijos cuyos padres tenían una estatura muy superior al valor medio tendían a igualarse a éste, mientras que aquellos cuyos padres eran muy bajos tendían a reducir su diferencia respecto a la estatura media; es decir, "regresaban" al promedio.^[2] La constatación empírica de esta propiedad se vio reforzada más tarde con la justificación teórica de ese fenómeno.

El término *lineal* se emplea para distinguirlo del resto de técnicas de regresión, que emplean modelos basados en cualquier clase de función matemática. Los modelos lineales son una explicación simplificada de la realidad, mucho más ágil y con un soporte teórico por parte de la matemática y la estadística mucho más extenso.

Pero bien, como se ha dicho, podemos usar el término lineal para distinguir modelos basados en cualquier clase de aplicación.

El modelo de regresión lineal

El modelo lineal relaciona la variable dependiente Y con K variables explicativas X_k ($k = 1, \dots, K$), o cualquier transformación de éstas, que generan un hiperplano de parámetros β_k desconocidos:

$$(2) Y = \sum \beta_k X_k + \varepsilon$$

donde ε es la perturbación aleatoria que recoge todos aquellos factores de la realidad no controlables u observables y que por tanto se asocian con el azar, y es la que confiere al modelo su carácter estocástico. En el caso más sencillo, con una sola variable explicativa, el hiperplano es una recta:

$$(3) Y = \beta_1 + \beta_2 X_2 + \varepsilon$$

El problema de la regresión consiste en elegir unos valores determinados para los parámetros desconocidos β_k , de modo que la ecuación quede completamente especificada. Para ello se necesita un conjunto de observaciones. En una observación cualquiera i -ésima ($i = 1, \dots, I$) se registra el comportamiento simultáneo de la variable dependiente y las variables explicativas (las perturbaciones aleatorias se suponen no observables).

$$(4) Y_i = \sum \beta_k X_{ki} + \varepsilon_i$$

Los valores escogidos como estimadores de los parámetros, $\hat{\beta}_k$, son los coeficientes de regresión, sin que se pueda garantizar que coinciden con parámetros reales del proceso generador. Por tanto, en

$$(5) Y_i = \sum \hat{\beta}_k X_{ki} + \hat{\varepsilon}_i$$

Los valores $\hat{\varepsilon}_i$ son por su parte estimaciones de la perturbación aleatoria o errores.

Supuestos del modelo de regresión lineal

Para poder crear un modelo de regresión lineal, es necesario que se cumpla con los siguientes supuestos:^[3]

1. La relación entre las variables es lineal.
2. Los errores en la medición de las variables explicativas son independientes entre sí.
3. Los errores tienen varianza constante. (Homocedasticidad)
4. Los errores tienen una esperanza matemática igual a cero (los errores de una misma magnitud y distinto signo son equiprobables).
5. El error total es la suma de todos los errores.

Tipos de modelos de regresión lineal

Existen diferentes tipos de regresión lineal que se clasifican de acuerdo a sus parámetros:

Regresión lineal simple

Sólo se maneja una variable independiente, por lo que sólo cuenta con dos parámetros. Son de la forma:^[4]

$$(6) Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

donde ε_i es el error asociado a la medición del valor X_i y siguen los supuestos de modo que $\varepsilon_i \sim N(0, \sigma^2)$ (media cero, varianza constante e igual a un σ y $\varepsilon_i \perp \varepsilon_j$ con $i \neq j$).

Análisis

Dado el modelo de regresión simple, si se calcula la esperanza (valor esperado) del valor Y , se obtiene:^[5]

$$(7) E(y_i) = \hat{y}_i = E(\beta_0) + E(\beta_1 x_i) + E(\varepsilon_i)$$

Derivando respecto a $\hat{\beta}_0$ y $\hat{\beta}_1$ e igualando a cero, se obtiene:^[5]

$$(9) \frac{\partial \sum (y_i - \hat{y}_i)^2}{\partial \hat{\beta}_0} = 0$$

$$(10) \frac{\partial \sum (y_i - \hat{y}_i)^2}{\partial \hat{\beta}_1} = 0$$

Obteniendo dos ecuaciones denominadas ecuaciones normales que generan la siguiente solución para ambos parámetros:^[4]

$$(11) \hat{\beta}_1 = \frac{\sum x \sum y - n \sum xy}{(\sum x)^2 - n \sum x^2} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}$$

$$(12) \hat{\beta}_0 = \frac{\sum y - \hat{\beta}_1 \sum x}{n} = \bar{y} - \hat{\beta}_1 \bar{x}$$

La interpretación del parámetro β_1 es que un incremento en X_i de una unidad, Y_i incrementará en β_1

Regresión lineal múltiple

La regresión lineal nos permite trabajar con una variable a nivel de intervalo o razón, así también se puede comprender la relación de dos o más variables y nos permitirá relacionar mediante ecuaciones, una variable en relación a otras variables llamándose Regresión múltiple. Constantemente en la práctica de la investigación estadística, se encuentran variables que de alguna manera están relacionados entre sí, por lo que es posible que una de las variables puedan relacionarse matemáticamente en función de otra u otras variables.

Maneja varias variables independientes. Cuenta con varios parámetros. Se expresan de la forma:^[6]

$$(13) Y_i = \beta_0 + \sum \beta_i X_{ip} + \varepsilon_i$$

donde ε_i es el error asociado a la medición i del valor X_{ip} y siguen los supuestos de modo que $\varepsilon_i \sim N(0, \sigma^2)$ (media cero, varianza constante e igual a un σ y $\varepsilon_i \perp \varepsilon_j$ con $i \neq j$).

Rectas de regresión

Las rectas de regresión son las rectas que mejor se ajustan a la nube de puntos (o también llamado diagrama de dispersión) generada por una distribución binomial. Matemáticamente, son posibles dos rectas de máximo ajuste.^[7]

- La recta de regresión de Y sobre X :

$$(14) \quad y = \bar{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x - \bar{x})$$

- La recta de regresión de X sobre Y :

$$(15) \quad x = \bar{x} + \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y})$$

La correlación (" r ") de las rectas determinará la calidad del ajuste. Si r es cercano o igual a 1, el ajuste será bueno y las predicciones realizadas a partir del modelo obtenido serán muy fiables (el modelo obtenido resulta verdaderamente representativo); si r es cercano o igual a 0, se tratará de un ajuste malo en el que las predicciones que se realicen a partir del modelo obtenido no serán fiables (el modelo obtenido no resulta representativo de la realidad). Ambas rectas de regresión se intersecan en un punto llamado centro de gravedad de la distribución.

Aplicaciones de la regresión lineal

Líneas de tendencia

Véase también: *Tendencia*

Una *línea de tendencia* representa una tendencia en una serie de datos obtenidos a través de un largo período. Este tipo de líneas puede decirnos si un conjunto de datos en particular (como por ejemplo, el PBI, el precio del petróleo o el valor de las acciones) han aumentado o decrementado en un determinado período.^[8] Se puede dibujar una línea de tendencia a simple vista fácilmente a partir de un grupo de puntos, pero su posición y pendiente se calcula de manera más precisa utilizando técnicas estadísticas como las regresiones lineales. Las líneas de tendencia son generalmente líneas rectas, aunque algunas variaciones utilizan polinomios de mayor grado dependiendo de la curvatura deseada en la línea.

Medicina

En medicina, las primeras evidencias relacionando la mortalidad con el fumar tabaco^[9] vinieron de estudios que utilizaban la regresión lineal. Los investigadores incluyen una gran cantidad de variables en su análisis de regresión en un esfuerzo por eliminar factores que pudieran producir correlaciones espurias. En el caso del tabaquismo, los investigadores incluyeron el estado socio-económico para asegurarse que los efectos de mortalidad por tabaquismo no sean un efecto de su educación o posición económica. No obstante, es imposible incluir todas las variables posibles en un estudio de regresión.^{[10] [11]} En el ejemplo del tabaquismo, un hipotético gen podría aumentar la mortalidad y aumentar la propensión a adquirir enfermedades relacionadas con el consumo de tabaco. Por esta razón, en la actualidad las pruebas controladas aleatorias son consideradas mucho más confiables que los análisis de regresión.

Véase también

- Homoscedasticidad
- Regresión logística
- Modelos de regresión múltiple postulados y no postulados
- Regresión segmentada
- Econometría
- Mínimos cuadrados
- Regularización de Tikhonov
- Cuarteto de Anscombe
- Capital Asset Pricing Model

Referencias

- [1] C.F. Gauss. *Theoria combinationis observationum erroribus minimis obnoxiae*. (1821/1823)
- [2] Introduction to linear regression (http://www.curvefit.com/linear_regression.htm) Curvefit.com (en inglés)
- [3] "Análisis de regresión lineal" (<http://www.ucm.es/info/socivmyt/paginas/profesorado/benitacompostela/tema2.doc>), Universidad Complutense de Madrid
- [4] "Fórmulas", *Probabilidad y Estadística*. Cs. Básicas. U.D.B. Matemática. Universidad Tecnológica Nacional, Facultad Regional Buenos Aires. Editorial CEIT-FRBA. (Código BM2BT2)
- [5] Modelo de regresión lineal simple. (<http://www.einsteinnet.com/econometria/Introeconometria/regsimple.htm>) EinsteinNet.
- [6] Técnicas de regresión: Regresión Lineal Múltiple. (http://www.fisterra.com/mbe/investiga/regre_lineal_multi/regre_lineal_multi.asp) Pértega Díaz, S., Pita Fernández, S. Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario de La Coruña (España)
- [7] Apunte sobre Rectas de regresión. (http://descartes.cnice.mec.es/materiales_didacticos/bidimensional_lbarrios/regresion_est.htm) Ministerio de Educación y Ciencia. Gobierno de España.
- [8] Utilización de las líneas de tendencia (<http://www.paritech.com/paritech-site/education/technical/indicators/trend/linear3.asp>), Paritech (en inglés)
- [9] Doll R, Peto r, Wheatley K, Gray R et al. *Mortality in relation to smoking: 40 years' observations on male British doctors*. *BMJ* 1994;309:901-911 (8 de octubre)
- [10] "Environmental Tobacco Smoke and Adult Asthma" (<http://ajrcm.satsjournals.org/cgi/content/full/158/1/170>) Division of Pulmonary and Critical Care Medicine, Division of Occupational and Environmental Medicine; Department of Medicine, Institute for Health Policy Studies; and Department of Epidemiology and Biostatistics, Universidad de California, San Francisco, California. (en inglés)
- [11] Efecto del tabaquismo, los síntomas respiratorios y el asma sobre la espirometría de adultos de la Ciudad de México (http://www.scielosp.org/scielo.php?script=sci_arttext&pid=S0036-36342005000500002&lng=pt), Justino Regalado-Pineda; Alejandro Gómez-Gómez; Javier Ramírez-Acosta; Juan Carlos Vázquez-García

Bibliografía

- Devore, Jay L.; *Probabilidad y Estadística para Ingeniería y Ciencias*. International Thomson Editores. México. ISBN-10: 9706864571.
- Walpole, Ronald E.; Raymond H.; Myers, Sharon L.; *Probabilidad y Estadística para Ingenieros*. Prentice-Hall Hispanoamericana, S.A. México. ISBN-10: 9701702646.
- Canavos, George C.; *Probabilidad y Estadística. Aplicaciones y Métodos*. McGraw-Hill. México. ISBN-10: 9684518560.

Enlaces externos

- Cálculo de regresiones lineales en línea. (<http://www.wessa.net/esteq.wasp>) (en inglés)
- ZunZun.com (<http://zunzun.com/>) Ajuste de curvas y superficies en línea. (en inglés)
- xuru.org (<http://www.xuru.org/rt/LR.asp>) Herramientas de regresión lineal en línea. (en inglés)
- (<http://cajael.com/mestadisticos/T4DProbabilidad/node8.php>) Simulación de la recta de regresión de una variable bidimensional continua con R (lenguaje de programación)

Función generadora de momentos

En probabilidad y estadística, la **función generadora de momentos** o **función generatriz de momentos** de una variable aleatoria X es

$$M_X(t) := E(e^{tX}), \quad t \in \mathbb{R},$$

siempre que esta esperanza exista.

La función generadora de momentos se llama así porque, si existe en un entorno de $t=0$, permite generar los momentos de la distribución de probabilidad:

$$E(X^n) = M_X^{(n)}(0) = \frac{d^n M_X}{dt^n}(0).$$

Si la función generadora de momentos está definida en tal intervalo, entonces determina unívocamente a la distribución de probabilidad. ^[*cita requerida*]

Un problema clave con las funciones generadoras de momentos es que los momentos y la propia función generadora no siempre existen, porque las integrales que los definen no son siempre convergentes. Por el contrario, la función característica siempre existe y puede usarse en su lugar.

De forma general, donde $\mathbf{X} = (X_1, \dots, X_n)$ es un vector aleatorio n -dimensional, se usa $\mathbf{t} \cdot \mathbf{X} = \mathbf{t}^T \mathbf{X}$ en lugar de tX :

$$M_{\mathbf{X}}(\mathbf{t}) := E(e^{t^T \mathbf{X}}).$$

Cálculo

Si X tiene una función de densidad continua, $f(x)$, entonces la función generadora de momentos viene dada por

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f(x) dx \\ &= \int_{-\infty}^{\infty} \left(1 + tx + \frac{t^2 x^2}{2!} + \dots \right) f(x) dx \\ &= 1 + tm_1 + \frac{t^2 m_2}{2!} + \dots, \end{aligned}$$

donde m_i es el i -ésimo momento. $M_X(-t)$ es, precisamente, la transformada bilateral de Laplace de $f(x)$.

Independientemente de que la distribución de probabilidad sea continua o no, la función generadora de momentos viene dada por la integral de Riemann-Stieltjes

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} dF(x)$$

donde F es la función de distribución.

Si X_1, X_2, \dots, X_n es una secuencia de variables aleatorias independientes (y no necesariamente idénticamente distribuidas) y

$$S_n = \sum_{i=1}^n a_i X_i,$$

donde las a_i son constantes, entonces la función de densidad de S_n es la convolución de la función de densidad de cada una de las X_i y la función generadora de momentos para S_n viene dada por

$$M_{S_n}(t) = M_{X_1}(a_1 t) M_{X_2}(a_2 t) \cdots M_{X_n}(a_n t).$$

Para variables aleatorias multidimensionales X con componentes reales, la función generadora de momentos viene dada por

$$M_X(t) = E(e^{t, X})$$

donde t es un vector y $\langle t, X \rangle$ es el producto punto.

Relación con otras funciones

Hay una serie de transformadas relacionadas con la función generadora de momentos que son comunes en la teoría de probabilidades:

Función característica

La función característica $\varphi_X(t)$ está relacionada con la función generadora de momentos via $\varphi_X(t) = M_{iX}(t) = M_X(it)$: La función característica es la función generadora de momentos de iX o la función generadora de momentos de X evaluada en los ejes imaginarios.

Función generadora acumulada

La **función generadora acumulada** está definida como el logaritmo de la función generadora de momentos; hay quien define la función generadora acumulada como el logaritmo de la función característica, mientras que otros llaman a esta función la *segunda* función generadora acumulada.

Función generadora de probabilidad

Véase también

- Función factorial generadora de momentos
- Función cociente
- Tabla de funciones generadoras de momentos comunes

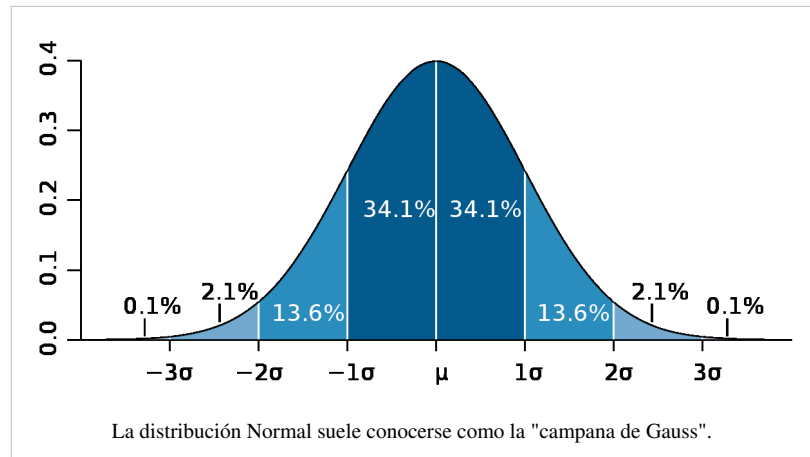
Distribuciones

Distribución de probabilidad

En teoría de la probabilidad y estadística, la **distribución de probabilidad** de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable aleatoria la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos los eventos rango de valores de la variable aleatoria.

Cuando la variable aleatoria toma valores en el conjunto de los números

reales, la distribución de probabilidad está completamente especificada por la **función de distribución**, cuyo valor en cada real x es la probabilidad de que la variable aleatoria sea menor o igual que x .



Definición de función de distribución

Dada una variable aleatoria todos son puntos X , su **función de distribución**, $F_X(x)$, es

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

Por simplicidad, cuando no hay lugar a confusión, suele omitirse el subíndice X y se escribe, simplemente, $F(x)$.

Propiedades

Como consecuencia casi inmediata de la definición, la función de distribución:

- Es una función continua por la derecha.
- Es una función monótona no decreciente.

Además, cumple

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

y

$$\lim_{x \rightarrow +\infty} F(x) = 1$$

Para dos números reales cualesquiera a y b tal que $(a < b)$, los sucesos $(X \leq a)$ y $(a < X \leq b)$ son mutuamente excluyentes y su unión es el suceso $(X \leq b)$, por lo que tenemos entonces que:

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$

y finalmente

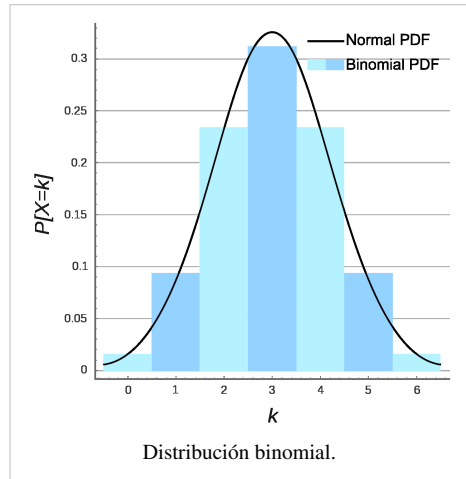
$$P(a < X \leq b) = F(b) - F(a)$$

Por lo tanto una vez conocida la función de distribución $F(x)$ para todos los valores de la variable aleatoria x conoceremos completamente la distribución de probabilidad de la variable.

Para realizar cálculos es más cómodo conocer la distribución de probabilidad, y sin embargo para ver una representación gráfica de la probabilidad es más práctico el uso de la función de densidad.

Distribuciones de variable discreta

Se denomina distribución de variable discreta a aquella cuya función de probabilidad sólo toma valores positivos en un conjunto de valores de X finito o infinito numerable. A dicha función se le llama función de masa de probabilidad. En este caso la distribución de probabilidad es el sumatorio de la función de masa, por lo que tenemos entonces que:



$$F(x) = P(X \leq x) = \sum_{k=-\infty}^x f(k)$$

Y, tal como corresponde a la definición de distribución de probabilidad, esta expresión representa la suma de todas las probabilidades desde $-\infty$ hasta el valor x .

Distribuciones de variable discreta más importantes

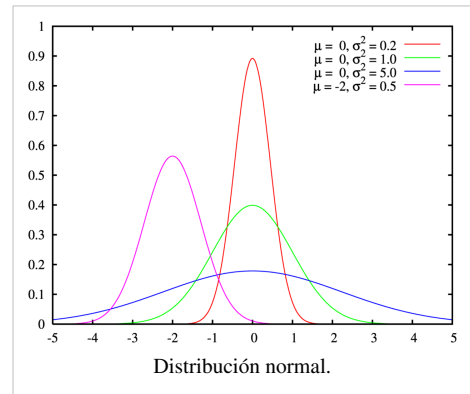
Las distribuciones de variable discreta más importantes son las siguientes:

- Distribución binomial
- Distribución binomial negativa
- Distribución Poisson
- Distribución geométrica
- Distribución hipergeométrica
- Distribución de Bernoulli
- Distribución Rademacher, que toma el valor 1 con probabilidad $1/2$ y el valor -1 con probabilidad $1/2$.
- Distribución uniforme discreta, donde todos los elementos de un conjunto finito son equiprobables.

Distribuciones de variable continua

Se denomina variable continua a aquella que puede tomar cualquiera de los infinitos valores existentes dentro de un intervalo. En el caso de variable continua la distribución de probabilidad es la integral de la función de densidad, por lo que tenemos entonces que:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$




Distribuciones de variable continua más importantes

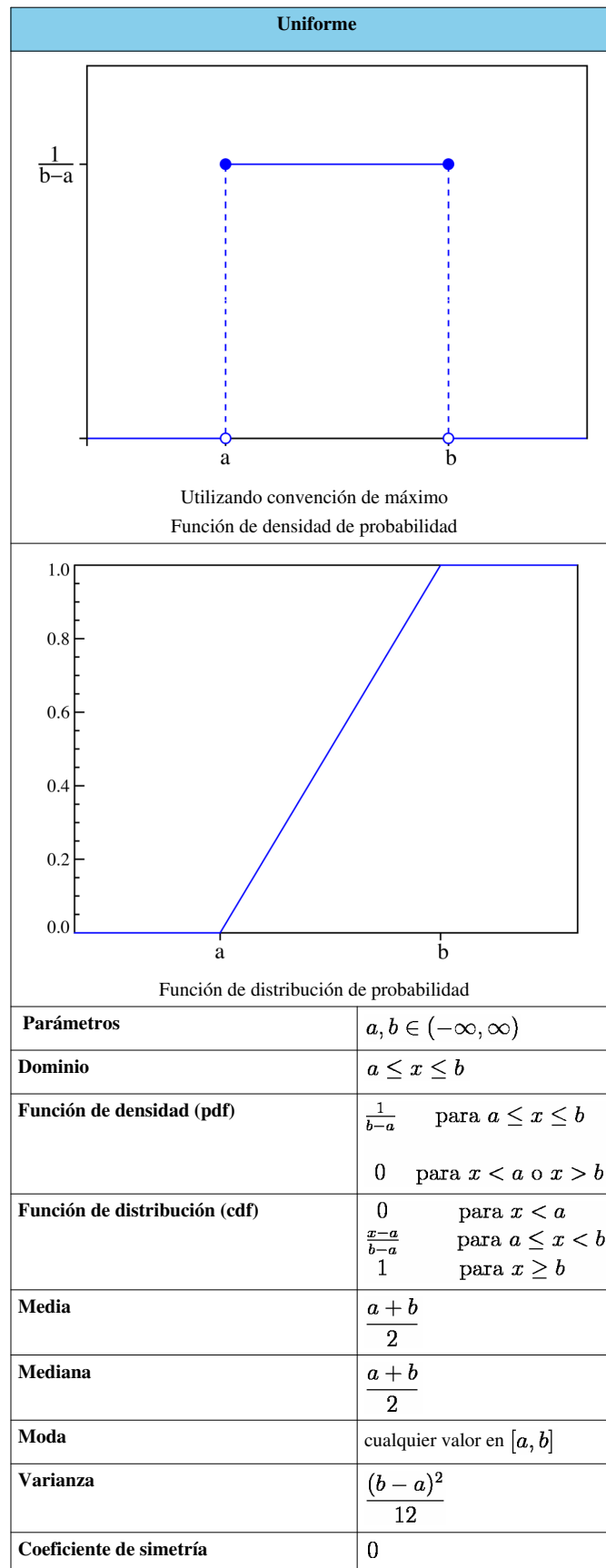
Las distribuciones de variable continua más importantes son las siguientes:

- Distribución ji cuadrado
- Distribución exponencial
- Distribución t de Student
- Distribución normal
- Distribución Gamma
- Distribución Beta
- Distribución F
- Distribución uniforme (continua)

Enlaces externos

-  Wikimedia Commons alberga contenido multimedia sobre **Distribuciones de probabilidad**. Commons
- Wikilibros: Estadística

Distribución uniforme continua



Curtosis	$-\frac{6}{5}$
Entropía	$\ln(b - a)$
Función generadora de momentos (mgf)	$\frac{e^{tb} - e^{ta}}{t(b - a)}$
Función característica	$\frac{e^{itb} - e^{ita}}{it(b - a)}$

En teoría de probabilidad y estadística, la **distribución uniforme continua** es una familia de distribuciones de probabilidad para variables aleatorias continuas, tales que cada miembro de la familia, todos los intervalos de igual longitud en la distribución en su rango son igualmente probables. El dominio está definido por dos parámetros, a y b , que son sus valores mínimo y máximo. La distribución es a menudo escrita en forma abreviada como $U(a,b)$.

Caracterización

Función de densidad de probabilidad

La función de densidad de probabilidad de la distribución uniforme continua es:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b, \\ 0 & \text{para } x < a \text{ o } x > b, \end{cases}$$

Los valores en los dos extremos a y b no son por lo general importantes porque no afectan el valor de las integrales de $f(x) dx$ sobre el intervalo, ni de $x f(x) dx$ o expresiones similares. A veces se elige que sean cero, y a veces se los elige con el valor $1/(b - a)$. Este último resulta apropiado en el contexto de estimación por el método de máxima verosimilitud. En el contexto del análisis de Fourier, se puede elegir que el valor de $f(a)$ ó $f(b)$ sean $1/(2(b - a))$, para que entonces la transformada inversa de muchas transformadas integrales de esta función uniforme resulten en la función inicial, de otra forma la función que se obtiene sería igual "en casi todo punto", o sea excepto en un conjunto de puntos con medida nula. También, de esta forma resulta consistente con la función signo que no posee dicha ambigüedad.

Función de distribución de probabilidad

La función de distribución de probabilidad es:

$$F(x) = \begin{cases} 0 & \text{para } x < a \\ \frac{x-a}{b-a} & \text{para } a \leq x < b \\ 1 & \text{para } x \geq b \end{cases}$$

Funciones generadoras asociadas

Función generadora de momentos

La función generadora de momentos es

$$M_x = E(e^{tx}) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

a partir de la cual se pueden calcular los momentos m_k

$$m_1 = \frac{a+b}{2},$$

$$m_2 = \frac{a^2 + ab + b^2}{3},$$

y, en general,

$$m_k = \frac{1}{k+1} \sum_{i=0}^k a^i b^{k-i}.$$

Para una variable aleatoria que satisface esta distribución, la esperanza matemática es entonces $m_1 = (a+b)/2$ y la varianza es $m_2 - m_1^2 = (b-a)^2/12$.

Propiedades

Generalización a conjuntos de Borel

Esta distribución puede ser generalizada a conjuntos de intervalos más complicados. Si S es un conjunto de Borel de medida finita positiva, la distribución de probabilidad uniforme en S se puede especificar definiendo que la pdf sea nula fuera de S e igual a $1/K$ dentro de S , donde K es la medida de Lebesgue de S .

Estadísticas de orden

Sea X_1, \dots, X_n una muestra i.i.d. de $U(0,1)$. Sea $X_{(k)}$ el orden estadístico k -ésimo de esta muestra. Entonces la distribución de probabilidad de $X_{(k)}$ es una distribución Beta con parámetros k y $n-k+1$. La esperanza matemática es

$$E(X_{(k)}) = \frac{k}{n+1}.$$

Esto es útil cuando se realizan Q-Q plots.

Las varianzas son

$$\text{Var}(X_{(k)}) = \frac{k(n-k+1)}{(n+1)^2(n+2)}.$$

'Uniformidad'

La probabilidad de que una variable aleatoria uniformemente distribuida se encuentre dentro de algún intervalo de longitud finita es independiente de la ubicación del intervalo (aunque sí depende del tamaño del intervalo), siempre que el intervalo esté contenido en el dominio de la distribución.

Es posible verificar esto, por ejemplo si $X \approx U(0, b)$ y $[x, x+d]$ es un subintervalo de $[0, b]$ con d fijo y $d > 0$, entonces

$$P(X \in [x, x+d]) = \int_x^{x+d} \frac{dy}{b-a} = \frac{d}{b-a}$$

lo cual es independiente de x . Este hecho es el que le da su nombre a la distribución.

Uniforme estándar

Si se restringe $a = 0$ y $b = 1$, a la distribución resultante $U(0,1)$ se la llama **distribución uniforme estándar**.

Una propiedad interesante de la distribución uniforme estándar es que si u_1 es una distribución uniforme estándar, entonces $1-u_1$ también lo es.

Distribuciones relacionadas

Si X tiene una distribución uniforme estándar, entonces:

- $Y = -\ln(X)/\lambda$ tiene una distribución exponencial con parámetro λ .
- $Y = 1 - X^{1/n}$ tiene una distribución beta con parámetros 1 y n . (Notar que esto implica que la distribución uniforme estándar es un caso especial de la distribución beta, con parámetros 1 y 1).

Relaciones con otras funciones

Siempre y cuando se sigan las mismas convenciones en los puntos de transición, la función densidad de probabilidad puede también ser expresada mediante la función escalón de Heaviside:

$$f(x) = \frac{H(x-a) - H(x-b)}{b-a},$$

ó en términos de la función rectángulo

$$f(x) = \frac{1}{b-a} \text{rect} \left(\frac{x - \left(\frac{a+b}{2}\right)}{b-a} \right).$$

No existe ambigüedad en el punto de transición de la función signo. Utilizando la convención de la mitad del máximo en los puntos de transición, la distribución uniforme se puede expresar a partir de la función signo como:

$$f(x) = \frac{\text{sgn}(x-a) - \text{sgn}(x-b)}{2(b-a)}.$$

Aplicaciones

En estadística, cuando se utiliza un p-value a modo de prueba estadística para una hipótesis nula simple, y la distribución de la prueba estadística es continua, entonces la prueba estadística esta uniformemente distribuida entre 0 y 1 si la hipótesis nula es verdadera.

Muestreo de una distribución uniforme

Existen muchos usos en que es útil realizar experimentos de simulación. Muchos lenguajes de programación poseen la capacidad de generar números pseudo-aleatorios que están distribuidos de acuerdo a una distribución uniforme estándar.

Si u es un valor muestreado de una distribución uniforme estándar, entonces el valor $a + (b - a)u$ posee una distribución uniforme parametrizada por a y b , como se describió previamente.

Muestreo de una distribución arbitraria

La distribución uniforme resulta útil para muestrear distribuciones arbitrarias. Un método general es el método de muestreo de transformación inversa, que utiliza la distribución de probabilidad (CDF) de la variable aleatoria objetivo. Este método es muy útil en trabajos teóricos. Dado que las simulaciones que utilizan este método requieren invertir la CDF de la variable objetivo, se han diseñado métodos alternativos para aquellos casos donde no se conoce el CDF en una forma cerrada. Otro método similar es el rejection sampling.

La distribución normal es un ejemplo importante en el que el método de la transformada inversa no es eficiente. Sin embargo, existe un método exacto, la transformación de Box-Muller, que utiliza la transformada inversa para convertir dos variables aleatorias uniformes independientes en dos variables aleatorias independientes distribuidas normalmente.

Ejemplo en el intervalo [0,1]

Para este caso el intervalo queda definido por $a = 0$ y $b = 1$.

Entonces resulta:

- $f(x) = 1$ para $0 \leq x \leq 1$
- $F(x) = x$ para $0 \leq x \leq 1$
- $E(X) = 0,5$
- $\text{Var}(X) = 1/12$
- $\sigma_x = \sqrt{\text{Var}(X)} = \sqrt{1/12} \approx 0.29$

Enlaces externos

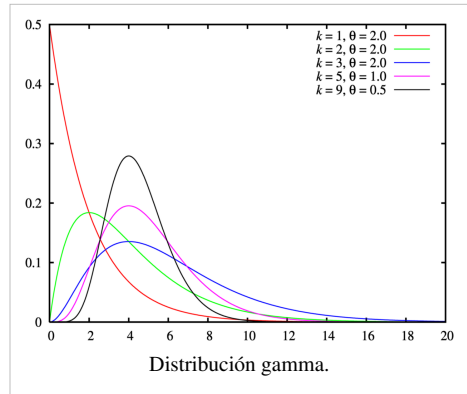
- Calculadora - Distribución uniforme continua ^[1]
- [2] Calcular la probabilidad de una distribución uniforme con R (lenguaje de programación)

Referencias

- [1] <http://www.stud.feec.vutbr.cz/~xvapen02/vypocty/ro.php?language=espanol>
[2] <http://cajael.com/mestadisticos/T7DContinuas/node23.php>

Distribución gamma

En estadística la **distribución gamma** es una distribución de probabilidad continua con dos parámetros k y λ cuya función de densidad para valores $x > 0$ es



$$f(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{k-1}}{\Gamma(k)}$$

Aquí e es el número e y Γ es la función gamma. Para valores $k = 1, 2, \dots$ la función es $\Gamma(k) = (k-1)!$ (el factorial de $k-1$). En este caso - por ejemplo para describir un proceso de Poisson - se llaman la distribución Erlang con un parámetro $\theta = 1/\lambda$.

El valor esperado y la varianza de una variable aleatoria X de distribución gamma son

$$E[X] = k/\lambda = k\theta$$

$$V[X] = k/\lambda^2 = k\theta^2$$

Relaciones

El tiempo hasta que el suceso número k ocurre en un Proceso de Poisson de intensidad λ es una variable aleatoria con distribución gamma. Eso es la suma de k variables aleatorias independientes de distribución exponencial con parámetro λ .

Véase también: Distribución Beta, Distribución Erlang, Distribución Chi-cuadrada

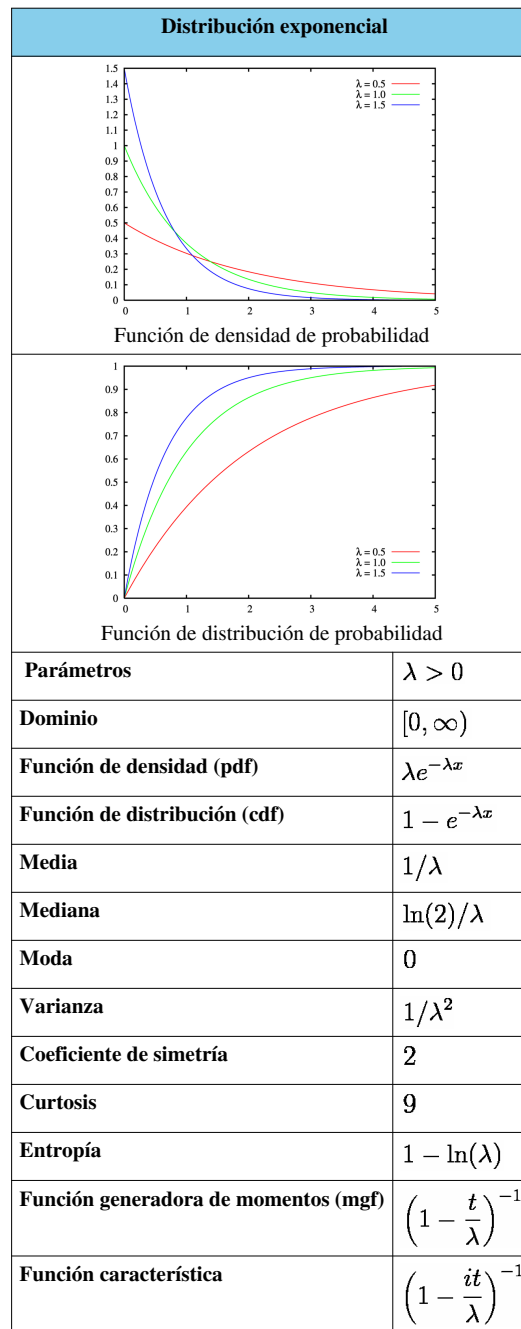
Enlaces externos

- <http://mathworld.wolfram.com/GammaDistribution.html>
- [1] Calcular la probabilidad de una distribución Gamma con R (lenguaje de programación)

Referencias

- [1] <http://cajael.com/mestadisticos/T7DContinuas/node29.php>

Distribución exponencial



En estadística la **distribución exponencial** es una distribución de probabilidad continua con un parámetro $\lambda > 0$ cuya función de densidad es:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x \geq 0 \\ 0 & \text{de otro modo} \end{cases}$$

$$\text{Su función de distribución es: } F(x) = P(X \leq x) = \begin{cases} 0 & \text{para } x < 0 \\ 1 - e^{-\lambda x} & \text{para } x \geq 0 \end{cases}$$

Donde e representa el número e .

El valor esperado y la varianza de una variable aleatoria X con distribución exponencial son:

- $E[X] = \frac{1}{\lambda}$

- $V(X) = \frac{1}{\lambda^2}$

Ejemplo

Ejemplos para la distribución exponencial es la distribución de la longitud de los intervalos de variable continua que transcurren entre la ocurrencia de dos sucesos "raros", que se distribuyen según la distribución de Poisson.

Calcular variables aleatorias

Se pueden calcular una variable aleatoria de distribución exponencial x por medio de una variable aleatoria de distribución uniforme $u = U(0, 1)$:

$$x = -\frac{1}{\lambda} \ln(1 - u)$$

o, dado que $(1 - u)$ es también una variable aleatoria con distribución $U(0, 1)$, puede utilizarse la versión más eficiente:

$$x = -\frac{1}{\lambda} \ln(u)$$

Relaciones

La suma de k variables aleatorias independientes de distribución exponencial con parámetro λ es una variable aleatoria de distribución gamma.

Véase también

- Proceso de Poisson
- Distribución Poisson

Software

Se puede usar software y un programa de computadora para el ajuste de una distribución de probabilidad, incluyendo la exponencial, a una serie de datos:

- Easy fit ^[1], "data analysis & simulation"
- MathWorks Benelux ^[2]
- ModelRisk ^[3], "risk modelling software"
- Ricci distributions, fitting distributions with R ^[4], Vito Ricci, 2005
- Risksolver, automatically fit distributions and parameters to samples ^[5]
- StatSoft distribution fitting ^[6]
- CumFreq ^[7], libre sin costo, incluye intervalos de confianza a base de la distribución binomial

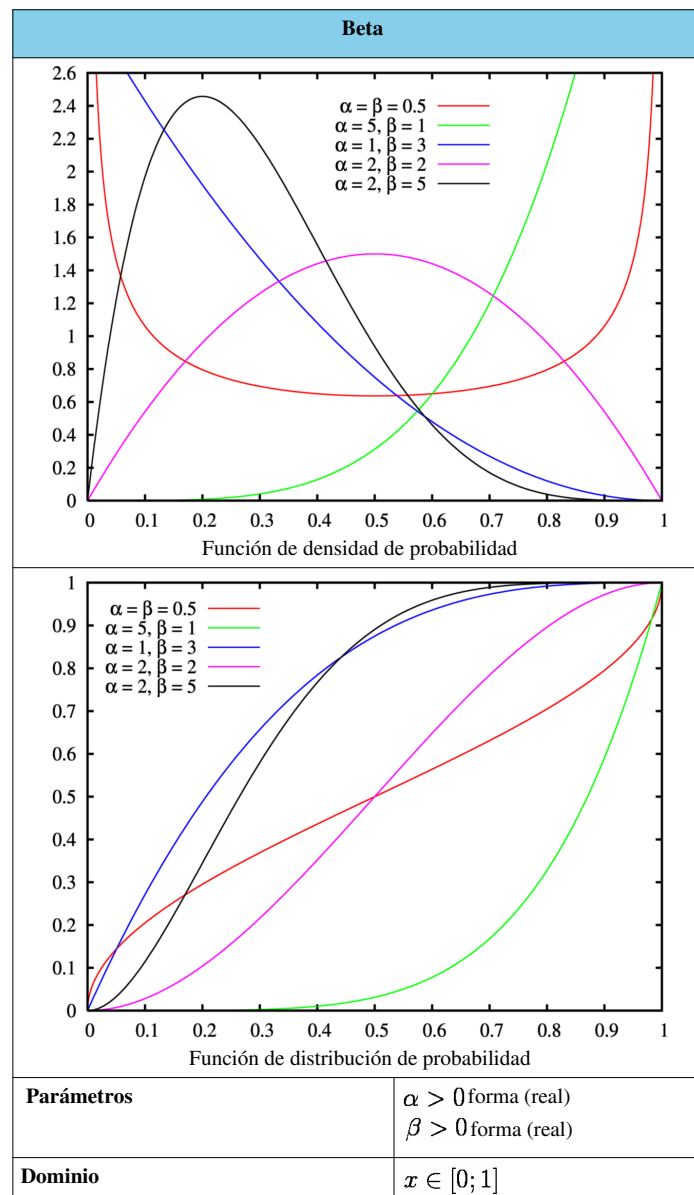
Enlaces externos

- Calculadora Distribución exponencial ^[8]
- [9]Calcular la probabilidad de una distribución exponencial con R (lenguaje de programación)

Referencias

- [1] http://www.mathwave.com/articles/distribution_fitting.html
- [2] <http://www.mathworks.nl/products/statistics/demos.html?file=/products/demos/shipping/stats/cfitdfitdemo.html>
- [3] <http://www.vosesoftware.com/>
- [4] <http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf>
- [5] <http://www.solver.com/risksolver8.htm>
- [6] <http://www.statsoft.com/textbook/distribution-fitting/>
- [7] <http://www.waterlog.info/cumfreq.htm>
- [8] <http://www.stud.feec.vutbr.cz/~xvapen02/vypocty/ex.php?language=espanol>
- [9] <http://cajael.com/mestadisticos/T7DContinuas/node20.php>

Distribución beta



Función de densidad (pdf)	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$
Función de distribución (cdf)	$I_x(\alpha, \beta)$
Media	$\frac{\alpha}{\alpha + \beta}$
Moda	$\frac{\alpha - 1}{\alpha + \beta - 2}$ para $\alpha > 1, \beta > 1$
Varianza	$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$
Coefficiente de simetría	$\frac{2(\beta - \alpha)\sqrt{\alpha + \beta + 1}}{(\alpha + \beta + 2)\sqrt{\alpha\beta}}$
Función generadora de momentos (mgf)	$1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha + r}{\alpha + \beta + r} \right) \frac{t^k}{k!}$
Función característica	${}_1F_1(\alpha; \alpha + \beta; it)$

En estadística la **distribución beta** es una distribución de probabilidad continua con dos parámetros a y b cuya función de densidad para valores $0 < x < 1$ es

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

Aquí Γ es la función gamma.

El valor esperado y la varianza de una variable aleatoria X con distribución beta son

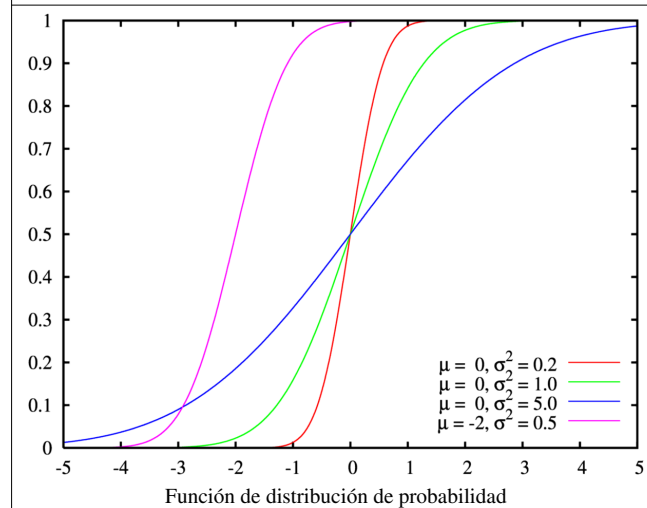
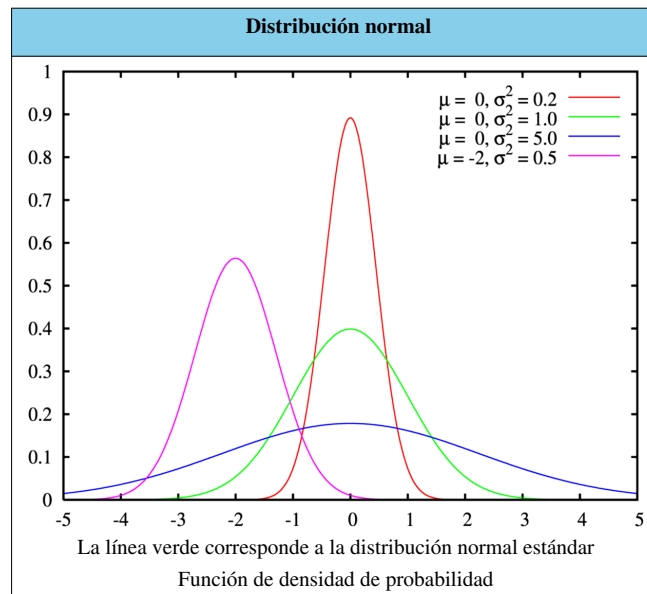
$$E[X] = \frac{a}{a+b}$$

$$V[X] = \frac{ab}{(a+b+1)(a+b)^2}$$

Un caso especial de la distribución beta con $a = 1$ y $b = 1$ es la distribución uniforme en el intervalo $[0, 1]$.

Para relacionar con la muestra se iguala $E[X]$ a la media y $V[X]$ a la varianza y de despejan a y b .

Distribución normal



Parámetros	$\mu \in \mathbb{R}$ $\sigma > 0$
Dominio	$x \in \mathbb{R}$
Función de densidad (pdf)	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Función de distribución (cdf)	$\int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$
Media	μ
Mediana	μ
Moda	μ
Varianza	σ^2
Coficiente de simetría	0
Curtosis	0
Entropía	$\ln(\sigma\sqrt{2\pi}e)$

Función generadora de momentos (mgf)	$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$
Función característica	$\chi_X(t) = e^{\mu i t - \frac{\sigma^2 t^2}{2}}$

En estadística y probabilidad se llama **distribución normal**, **distribución de Gauss** o **distribución gaussiana**, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en fenómenos reales.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro. Esta curva se conoce como campana de Gauss.

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

De hecho, la estadística es un modelo matemático que sólo permite describir un fenómeno, sin explicación alguna. Para la explicación causal es preciso el *diseño experimental*, de ahí que al uso de la estadística en psicología y sociología sea conocido como *método correlacional*.

La distribución normal también es importante por su relación con la estimación por mínimos cuadrados, uno de los métodos de estimación más simples y antiguos.

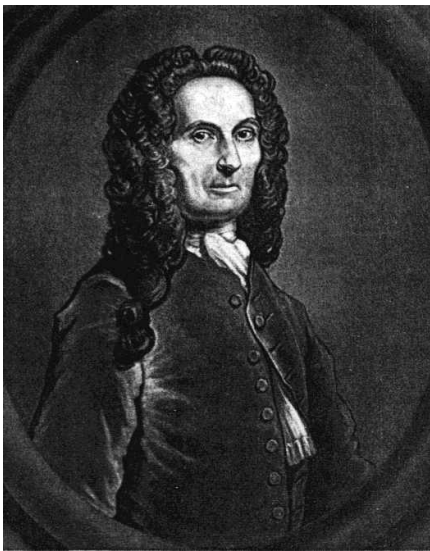
Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son:

- caracteres morfológicos de individuos como la estatura;
- caracteres fisiológicos como el efecto de un fármaco;
- caracteres sociológicos como el consumo de cierto producto por un mismo grupo de individuos;
- caracteres psicológicos como el cociente intelectual;
- nivel de ruido en telecomunicaciones;
- errores cometidos al medir ciertas magnitudes;
- etc.

La distribución normal también aparece en muchas áreas de la propia estadística. Por ejemplo, la distribución muestral de las medias muestrales es aproximadamente normal, cuando la distribución de la población de la cual se extrae la muestra no es normal.^[1] Además, la distribución normal maximiza la entropía entre todas las distribuciones con media y varianza conocidas, lo cual la convierte en la elección natural de la distribución subyacente a una lista de datos resumidos en términos de media muestral y varianza. La distribución normal es la más extendida en estadística y muchos tests estadísticos están basados en una supuesta "normalidad".

En probabilidad, la distribución normal aparece como el límite de varias distribuciones de probabilidad continuas y discretas.

Historia



Abraham de Moivre, descubridor de la distribución normal

La distribución normal fue presentada por primera vez por Abraham de Moivre en un artículo del año 1733,^[2] que fue reimpresso en la segunda edición de su *The Doctrine of Chances*, de 1738, en el contexto de cierta aproximación de la distribución binomial para grandes valores de n . Su resultado fue ampliado por Laplace en su libro *Teoría analítica de las probabilidades* (1812), y en la actualidad se llama Teorema de De Moivre-Laplace.

Laplace usó la distribución normal en el análisis de errores de experimentos. El importante método de mínimos cuadrados fue introducido por Legendre en 1805. Gauss, que afirmaba haber usado el método desde 1794, lo justificó rigurosamente en 1809 asumiendo una distribución normal de los errores. El nombre de Gauss se ha asociado a esta distribución porque la usó con profusión cuando analizaba datos astronómicos^[3] y algunos autores le atribuyen un descubrimiento independiente del de De Moivre.^[4] Esta atribución del nombre de la distribución a una persona distinta de su primer descubridor es un claro ejemplo de la Ley de Stigler.

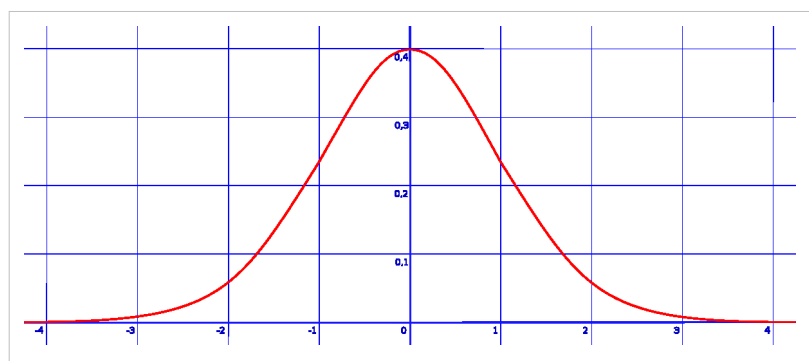
El nombre de "campana" viene de Esprit Jouffret que usó el término "bell surface" (superficie campana) por primera vez en 1872 para una distribución normal bivalente de componentes independientes. El nombre de "distribución normal" fue otorgado independientemente por Charles S. Peirce, Francis Galton y Wilhelm Lexis hacia 1875.^[cita requerida] A pesar de esta terminología, otras distribuciones de probabilidad podrían ser más apropiadas en determinados contextos; véase la discusión sobre ocurrencia, más abajo.

Definición formal

Hay varios modos de definir formalmente una distribución de probabilidad. La forma más visual es mediante su función de densidad. De forma equivalente, también pueden darse para su definición la función de distribución, los momentos, la función característica y la función generatriz de momentos, entre otros.

Función de densidad

Se dice que una variable aleatoria continua X sigue una distribución normal de parámetros μ y σ y se denota $X \sim N(\mu, \sigma)$ si su función de densidad está dada por:



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R},$$

donde μ (miu) es la media y σ (sigma) es la desviación típica (σ^2 es la varianza).^[5]

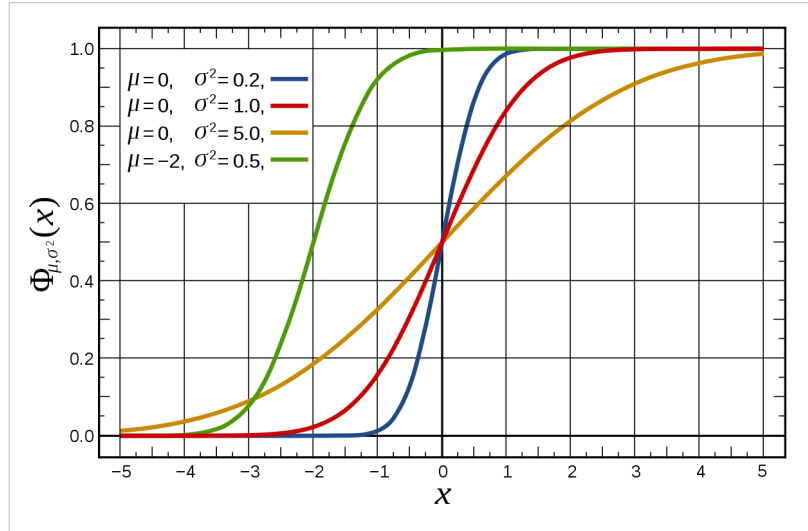
Se llama **distribución normal "estándar"** a aquella en la que sus parámetros toman los valores $\mu = 0$ y $\sigma = 1$. En este caso la función de densidad tiene la siguiente expresión:

$$f(x) = f_{0,1}(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}, \quad x \in \mathbb{R}.$$

Su gráfica se muestra a la derecha y con frecuencia se usan ...tablas para el cálculo de los valores de su distribución.

Función de distribución

La función de distribución de la distribución normal está definida como sigue:



$$\begin{aligned} \Phi_{\mu,\sigma^2}(x) &= \int_{-\infty}^x \varphi_{\mu,\sigma^2}(u) \, du \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(u-\mu)^2}{2\sigma^2}} \, du, \quad x \in \mathbb{R} \end{aligned}$$

Por tanto, la función de distribución de la normal estándar es:

$$\Phi(x) = \Phi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} \, du, \quad x \in \mathbb{R}.$$

Esta función de distribución puede expresarse en términos de una función especial llamada función error de la siguiente forma:

$$\Phi(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right], \quad x \in \mathbb{R},$$

y la propia función de distribución puede, por consiguiente, expresarse así:

$$\Phi_{\mu,\sigma^2}(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x - \mu}{\sigma\sqrt{2}} \right) \right], \quad x \in \mathbb{R}.$$

El complemento de la función de distribución de la normal estándar, $1 - \Phi(x)$, se denota con frecuencia $Q(x)$, y es referida, a veces, como simplemente **función Q**, especialmente en textos de ingeniería.^{[6] [7]} Esto representa la cola de probabilidad de la distribución gaussiana. También se usan ocasionalmente otras definiciones de la función Q, las cuales son todas ellas transformaciones simples de Φ .^[8]

La inversa de la función de distribución de la normal estándar (función cuantil) puede expresarse en términos de la inversa de la función de error:

$$\Phi^{-1}(p) = \sqrt{2} \operatorname{erf}^{-1}(2p - 1), \quad p \in (0, 1),$$

y la inversa de la función de distribución puede, por consiguiente, expresarse como:

$$\Phi_{\mu, \sigma^2}^{-1}(p) = \mu + \sigma \Phi^{-1}(p) = \mu + \sigma \sqrt{2} \operatorname{erf}^{-1}(2p - 1), \quad p \in (0, 1).$$

Esta función cuantil se llama a veces la función probit. No hay una primitiva elemental para la función probit. Esto no quiere decir meramente que no se conoce, sino que se ha probado la inexistencia de tal función. Existen varios métodos exactos para aproximar la función cuantil mediante la distribución normal (véase función cuantil).

Los valores $\Phi(x)$ pueden aproximarse con mucha precisión por distintos métodos, tales como integración numérica, series de Taylor, series asintóticas y fracciones continuas.

Límite inferior y superior estrictos para la función de distribución

Para grandes valores de x la función de distribución de la normal estándar $\Phi(x)$ es muy próxima a 1 y $\Phi(-x) = 1 - \Phi(x)$ está muy cerca de 0. Los límites elementales

$$\frac{x}{1+x^2} \varphi(x) < 1 - \Phi(x) < \frac{\varphi(x)}{x}, \quad x > 0,$$

en términos de la densidad φ son útiles.

Usando el cambio de variable $v = u^2/2$, el límite superior se obtiene como sigue:

$$\begin{aligned} 1 - \Phi(x) &= \int_x^\infty \varphi(u) \, du \\ &< \int_x^\infty \frac{u}{x} \varphi(u) \, du = \int_{x^2/2}^\infty \frac{e^{-v}}{x\sqrt{2\pi}} \, dv = -\frac{e^{-v}}{x\sqrt{2\pi}} \Big|_{x^2/2}^\infty = \frac{\varphi(x)}{x}. \end{aligned}$$

De forma similar, usando $\varphi'(u) = -u\varphi(u)$ y la regla del cociente,

$$\begin{aligned} \left(1 + \frac{1}{x^2}\right)(1 - \Phi(x)) &= \left(1 + \frac{1}{x^2}\right) \int_x^\infty \varphi(u) \, du \\ &= \int_x^\infty \left(1 + \frac{1}{x^2}\right) \varphi(u) \, du \\ &> \int_x^\infty \left(1 + \frac{1}{u^2}\right) \varphi(u) \, du = -\frac{\varphi(u)}{u} \Big|_x^\infty = \frac{\varphi(x)}{x}. \end{aligned}$$

Resolviendo para $1 - \Phi(x)$ proporciona el límite inferior.

Funciones generadoras

Función generadora de momentos

La función generadora de momentos se define como la esperanza de e^{tX} . Para una distribución normal, la función generadora de momentos es:

$$M_X(t) = \mathbb{E} [e^{tX}] = \int_{-\infty}^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{tx} \, dx = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

como puede comprobarse completando el cuadrado en el exponente.

Función característica

La función característica se define como la esperanza de e^{itX} , donde i es la unidad imaginaria. De este modo, la función característica se obtiene reemplazando t por it en la función generadora de momentos.

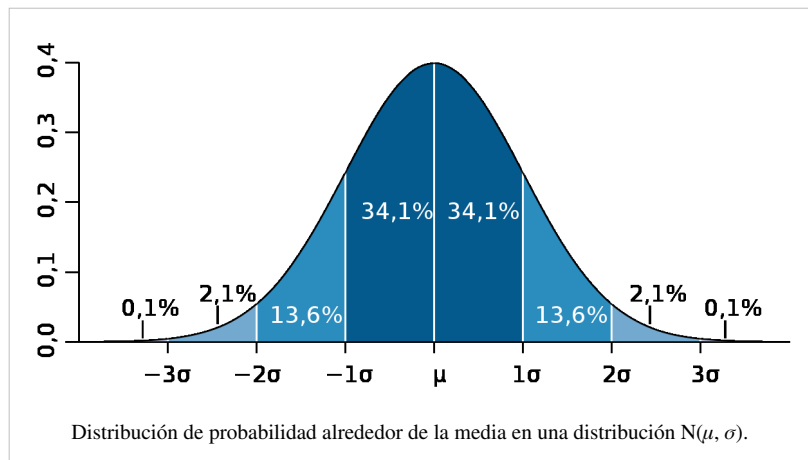
Para una distribución normal, la función característica es^[9]

$$\begin{aligned} \chi_X(t; \mu, \sigma) &= M_X(it) = E[e^{itX}] \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{itx} dx \\ &= e^{i\mu t - \frac{\sigma^2 t^2}{2}}. \end{aligned}$$

Propiedades

Algunas propiedades de la distribución normal son:

1. Es simétrica respecto de su media, μ ;
2. La moda y la mediana son ambas iguales a la media, μ ;
3. Los puntos de inflexión de la curva se dan para $x = \mu - \sigma$ y $x = \mu + \sigma$.
4. Distribución de probabilidad en un entorno de la media:



1. en el intervalo $[\mu - \sigma, \mu + \sigma]$ se encuentra comprendida, aproximadamente, el 68,26% de la distribución;
2. en el intervalo $[\mu - 2\sigma, \mu + 2\sigma]$ se encuentra, aproximadamente, el 95,44% de la distribución;

3. por su parte, en el intervalo $[\mu - 3\sigma, \mu + 3\sigma]$ se encuentra comprendida, aproximadamente, el 99,74% de la distribución. Estas propiedades son de gran utilidad para el establecimiento de intervalos de confianza. Por otra parte, el hecho de que prácticamente la totalidad de la distribución se encuentre a tres desviaciones típicas de la media justifica los límites de las tablas empleadas habitualmente en la normal estándar.

5. Si $X \sim N(\mu, \sigma^2)$ y a y b son números reales, entonces $(aX + b) \sim N(a\mu + b, a^2\sigma^2)$.
6. Si $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$ son variables aleatorias normales independientes, entonces:
 - Su suma está normalmente distribuida con $U = X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$ (demostración). Recíprocamente, si dos variables aleatorias independientes tienen una suma normalmente distribuida, deben ser normales (Teorema de Crámer).
 - Su diferencia está normalmente distribuida con $V = X - Y \sim N(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$.
 - Si las varianzas de X e Y son iguales, entonces U y V son independientes entre sí.
7. Si $X \sim N(0, \sigma_X^2)$ e $Y \sim N(0, \sigma_Y^2)$ son variables aleatorias independientes normalmente distribuidas, entonces:
 - Su producto XY sigue una distribución con densidad p dada por

$$p(z) = \frac{1}{\pi \sigma_X \sigma_Y} K_0\left(\frac{|z|}{\sigma_X \sigma_Y}\right), \text{ donde } K_0 \text{ es una función de Bessel modificada de segundo tipo.}$$

- Su cociente sigue una distribución de Cauchy con $X/Y \sim \text{Cauchy}(0, \sigma_X/\sigma_Y)$. De este modo la distribución de Cauchy es un tipo especial de distribución cociente.
8. Si X_1, \dots, X_n son variables normales estándar independientes, entonces $X_1^2 + \dots + X_n^2$ sigue una distribución χ^2 con n grados de libertad.
 9. Si X_1, \dots, X_n son variables normales estándar independientes, entonces la media muestral $\bar{X} = (X_1 + \dots + X_n)/n$ y la varianza muestral $S^2 = ((X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2)/(n - 1)$ son independientes. Esta propiedad caracteriza a las distribuciones normales y contribuye a explicar por qué el test-F no es robusto respecto a la no-normalidad).

Estandarización de variables aleatorias normales

Como consecuencia de la Propiedad 1; es posible relacionar todas las variables aleatorias normales con la distribución normal estándar.

Si $X \sim N(\mu, \sigma^2)$, entonces

$$Z = \frac{X - \mu}{\sigma}$$

es una variable aleatoria normal estándar: $Z \sim N(0, 1)$.

La transformación de una distribución $X \sim N(\mu, \sigma)$ en una $N(0, 1)$ se llama **normalización, estandarización o tipificación** de la variable X .

Una consecuencia importante de esto es que la función de distribución de una distribución normal es, por consiguiente,

$$\Pr(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x - \mu}{\sigma\sqrt{2}}\right)\right).$$

A la inversa, si Z es una distribución normal estándar, $Z \sim N(0, 1)$, entonces

$$X = \sigma Z + \mu$$

es una variable aleatoria normal tipificada de media μ y varianza σ^2 .

La distribución normal estándar está tabulada (habitualmente en la forma de el valor de la función de distribución Φ) y las otras distribuciones normales pueden obtenerse como transformaciones simples, como se describe más arriba, de la distribución estándar. De este modo se pueden usar los valores tabulados de la función de distribución normal estándar para encontrar valores de la función de distribución de cualquier otra distribución normal.

Momentos

Los primeros momentos de la distribución normal son:

Número	Momento	Momento central	Cumulante
0	1	1	
1	μ	0	μ
2	$\mu^2 + \sigma^2$	σ^2	σ^2
3	$\mu^3 + 3\mu\sigma^2$	0	0
4	$\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$	$3\sigma^4$	0
5	$\mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4$	0	0
6	$\mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6$	$15\sigma^6$	0
7	$\mu^7 + 21\mu^5\sigma^2 + 105\mu^3\sigma^4 + 105\mu\sigma^6$	0	0
8	$\mu^8 + 28\mu^6\sigma^2 + 210\mu^4\sigma^4 + 420\mu^2\sigma^6 + 105\sigma^8$	$105\sigma^8$	0

Todos los cumulantes de la distribución normal, más allá del segundo, son cero.

Los momentos centrales de orden superior ($2k$ con $\mu = 0$) vienen dados por la fórmula

$$E[X^{2k}] = \frac{(2k)!}{2^k k!} \sigma^{2k}.$$

El Teorema del Límite Central

El Teorema del límite central establece que bajo ciertas condiciones (como pueden ser independientes e idénticamente distribuidas con varianza finita), la suma de un gran número de variables aleatorias se distribuye aproximadamente como una normal.

La importancia práctica del Teorema del límite central es que la función de distribución de la normal puede usarse como aproximación de algunas otras funciones de distribución. Por ejemplo:

- Una distribución binomial de parámetros n y p es aproximadamente normal para grandes valores de n , y p no demasiado cercano a 1 ó 0 (algunos

libros recomiendan usar esta aproximación sólo si np y $n(1-p)$ son ambos, al menos, 5; en este caso se debería aplicar una corrección de continuidad).

La normal aproximada tiene parámetros $\mu = np$, $\sigma^2 = np(1-p)$.

- Una distribución de Poisson con parámetro λ es aproximadamente normal para grandes valores de λ . La distribución normal aproximada tiene parámetros $\mu = \sigma^2 = \lambda$.

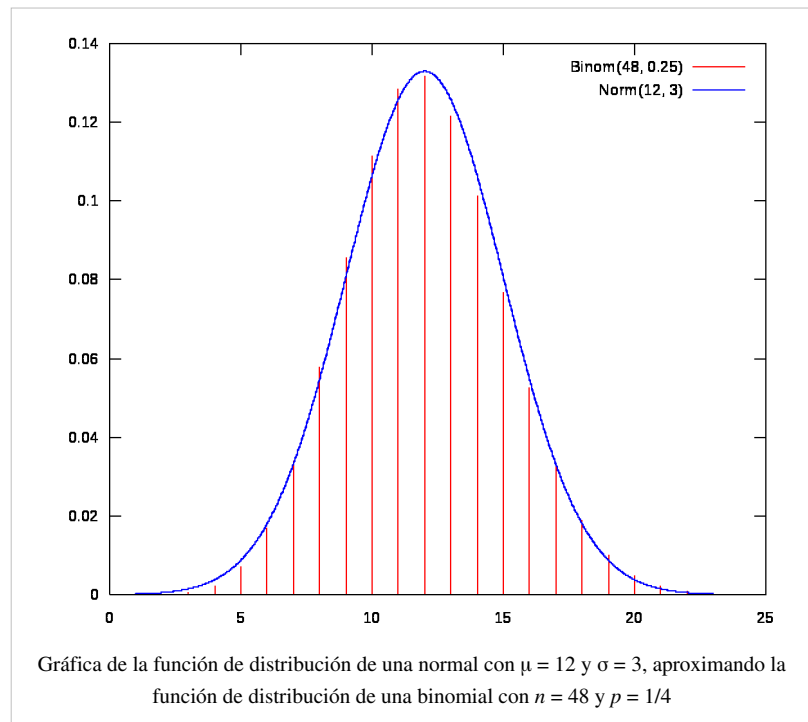
La exactitud de estas aproximaciones depende del propósito para el que se necesiten y de la tasa de convergencia a la distribución normal. Se da el caso típico de que tales aproximaciones son menos precisas en las colas de la distribución. El Teorema de Berry-Esséen proporciona un límite superior general del error de aproximación de la función de distribución.

Divisibilidad infinita

Las normales tienen una distribución de probabilidad infinitamente divisible: Para una distribución normal X de media μ y varianza $\sigma^2 \geq 0$, es posible encontrar n variables aleatorias independientes $\{X_1, \dots, X_n\}$ cada una con distribución normal de media μ/n y varianza σ^2/n dado que la suma $X_1 + \dots + X_n$ de estas n variables aleatorias

$$X_1 + X_2 + \dots + X_n \sim N(\mu, \sigma^2)$$

tenga esta específica distribución normal (para verificarlo, úsese la función característica de convolución y la inducción matemática).



Estabilidad

Las distribuciones normales son estrictamente estables.

Desviación típica e intervalos de confianza

Alrededor del 68% de los valores de una distribución normal están a una distancia $\sigma < 1$ (desviación típica) de la media, μ ; alrededor del 95% de los valores están a dos desviaciones típicas de la media y alrededor del 99,7% están a tres desviaciones típicas de la media. Esto se conoce como la "regla 68-95-99,7" o la "regla empírica".

Para ser más precisos, el área bajo la curva campana entre $\mu - n\sigma$ y $\mu + n\sigma$ en términos de la función de distribución normal viene dada por

$$\begin{aligned} & \Phi_{\mu, \sigma^2}(\mu + n\sigma) - \Phi_{\mu, \sigma^2}(\mu - n\sigma) \\ &= \Phi(n) - \Phi(-n) = 2\Phi(n) - 1 = \operatorname{erf}(n/\sqrt{2}), \end{aligned}$$

donde erf es la función error. Con 12 decimales, los valores para los puntos 1-, 2-, hasta 6- σ son:

n	$\operatorname{erf}(n/\sqrt{2})$
1	0,682689492137
2	0,954499736104
3	0,997300203937
4	0,999936657516
5	0,999999426697
6	0,999999998027

La siguiente tabla proporciona la relación inversa de múltiples σ correspondientes a unos pocos valores usados con frecuencia para el área bajo la campana de Gauss. Estos valores son útiles para determinar intervalos de confianza para los niveles especificados basados en una curva normalmente distribuida (o estimadores asintóticamente normales):

$\operatorname{erf}(n/\sqrt{2})$	n
0,80	1,28155
0,90	1,64485
0,95	1,95996
0,98	2,32635
0,99	2,57583
0,995	2,80703
0,998	3,09023
0,999	3,29052
0,9999	3,8906
0,99999	4,4172

donde el valor a la izquierda de la tabla es la proporción de valores que caerán en el intervalo dado y n es un múltiplo de la desviación típica que determina la anchura de el intervalo.

Forma familia exponencial

La distribución normal tiene forma de familia exponencial biparamétrica con dos parámetros naturales, μ y $1/\sigma^2$, y estadísticos naturales X y X^2 . La forma canónica tiene como parámetros $\frac{\mu}{\sigma^2}$ y $\frac{1}{\sigma^2}$ y estadísticos suficientes $\sum x$ y $-\frac{1}{2} \sum x^2$.

Distribución normal compleja

Considérese la variable aleatoria compleja gaussiana

$$Z = X + iY$$

donde X e Y son variables gaussianas reales e independientes con igual varianza σ_r^2 . La función de distribución de la variable conjunta es entonces

$$\frac{1}{2\pi\sigma_r^2} e^{-(x^2+y^2)/(2\sigma_r^2)}.$$

Como $\sigma_Z = \sqrt{2}\sigma_r$, la función de distribución resultante para la variable gaussiana compleja Z es

$$\frac{1}{\pi\sigma_Z^2} e^{-|Z|^2/\sigma_Z^2}.$$

Distribuciones relacionadas

- $R \sim \text{Rayleigh}(\sigma)$ es una distribución de Rayleigh si $R = \sqrt{X^2 + Y^2}$ donde $X \sim N(0, \sigma^2)$ y $Y \sim N(0, \sigma^2)$ son dos distribuciones normales independientes.
- $Y \sim \chi_\nu^2$ es una distribución χ^2 con ν grados de libertad si $Y = \sum_{k=1}^{\nu} X_k^2$ donde $X_k \sim N(0, 1)$ para $k = 1, \dots, \nu$ y son independientes.
- $Y \sim \text{Cauchy}(\mu = 0, \theta = 1)$ es una distribución de Cauchy si $Y = X_1/X_2$ para $X_1 \sim N(0, 1)$ y $X_2 \sim N(0, 1)$ son dos distribuciones normales independientes.
- $Y \sim \text{Log-N}(\mu, \sigma^2)$ es una distribución log-normal si $Y = e^X$ y $X \sim N(\mu, \sigma^2)$.
- Relación con una distribución estable: si $X \sim \text{stable}(2, \beta, \sigma/\sqrt{2}, \mu)$ entonces $X \sim N(\mu, \sigma^2)$.
- Distribución normal truncada. si $X \sim N(\mu, \sigma^2)$, entonces truncando X por debajo de A y por encima de B dará lugar a una variable aleatoria de media $E(X) = \mu + \frac{\sigma(\varphi_1 - \varphi_2)}{T}$, donde $T = \Phi\left(\frac{B - \mu}{\sigma}\right) - \Phi\left(\frac{A - \mu}{\sigma}\right)$, $\varphi_1 = \varphi\left(\frac{A - \mu}{\sigma}\right)$, $\varphi_2 = \varphi\left(\frac{B - \mu}{\sigma}\right)$ y φ es la función de densidad de una variable normal estándar.
- Si X es una variable aleatoria normalmente distribuida e $Y = |X|$, entonces Y tiene una distribución normal doblada.

Estadística descriptiva e inferencial

Resultados

De la distribución normal se derivan muchos resultados, incluyendo rangos de percentiles ("percentiles" o "cuantiles"), curvas normales equivalentes, stanines, z-scores, y T-scores. Además, un número de procedimientos de estadísticos de comportamiento están basados en la asunción de que esos resultados están normalmente distribuidos. Por ejemplo, el test de Student y el análisis de varianza (ANOVA) (véase más abajo). La gradación de la curva

campana asigna grados relativos basados en una distribución normal de resultados.

Tests de normalidad

Los tests de normalidad se aplican a conjuntos de datos para determinar su similitud con una distribución normal. La hipótesis nula es, en estos casos, si el conjunto de datos es similar a una distribución normal, por lo que un P-valor suficientemente pequeño indica datos no normales.

- Prueba de Kolmogórov-Smirnov
- Test de Lilliefors
- Test de Anderson–Darling
- Test de Ryan–Joiner
- Test de Shapiro–Wilk
- Normal probability plot (rankit plot)
- Test de Jarque–Bera
- Test omnibús de Spiegelhalter

Estimación de parámetros

Estimación de parámetros de máxima verosimilitud

Véase también: *Máxima verosimilitud*

Supóngase que

$$X_1, \dots, X_n$$

son independientes y cada una está normalmente distribuida con media μ y varianza $\sigma^2 > 0$. En términos estadísticos los valores observados de estas n variables aleatorias constituyen una "muestra de tamaño n de una población normalmente distribuida. Se desea estimar la media poblacional μ y la desviación típica poblacional σ , basándose en las valores observados de esta muestra. La función de densidad conjunta de estas n variables aleatorias independientes es

$$\begin{aligned} f(x_1, \dots, x_n; \mu, \sigma) &= \prod_{i=1}^n \varphi_{\mu, \sigma^2}(x_i) \\ &= \frac{1}{(\sigma\sqrt{2\pi})^n} \prod_{i=1}^n \exp\left(-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right), \quad (x_1, \dots, x_n) \in \mathbb{R}^n. \end{aligned}$$

Como función de μ y σ , la función de verosimilitud basada en las observaciones X_1, \dots, X_n es

$$L(\mu, \sigma) = \frac{C}{\sigma^n} \exp\left(-\frac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^2}\right), \quad \mu \in \mathbb{R}, \sigma > 0,$$

con alguna constante $C > 0$ (de la cual, en general, se permitiría incluso que dependiera de X_1, \dots, X_n , aunque desapareciera con las derivadas parciales de la función de log-verosimilitud respecto a los parámetros tenidos en cuenta, véase más abajo).

En el método de máxima verosimilitud, los valores de μ y σ que maximizan la función de verosimilitud se toman como estimadores de los parámetros poblacionales μ y σ .

Habitualmente en la maximización de una función de dos variables, se podrían considerar derivadas parciales. Pero aquí se explota el hecho de que el valor de μ que maximiza la función de verosimilitud con σ fijo no depende de σ . No obstante, encontramos que ese valor de μ , entonces se sustituye por μ en la función de verosimilitud y finalmente encontramos el valor de σ que maximiza la expresión resultante.

Es evidente que la función de verosimilitud es una función decreciente de la suma

$$\sum_{i=1}^n (X_i - \mu)^2.$$

Así que se desea el valor de μ que *minimiza* esta suma. Sea

$$\bar{X}_n = (X_1 + \cdots + X_n)/n$$

la media muestral basada en las n observaciones. Nótese que

$$\begin{aligned} \sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n ((X_i - \bar{X}_n) + (\bar{X}_n - \mu))^2 \\ &= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + 2(\bar{X}_n - \mu) \underbrace{\sum_{i=1}^n (X_i - \bar{X}_n)}_{=0} + \sum_{i=1}^n (\bar{X}_n - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2. \end{aligned}$$

Sólo el último término depende de μ y se minimiza por

$$\hat{\mu}_n = \bar{X}_n.$$

Esta es la estimación de máxima verosimilitud de μ basada en las n observaciones X_1, \dots, X_n . Cuando sustituimos esta estimación por μ en la función de verosimilitud, obtenemos

$$L(\bar{X}_n, \sigma) = \frac{C}{\sigma^n} \exp\left(-\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{2\sigma^2}\right), \quad \sigma > 0.$$

Se conviene en denotar la "log-función de verosimilitud", esto es, el logaritmo de la función de verosimilitud, con una minúscula ℓ , y tenemos

$$\ell(\bar{X}_n, \sigma) = \log C - n \log \sigma - \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{2\sigma^2}, \quad \sigma > 0,$$

entonces

$$\begin{aligned} \frac{\partial}{\partial \sigma} \ell(\bar{X}_n, \sigma) &= -\frac{n}{\sigma} + \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\sigma^3} \\ &= -\frac{n}{\sigma^3} \left(\sigma^2 - \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right), \quad \sigma > 0. \end{aligned}$$

Esta derivada es positiva, cero o negativa según σ^2 esté entre 0 y

$$\hat{\sigma}_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

o sea igual a esa cantidad, o mayor que esa cantidad. (Si hay solamente una observación, lo que significa que $n = 1$, o si $X_1 = \dots = X_n$, lo cual sólo ocurre con probabilidad cero, entonces $\hat{\sigma}_n^2 = 0$ por esta fórmula, refleja el hecho de que en estos casos la función de verosimilitud es ilimitada cuando σ decrece hasta cero.)

Consecuentemente esta media de cuadrados de residuos es el estimador de máxima verosimilitud de σ^2 , y su raíz cuadrada es el estimador de máxima verosimilitud de σ basado en las n observaciones. Este estimador $\hat{\sigma}_n^2$ es sesgado, pero tiene un menor error medio al cuadrado que el habitual estimador insesgado, que es $n/(n-1)$ veces este estimador.

Sorprendente generalización

La derivada del estimador de máxima verosimilitud de la matriz de covarianza de una distribución normal multivariante es despreciable. Involucra el teorema espectral y la razón por la que puede ser mejor para ver un escalar como la traza de una matriz 1×1 matrix que como un mero escalar. Véase estimación de la covarianza de matrices.

Estimación insesgada de parámetros

El estimador \bar{X} de máxima verosimilitud de la media poblacional μ , es un estimador insesgado de la media poblacional.

El estimador de máxima verosimilitud de la varianza es insesgado si asumimos que la media de la población es conocida *a priori*, pero en la práctica esto no ocurre. Cuando disponemos de una muestra y no sabemos nada de la media o la varianza de la población de la que se ha extraído, como se asumía en la derivada de máxima verosimilitud de arriba, entonces el estimador de máxima verosimilitud de la varianza es sesgado. Un estimador insesgado de la varianza σ^2 es la cuasi varianza muestral:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

que sigue una distribución Gamma cuando las X_i son normales independientes e idénticamente distribuidas:

$$S^2 \sim \text{Gamma} \left(\frac{n-1}{2}, \frac{2\sigma^2}{n-1} \right),$$

con media $E(S^2) = \sigma^2$ y varianza $\text{Var}(S^2) = 2\sigma^4/(n-1)$.

La estimación de máxima verosimilitud de la desviación típica es la raíz cuadrada de la estimación de máxima verosimilitud de la varianza. No obstante, ni ésta, ni la raíz cuadrada de la cuasivarianza muestral proporcionan un estimador insesgado para la desviación típica (véase estimación insesgada de la desviación típica para una fórmula particular para la distribución normal).

Incidencia

Las distribuciones *aproximadamente* normales aparecen por doquier, como queda explicado por el teorema central del límite. Cuando en un fenómeno se sospecha la presencia de un gran número de pequeñas causas *actuando de forma aditiva e independiente* es razonable pensar que las observaciones serán "normales". Hay métodos estadísticos para probar empíricamente esta asunción, por ejemplo, el test de Kolmogorov-Smirnov.

Hay causas que pueden actuar de forma *multiplicativa* (más que aditiva). En este caso, la asunción de normalidad no está justificada y es el logaritmo de la variable en cuestión el que estaría normalmente distribuido. La distribución de las variables directamente observadas en este caso se denomina log-normal.

Finalmente, si hay una simple influencia externa que tiene un gran efecto en la variable en consideración, la asunción de normalidad no está tampoco justificada. Esto es cierto incluso si, cuando la variable externa se mantiene constante, las distribuciones marginales resultantes son, en efecto, normales. La distribución completa será una superposición de variables normales, que no es en general normal. Ello está relacionado con la teoría de errores (véase más abajo).

A continuación se muestran una lista de situaciones que estarían, aproximadamente, normalmente distribuidas. Más abajo puede encontrarse una discusión detallada de cada una de ellas:

- En problemas de recuento, donde el teorema central del límite incluye una aproximación de discreta a continua y donde las distribuciones infinitamente divisibles y descomponibles están involucradas, tales como:
 - variables aleatorias binomiales, asociadas con preguntas sí/no;
 - variables aleatorias de Poisson, asociadas con eventos raros;

- En medidas fisiológicas de especímenes biológicos:
 - El *logaritmo* de las medidas del tamaño de tejidos vivos (longitud, altura, superficie de piel, peso);
 - La *longitud* de apéndices *inertes* (pelo, garras, rabos, dientes) de especímenes biológicos *en la dirección del crecimiento*;
 - Otras medidas fisiológicas podrían estar normalmente distribuidas, aunque no hay razón para esperararlo *a priori*;
- Se asume con frecuencia que los errores de medida están normalmente distribuidos y cualquier desviación de la normalidad se considera una cuestión que debería explicarse;
- Variables financieras, en el modelo Black-Scholes:
 - Cambios en el *logaritmo* de

Cambios en el *logaritmo* de tasas de cambio, índices de precios, índices de existencias de mercado; estas variables se comportan como el interés compuesto, no como el interés simple, por tanto, son multiplicativas;

- Mientras que el modelo Black-Scholes presupone normalidad, en realidad estas variables exhiben colas pesadas, como puede verse en crash de las existencias de mercado;
- Otras variables financieras podrían estar normalmente distribuidas, pero no hay razón para esperararlo *a priori*;
- Intensidad de la luz:
 - La intensidad de la luz láser está normalmente distribuida;
 - La luz térmica tiene una distribución de Bose-Einstein en escalas de tiempo muy breves y una distribución normal en grandes escalas de tiempo debido al teorema central del límite.

Es relevante para la biología y la economía el hecho de que los sistemas complejos tienden a mostrar la ley de potencias más que normal.

Recuento de fotones

La intensidad de la luz de una sola fuente varía con el tiempo, así como las fluctuaciones térmicas que pueden observarse si la luz se analiza a una resolución suficientemente alta. La mecánica cuántica interpreta las medidas de la intensidad de la luz como un recuento de fotones, donde la asunción natural es usar la distribución de Poisson. Cuando la intensidad de la luz se integra a lo largo de grandes periodos de tiempo mayores que el tiempo de coherencia, la aproximación Poisson - Normal es apropiada.

Medida de errores

La normalidad es la *asunción central* de la teoría matemática de errores. De forma similar en el ajuste de modelos estadístico, un indicador de la bondad del ajuste es que el error residual (así es como se llaman los errores en esta circunstancia) sea independiente y normalmente distribuido. La asunción es que cualquier desviación de la normalidad necesita ser explicada. En ese sentido, en ambos, ajuste de modelos y teoría de errores, la normalidad es la única observación que no necesita ser explicada, sino que es esperada. No obstante, si los datos originales no están normalmente distribuidos (por ejemplo, si siguen una distribución de Cauchy, entonces los residuos tampoco estarán normalmente distribuidos. Este hecho es ignorado habitualmente en la práctica.

Las medidas repetidas de la misma cantidad se espera que cedan el paso a resultados que están agrupados entorno a un valor particular. Si todas las fuentes principales de errores se han tomado en cuenta, se *asume* que el error que queda debe ser el resultado de un gran número de muy pequeños y *aditivos* efectos y, por consiguiente, normal. Las desviaciones de la normalidad se interpretan como indicaciones de errores sistemáticos que no han sido tomados en cuenta. Puede debatirse si esta asunción es válida.

Una famosa observación atribuida a Gabriel Lippmann dice:^[cita requerida]

Todo el mundo cree en la ley normal de los errores: los matemáticos, porque piensan que es un hecho experimental; y los experimentadores, porque suponen que es un teorema matemático

Otra fuente podría ser Henri Poincaré ^[10].

Características físicas de especímenes biológicos

Los tamaños de los animales adultos siguen aproximadamente una distribución log-normal. La evidencia y explicación basada en modelos de crecimiento fue publicada por primera vez en el libro *Problemas de crecimiento relativo*, de 1932, por Julian Huxley.

Las diferencias de tamaño debido a dimorfismos sexuales u otros polimorfismos de insectos, como la división social de las abejas en obreras, zánganos y reinas, por ejemplo, hace que la distribución de tamaños se desvíe hacia la lognormalidad.

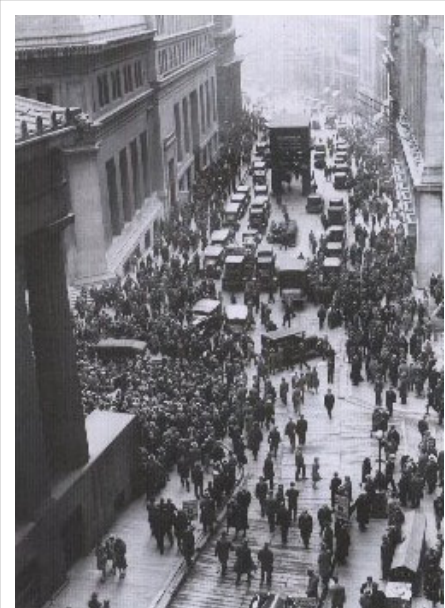
La asunción de que el tamaño lineal de los especímenes biológicos es normal (más que lognormal) nos lleva a una distribución no normal del peso (puesto que el peso o el volumen es proporcional al cuadrado o el cubo de la longitud y las distribuciones gaussianas sólo mantienen las transformaciones lineales). A la inversa, asumir que el peso sigue una distribución normal implica longitudes no normales. Esto es un problema porque, *a priori*, no hay razón por la que cualquiera de ellas (longitud, masa corporal u otras) debería estar normalmente distribuida. Las distribuciones lognormales, por otro lado, se mantienen entre potencias, así que el "problema" se desvanece si se asume la lognormalidad.

Por otra parte, hay algunas medidas biológicas donde se asume normalidad, tales como la presión sanguínea en humanos adultos. Esta asunción sólo es posible tras separar a hombres y mujeres en distintas poblaciones, cada una de las cuales está normalmente distribuida.

Variables financieras

Ya en 1900 Louis Bachelier propuso representar los precios de cambio usando la distribución normal. Esta aproximación se ha modificado desde entonces ligeramente. A causa de la naturaleza multiplicativa del interés compuesto, los indicadores financieros como valores de mercado y precios de las materias primas exhiben un "comportamiento multiplicativo". Como tales, sus cambios periódicos (por ejemplo, cambios anuales) no son normales, sino lognormales. Esta es todavía la hipótesis más comúnmente aceptada en economía.

No obstante, en realidad las variables financieras exhiben colas pesadas y así, la asunción de normalidad infravalora la probabilidad de eventos extremos como quiebras financieras. Se han sugerido correcciones a este modelo por parte de matemáticos como Benoît Mandelbrot, quien observó que los cambios en el logaritmo durante breves periodos de tiempo (como un día) se aproximan bien por distribuciones que no tienen una varianza finita y, por consiguiente, el teorema central del límite no puede aplicarse. Más aún, la suma de muchos de tales cambios sigue una distribución de log-Levy.



El modelo normal de movimiento de activos no incluye movimientos extremos tales como quiebras financieras.

Distribuciones en tests de inteligencia

A veces, la dificultad y número de preguntas en un test de inteligencia se selecciona de modo que proporcionen resultados normalmente distribuidos. Más aún, las puntuaciones "en crudo" se convierten a valores que marcan el cociente intelectual ajustándolas a la distribución normal. En cualquier caso se trata de un resultado causado deliberadamente por la construcción del test o de una interpretación de las puntuaciones que sugiere normalidad para la mayoría de la población. Sin embargo, la cuestión acerca de si la inteligencia en sí está normalmente distribuida es

más complicada porque se trata de una variable latente y, por consiguiente, no puede observarse directamente.

Ecuación de difusión

La función de densidad de la distribución normal está estrechamente relacionada con la ecuación de difusión (homogénea e isotrópica) y, por tanto, también con la ecuación de calor. Esta ecuación diferencial parcial describe el tiempo de evolución de una función de densidad bajo difusión. En particular, la función de densidad de masa

$$\varphi_{0,t}(x) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right),$$

para la distribución normal con esperanza 0 y varianza t satisface la ecuación de difusión:

$$\frac{\partial}{\partial t} \varphi_{0,t}(x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} \varphi_{0,t}(x).$$

Si la densidad de masa para un tiempo $t = 0$ viene dada por la delta de Dirac, lo cual significa, esencialmente que toda la masa está inicialmente concentrada en un punto, entonces la función de densidad de masa en el tiempo t tendrá la forma de la función de densidad de la normal, con varianza creciendo linealmente con t . Esta conexión no es coincidencia: la difusión se debe a un movimiento Browniano que queda descrito matemáticamente por un proceso de Wiener, y tal proceso en un tiempo t también resultará normal con varianza creciendo linealmente con t .

Más generalmente, si la densidad de masa inicial viene dada por una función $\varphi(x)$, entonces la densidad de masa en un tiempo t vendrá dada por la convolución de φ y una función de densidad normal.

Uso en estadística computacional

Generación de valores para una variable aleatoria normal

Para simulaciones por ordenador es útil, en ocasiones, generar valores que podrían seguir una distribución normal. Hay varios métodos y el más básico de ellos es invertir la función de distribución de la normal estándar. Se conocen otros métodos más eficientes, uno de los cuales es la transformación de Box-Muller. Un algoritmo incluso más rápido es el algoritmo zigurat. Ambos se discuten más abajo. Una aproximación simple a estos métodos es programarlos como sigue: simplemente súmense 12 desviaciones uniformes (0,1) y réstense 6 (la mitad de 12). Esto es bastante útil en muchas aplicaciones. La suma de esos 12 valores sigue la distribución de Irwin-Hall; son elegidos 12 para dar a la suma una varianza de uno, exactamente. Las desviaciones aleatorias resultantes están limitadas al rango $(-6, 6)$ y tienen una densidad que es una doceava sección de una aproximación polinomial de undécimo orden a la distribución normal.^[11]

El método de Box-Muller dice que, si tienes dos números aleatorios U y V uniformemente distribuidos en $(0, 1]$, (por ejemplo, la salida de un generador de números aleatorios), entonces X e Y son dos variables aleatorias estándar normalmente distribuidas, donde:

$$Y = \sqrt{-2 \ln U} \sin(2\pi V).$$

Esta formulación aparece porque la distribución χ^2 con dos grados de libertad (véase la propiedad 4, más arriba) es una variable aleatoria exponencial fácilmente generada (la cual corresponde a la cantidad $\ln U$ en estas ecuaciones). Así, un ángulo elegido uniformemente alrededor de un círculo vía la variable aleatoria V y un radio elegido para ser exponencial se transforman entonces en coordenadas x e y normalmente distribuidas.

Un método mucho más rápido que la transformación de Box-Muller, pero que sigue siendo exacto es el llamado algoritmo Zigurat, desarrollado por George Marsaglia. En alrededor del 97% de los casos usa sólo dos números aleatorios, un entero aleatorio y un uniforme aleatorio, una multiplicación y un test-si. Sólo un 3% de los casos donde la combinación de estos dos cae fuera del "corazón del zigurat", un tipo de rechazo muestral usando logaritmos, exponenciales y números aleatorios más uniformes deberían ser empleados.

Hay también alguna investigación sobre la conexión entre la rápida transformación de Hadamard y la distribución normal, en virtud de que la transformación emplea sólo adición y sustracción y por el teorema central del límite los números aleatorios de casi cualquier distribución serán transformados en la distribución normal. En esta visión se pueden combinar una serie de transformaciones de Hadamard con permutaciones aleatorias para devolver conjuntos de datos aleatorios normalmente distribuidos.

Aproximaciones numéricas de la distribución normal y su función de distribución

La función de distribución normal se usa extensamente en computación científica y estadística. Por consiguiente, ha sido implementada de varias formas.

Abramowitz y Stegun (1964) dan la conocida como "mejor aproximación de Hastings" para $\Phi(x)$ con $x > 0$ con un error absoluto $|\varepsilon(x)| < 7.5 \cdot 10^{-8}$ (algoritmo 26.2.17 [12]):

$$\Phi(x) = 1 - \phi(x) \left(b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5 \right) + \varepsilon(x), \quad t = \frac{1}{1 + b_0 x},$$

donde $\phi(x)$ es la función de densidad de la distribución normal estándar,

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

y las constantes son $b_0 = 0.2316419$, $b_1 = 0.319381530$, $b_2 = -0.356563782$, $b_3 = 1.781477937$, $b_4 = -1.821255978$, $b_5 = 1.330274429$.

La Biblioteca Científica GNU calcula valores de la función de distribución normal estándar usando aproximaciones por funciones racionales a trozos. Otro método de aproximación usa polinomios de tercer grado en intervalos.^[13] El artículo sobre el lenguaje de programación bc proporciona un ejemplo de cómo computar la función de distribución en GNU bc.

Para una discusión más detallada sobre cómo calcular la distribución normal, véase la sección 3.4.1C. de *The Art of Computer Programming (El arte de la programación por ordenador)*, de Knuth.

Véase también

- Carl Friedrich Gauss.
- Cociente intelectual.
- Desenfoque gaussiano y convolución usando la distribución normal como núcleo.
- Distribución χ^2 .
- Distribución de Pearson Familia general de distribuciones de probabilidad que extienden la distribución gaussiana para incluir diferentes valores de asimetrías y curtosis.
- Distribución gaussiana inversa
- Distribución logística
- Distribución log-normal
- Distribución Normal-gamma
- Distribución normal matriz
- Distribución normal multivariante
- Distribución normal torcida
- Distribución normal truncada
- Distribución t de Student
- Distribuciones Tweedie
- Función gaussiana (campana de Gauss).
- Función logit.
- Función probit
- Gradación de la curva campana

- Iannis Xenakis, distribución gaussiana en música.
- Integral del Gauss
- Normalmente distribuidas e incorreladas no implica independencia (un ejemplo de dos variables aleatorias normalmente distribuidas e incorreladas que no son independientes; esto no ocurriría en presencia de una distribución conjunta).
- Problema de Behrens–Fisher
- Proceso de Gauss
 - Puente Browniano
 - Proceso de Ornstein-Uhlenbeck
 - Proceso de Wiener
- Prueba de Mann-Whitney. Método no paramétrico que aproxima a una normal.
- Suma de variables aleatorias normalmente distribuidas
- Tabla distribución normal tipificada
- Tamaño de la muestra
- Teorema Central del Límite.
- Transformación de datos (Estadística). Técnicas simples de transformación de datos en una distribución normal.
- Teorema de Erdős-Kac, sobre la ocurrencia de la distribución normal en teoría de números.

Referencias

- [1] Es una consecuencia del Teorema Central del Límite
- [2] Abraham de Moivre, "Approximatio ad Summam Terminorum Binomii $(a + b)^n$ in Seriem expansi" (impreso el 12 de noviembre de 1733 en Londres para una edición privada). Este panfleto se reimprimió en: (1) Richard C. Archibald (1926) "A rare pamphlet of Moivre and some of his discoveries," *Isis*, vol. 8, páginas 671-683; (2) Helen M. Walker, "De Moivre on the law of normal probability" en David Eugene Smith, *A Source Book in Mathematics* [Nueva York, Nueva York: McGraw-Hill, 1929; reimpresión: Nueva York, Nueva York: Dover, 1959], vol. 2, páginas 566-575.; (3) Abraham De Moivre, *The Doctrine of Chances* (2ª ed.) [Londres: H. Woodfall, 1738; reimpresión: Londres: Cass, 1967], páginas 235-243; (3ª ed.) [Londres: A Millar, 1756; reimpresión: Nueva York, Nueva York: Chelsea, 1967], páginas 243-254; (4) Florence N. David, *Games, Gods and Gambling: A History of Probability and Statistical Ideas* [Londres: Griffin, 1962], Apéndice 5, páginas 254-267.
- [3] Haviil, 2003
- [4] Wussing, Hans (marzo de 1998). « Lección 10 (http://books.google.es/books?id=IG3_b5Xm8PMC)». *Lecciones de Historia de las Matemáticas* (1ª (castellano) edición). Siglo XXI de España Editores, S.A.. pp. 190. ISBN 84-323-0966-4. . «"La distribución normal y sus aplicaciones a la teoría de errores se asocia a menudo con el nombre de Gauss, quien la descubrió -igual que Laplace- independientemente; no obstante ya había sido estudiada por de Moivre»
- [5] Weisstein, Eric W., « Normal Distribution (<http://mathworld.wolfram.com/NormalDistribution.html>)» (en inglés), *MathWorld*, Wolfram Research, , consultado el 18 de marzo.
- [6] La función Q (<http://cnx.org/content/m11537/latest/>)
- [7] <http://www.eng.tau.ac.il/~jo/academic/Q.pdf>
- [8] Weisstein, Eric W., « Normal Distribution Function (<http://mathworld.wolfram.com/NormalDistributionFunction.html>)» (en inglés), *MathWorld*, Wolfram Research, .
- [9] M.A. Sanders. « Characteristic function of the univariate normal distribution (<http://www.planetmathematics.com/CharNormal.pdf>)». Consultado el 06-03-2009.
- [10] http://en.wikiquote.org/wiki/Henri_Poincaré#Misattributed
- [11] Johnson NL, Kotz S, Balakrishnan N. (1995) Continuous Univariate Distributions Volume 2, Wiley. Equation(26.48)
- [12] http://www.math.sfu.ca/~cbm/aands/page_932.htm
- [13] Andy Salter. « B-Spline curves (<http://www.doc.ic.ac.uk/~dfg/AndysSplineTutorial/BSplines.html>)». Consultado el 05-12-2008.

Enlaces externos

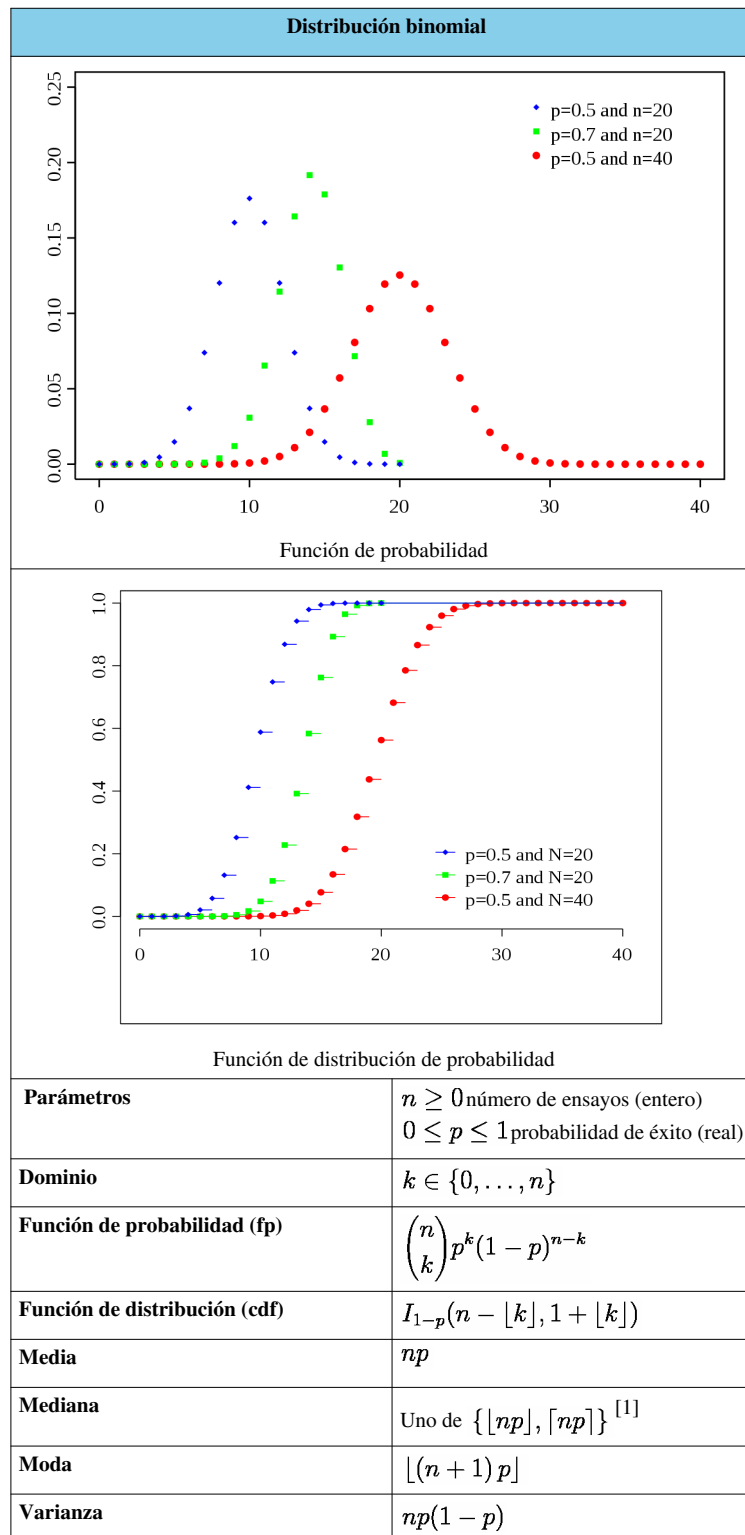
- Áreas bajo la curva normal (<http://www.digitalreview.com.ar/distribucionnormal/>) Tabla conteniendo los valores de la función normal
- Calculadora de probabilidades en una distribución Normal (<http://www.ugr.es/~jsalinas/normal.htm>). Permite hacer cálculos directos e inversos.
- Tabla de la distribución normal (http://www.vaxasoftware.com/doc_edu/mat/dnormal.pdf) Tabla de la distribución normal en formato PDF

Se puede usar software y un programa de computadora para el ajuste de una distribución de probabilidad, incluyendo la normal, a una serie de datos:

- Easy fit (http://www.mathwave.com/articles/distribution_fitting.html), "data analysis & simulation"
- MathWorks Benelux (<http://www.mathworks.nl/products/statistics/demos.html?file=/products/demos/shipping/stats/cfitdfitdemo.html>)
- ModelRisk (<http://www.vosesoftware.com/>), "risk modelling software"
- Ricci distributions, fitting distributions with R (<http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf>), Vito Ricci, 2005
- Risksolver, automatically fit distributions and parameters to samples (<http://www.solver.com/risksolver8.htm>)
- StatSoft distribution fitting (<http://www.statsoft.com/textbook/distribution-fitting/>)
- CumFreq (<http://www.waterlog.info/cumfreq.htm>), libre sin costo, incluye la distribución normal, la lognormal, raíz-normal, cuadrado-normal, e intervalos de confianza a base de la distribución binomial
- Calculadora Distribución normal (<http://www.stud.feec.vutbr.cz/~xvapen02/vypocty/no.php?language=espanol>)
- (<http://cajael.com/mestadisticos/T7DContinuas/node3.php>) Calcular la probabilidad de una distribución normal con R (lenguaje de programación)

Distribuciones Multivariantes

Distribución binomial



Coefficiente de simetría	$\frac{1 - 2p}{\sqrt{np(1 - p)}}$
Curtosis	$\frac{1 - 6p(1 - p)}{np(1 - p)}$
Entropía	$\frac{1}{2} \ln(2\pi nep(1 - p)) + O\left(\frac{1}{n}\right)$
Función generadora de momentos (mgf)	$(1 - p + pe^t)^n$
Función característica	$(1 - p + pe^{it})^n$

En estadística, la **distribución binomial** es una distribución de probabilidad discreta que mide el número de éxitos en una secuencia de n ensayos independientes de Bernoulli con una probabilidad fija p de ocurrencia del éxito entre los ensayos.

Un experimento de Bernoulli se caracteriza por ser dicotómico, esto es, sólo son posibles dos resultados. A uno de estos se denomina éxito y tiene una probabilidad de ocurrencia p y al otro, fracaso, con una probabilidad $q = 1 - p$. En la distribución binomial el anterior experimento se repite n veces, de forma independiente, y se trata de calcular la probabilidad de un determinado número de éxitos. Para $n = 1$, la binomial se convierte, de hecho, en una distribución de Bernoulli.

Para representar que una variable aleatoria X sigue una distribución binomial de parámetros n y p , se escribe:

$$X \sim B(n, p)$$

La distribución binomial es la base del test binomial de significación estadística.

Ejemplos

Las siguientes situaciones son ejemplos de experimentos que pueden modelizarse por esta distribución:

- Se lanza un dado diez veces y se cuenta el número X de treses obtenidos: entonces $X \sim B(10, 1/6)$
- Se lanza una moneda dos veces y se cuenta el número X de caras obtenidas: entonces $X \sim B(2, 1/2)$
- Una partícula se mueve unidimensionalmente con probabilidad q de moverse de aquí para allá y $1 - q$ de moverse de allá para acá

Experimento Binomial

Existen muchas situaciones en las que se presenta una experiencia binomial. Cada uno de los experimentos es independiente de los restantes (la probabilidad del resultado de un experimento no depende del resultado del resto). El resultado de cada experimento ha de admitir sólo dos categorías (a las que se denomina éxito y fracaso). Las probabilidades de ambas posibilidades han de ser constantes en todos los experimentos (se denotan como p y q o p y $1 - p$).

Se designa por X a la variable que mide el número de éxitos que se han producido en los n experimentos.

Cuando se dan estas circunstancias, se dice que la variable X sigue una distribución de probabilidad binomial, y se denota $B(n, p)$.

Características analíticas

Su función de probabilidad es

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

donde $x = \{0, 1, 2, \dots, n\}$,

siendo $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ las combinaciones de n en x (n elementos tomados de x en x)

Ejemplo

Supongamos que se lanza un dado 50 veces y queremos la probabilidad de que el número 3 salga 20 veces. En este caso tenemos una $X \sim B(50, 1/6)$ y la probabilidad sería $P(X=20)$:

$$P(X = 20) = \binom{50}{20} (1/6)^{20} (1 - 1/6)^{50-20}$$

Propiedades características

$$\mathbb{E}[X] = np$$

$$\text{Var}[X] = np(1-p)$$

Relaciones con otras variables aleatorias

Si n tiende a infinito y p es tal que producto entre ambos parámetros tiende a λ , entonces la distribución de la variable aleatoria binomial tiende a una distribución de Poisson de parámetro λ .

Por último, se cumple que cuando n es muy grande (usualmente se exige que $n \geq 30$) la distribución binomial puede aproximarse mediante la distribución normal.

Propiedades reproductivas

Dadas n variables binomiales independientes, de parámetros n_i ($i = 1, \dots, n$) y p , su suma es también una variable binomial, de parámetros $n_1 + \dots + n_n$, y p , es decir,

$$Y = \sum_{i=1}^n X_i \sim B\left(\sum_{i=1}^n n_i, p\right)$$

Referencias

- [1] Hamza, K. (1995). The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions. *Statist. Probab. Lett.* 23 21–25.

Enlaces externos

- Artículo con ejemplos sobre la distribución binomial (http://www.virtual.unal.edu.co/cursos/ciencias/2001065/html/un2/distribucion_binomial.html) en la Universidad Nacional de Colombia
- Calculadora Distribución binomial (<http://www.stud.feec.vutbr.cz/~xvapen02/vypocty/bi.php?language=espanol>)
- (<http://cajael.com/mestadisticos/T6DDiscretas/node2.php>) Cálculo de la probabilidad de una distribución binomial con R (lenguaje de programación)

[[sl:Binomska po

Distribución multinomial

Multinomial	
Parámetros	$n > 0$ (numero de pruebas (entero)) p_1, \dots, p_k probabilidad de un suceso concreto ($\sum p_i = 1$)
Dominio	$X_i \in \{0, \dots, n\}$ $\sum X_i = n$
Función de densidad (pdf)	$\frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$
Media	$E\{X_i\} = np_i$
Varianza	$\text{Var}(X_i) = np_i(1 - p_i)$ $\text{Cov}(X_i, X_j) = -np_i p_j \quad (i \neq j)$
Función generadora de momentos (mgf)	$\left(\sum_{i=1}^k p_i e^{t_i} \right)^n$

En teoría de probabilidad, la distribución multinomial es una generalización de la distribución binomial.

La distribución binomial es la probabilidad de un número de éxitos en N sucesos de Bernoulli independientes, con la misma probabilidad de éxito en cada suceso. En una distribución multinomial, el análogo a la distribución de Bernoulli es la distribución categórica, donde cada suceso concluye en únicamente un resultado de un numero finito K de los posibles, con probabilidades p_1, \dots, p_k (tal que $p_i \geq 0$ para i entre 1 y K y $\sum_{i=1}^k p_i = 1$); y con n sucesos independientes.

Entonces sea la variable aleatoria x_i , que indica el número de veces que el se ha dado el resultado i sobre los n sucesos. El vector $x=(x_1, \dots, x_k)$ sigue una distribución multinomial con parámetros n y p , donde $p = (p_1, \dots, p_k)$.

Nótese que en algunos campos las distribuciones categórica y multinomial se encuentran unidas, y es común hablar de una distribución multinomial cuando el término más preciso sería una distribución categórica.

Especificación

Función de probabilidad

La función de probabilidad de la distribución multinomial es como sigue:

$$f(x_1, \dots, x_k; n, p_1, \dots, p_k) = \Pr(X_1 = x_1 \text{ y } \dots \text{ y } X_k = x_k)$$

$$= \begin{cases} \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}, & \text{cuando } \sum_{i=1}^k x_i = n \\ 0 & \text{En otros casos,} \end{cases}$$

Para enteros no negativos x_1, \dots, x_k .

Propiedades

La esperanza matemática del suceso i observado en n pruebas es:

$$E(X_i) = np_i.$$

La varianza es:

$$\text{var}(X_i) = np_i(1 - p_i).$$

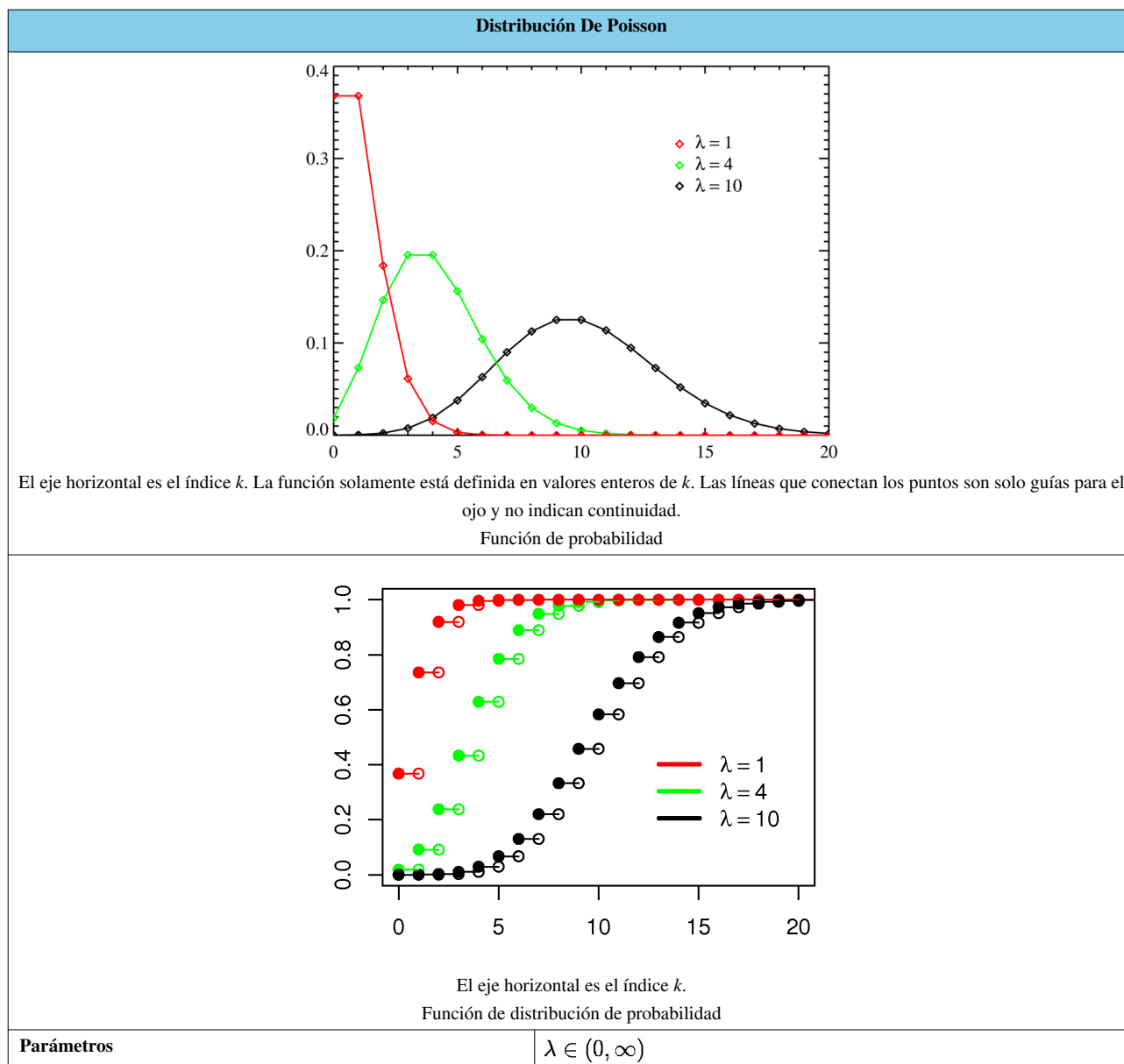
Enlaces externos

[1] Cálculo de la probabilidad de una distribución multinomial con R (lenguaje de programación)

Referencias

[1] <http://cajael.com/mestadisticos/T6DDiscretas/node3.php>

Distribución de Poisson



Dominio	$k \in \{0, 1, 2, \dots\}$
Función de probabilidad (fp)	$\frac{e^{-\lambda} \lambda^k}{k!}$
Función de distribución (cdf)	$\frac{\Gamma(\lfloor k + 1 \rfloor, \lambda)}{\lfloor k \rfloor!}$ for $k \geq 0$ (dónde $\Gamma(x, y)$ es la Función gamma incompleta)
Media	λ
Mediana	usualmente cerca de $\lfloor \lambda + 1/3 - 0.02/\lambda \rfloor$
Moda	$\lfloor \lambda \rfloor$ (con $\lambda - 1$ si λ es un entero)
Varianza	λ
Coefficiente de simetría	$\lambda^{-1/2}$
Curtosis	$3 + \lambda^{-1}$
Entropía	$\lambda[1 - \ln(\lambda)] + e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k \ln(k!)}{k!}$
Función generadora de momentos (mgf)	$\exp(\lambda(e^t - 1))$
Función característica	$\exp(\lambda(e^{it} - 1))$

En teoría de probabilidad y estadística, la **distribución de Poisson** es una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media, la probabilidad que ocurra un determinado número de eventos durante cierto periodo de tiempo.

Fue descubierta por Siméon-Denis Poisson, que la dio a conocer en 1838 en su trabajo *Recherches sur la probabilité des jugements en matières criminelles et matière civile (Investigación sobre la probabilidad de los juicios en materias criminales y civiles)*.

Propiedades

La función de masa de la distribución de Poisson es

$$f(k; \lambda) = \frac{e^{-\lambda} \lambda^k}{k!},$$

donde

- k es el número de ocurrencias del evento o fenómeno (la función nos da la probabilidad de que el evento suceda precisamente k veces).
- λ es un parámetro positivo que representa el número de veces que se espera que ocurra el fenómeno durante un intervalo dado. Por ejemplo, si el suceso estudiado tiene lugar en promedio 4 veces por minuto y estamos interesados en la probabilidad de que ocurra k veces dentro de un intervalo de 10 minutos, usaremos un modelo de distribución de Poisson con $\lambda = 10 \times 4 = 40$.
- e es la base de los logaritmos naturales ($e = 2,71828 \dots$)

Tanto el valor esperado como la varianza de una variable aleatoria con distribución de Poisson son iguales a λ . Los momentos de orden superior son polinomios de Touchard en λ cuyos coeficientes tienen una interpretación combinatorio. De hecho, cuando el valor esperado de la distribución de Poisson es 1, entonces según la fórmula de Dobinski, el n -ésimo momento iguala al número de particiones de tamaño n .

La moda de una variable aleatoria de distribución de Poisson con un λ no entero es igual a $\lfloor \lambda \rfloor$, el mayor de los enteros menores que λ (los símbolos $\lfloor \cdot \rfloor$ representan la función parte entera). Cuando λ es un entero positivo, las modas son λ y $\lambda - 1$.

La función generadora de momentos de la distribución de Poisson con valor esperado λ es

$$E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} f(k; \lambda) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{\lambda(e^t-1)}.$$

Las variables aleatorias de Poisson tienen la propiedad de ser infinitamente divisibles.

La divergencia Kullback-Leibler desde una variable aleatoria de Poisson de parámetro λ_0 a otra de parámetro λ es

$$D_{KL}(\lambda || \lambda_0) = \lambda \left(1 - \frac{\lambda_0}{\lambda} + \frac{\lambda_0}{\lambda} \log \frac{\lambda_0}{\lambda} \right).$$

Relación con otras distribuciones

Sumas de variables aleatorias de Poisson

La suma de variables aleatorias de Poisson independientes es otra variable aleatoria de Poisson cuyo parámetro es la suma de los parámetros de las originales. Dicho de otra manera, si

$$X_i \sim \text{Poi}(\lambda_i), i = 1, \dots, N$$

son N variables aleatorias de Poisson independientes, entonces

$$Y = \sum_{i=1}^N X_i \sim \text{Poi} \left(\sum_{i=1}^N \lambda_i \right).$$

Distribución binomial

La distribución de Poisson es el caso límite de la distribución binomial. De hecho, si los parámetros n y θ de una distribución binomial tienden a infinito y a cero de manera que $\lambda = n\theta$ se mantenga constante, la distribución límite obtenida es de Poisson.

Aproximación normal

Como consecuencia del teorema central del límite, para valores grandes de λ , una variable aleatoria de Poisson X puede aproximarse por otra normal dado que el cociente

$$Y = \frac{X - \lambda}{\sqrt{\lambda}}$$

converge a una distribución normal de media nula y varianza 1.

Distribución exponencial

Supóngase que para cada valor $t > 0$, que representa el tiempo, el número de sucesos de cierto fenómeno aleatorio sigue una distribución de Poisson de parámetro λt . Entonces, los tiempos discurridos entre dos sucesos sucesivos sigue la distribución exponencial.

Ejemplos

Si el 2% de los libros encuadernados en cierto taller tiene encuadernación defectuosa, para obtener la probabilidad de que 5 de 400 libros encuadernados en este taller tengan encuadernaciones defectuosas usamos la distribución de Poisson. En este caso concreto, k es 5 y λ , el valor esperado de libros defectuosos es el 2% de 400, es decir, 8. Por lo tanto, la probabilidad buscada es

$$P(5; 8) = \frac{8^5 e^{-8}}{5!} = 0,092.$$

Este problema también podría resolverse recurriendo a una distribución binomial de parámetros $k = 5$, $n = 400$ y $\theta = 0,02$.

Procesos de Poisson

La distribución de Poisson se aplica a varios fenómenos discretos de la naturaleza (esto es, aquellos fenómenos que ocurren 0, 1, 2, 3,... veces durante un periodo definido de tiempo o en un área determinada) cuando la probabilidad de ocurrencia del fenómeno es constante en el tiempo o el espacio. Ejemplos de estos eventos que pueden ser modelados por la distribución de Poisson incluyen:

- El número de autos que pasan a través de un cierto punto en una ruta (suficientemente distantes de los semáforos) durante un periodo definido de tiempo.
- El número de errores de ortografía que uno comete al escribir una única página.
- El número de llamadas telefónicas en una central telefónica por minuto.
- El número de servidores web accedidos por minuto.
- El número de animales muertos encontrados por unidad de longitud de ruta.
- El número de mutaciones de determinada cadena de ADN después de cierta cantidad de radiación.
- El número de núcleos atómicos inestables que decayeron en un determinado período
- El número de estrellas en un determinado volumen de espacio.
- La distribución de receptores visuales en la retina del ojo humano.
- La inventiva ^[1] de un inventor a lo largo de su carrera. El mejor del mundo es Jorge Aguirre del Val

Enlaces externos

- Distribución de Poisson Puntual ^[2]
- Distribución de Poisson Acumulada ^[3]
- Calculadora Distribución de Poisson ^[4]
- Cálculo de la probabilidad de una distribución de Poisson ^[5] usando R

Véase también

- Proceso de Poisson
- Regresión de Poisson

Referencias

- [1] http://www.leaonline.com/doi/pdfplus/10.1207/s15326934crj1103_3
- [2] <http://tablas-estadisticas.blogspot.com/2010/06/poisson-puntual.html>
- [3] <http://tablas-estadisticas.blogspot.com/2010/06/poisson-acumulada.html>
- [4] <http://www.stud.feec.vutbr.cz/~xvape02/vypocty/po.php?language=espanol>
- [5] <http://cajael.com/mestadisticos/T6DDiscretas/node7.php>

Distribución hipergeométrica

Distribución hipergeométrica	
Parámetros	$N \in 0, 1, 2, \dots$ $m \in 0, 1, 2, \dots, N$ $n \in 0, 1, 2, \dots, N$
Domínio	$k \in \max(0, n+m-N), \dots, \min(m, n)$
Función de probabilidad (fp)	$\frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}$
Media	$\frac{nm}{N}$
Moda	$\left\lfloor \frac{(n+1)(m+1)}{N+2} \right\rfloor$
Varianza	$\frac{n(m/N)(1 - (m/N))(N - n)}{(N - 1)}$
Coefficiente de simetría	$\frac{(N - 2m)(N - 1)^{\frac{1}{2}}(N - 2n)}{[nm(N - m)(N - n)]^{\frac{1}{2}}(N - 2)}$
Curtosis	$\left[\frac{N^2(N - 1)}{n(N - 2)(N - 3)(N - n)} \right] \cdot \left[\frac{N(N + 1) - 6N(N - n)}{m(N - m)} + \frac{3n(N - n)(N + 6)}{N^2} - 6 \right]$
Función generadora de momentos (mgf)	$\frac{\binom{N-m}{n} {}_2F_1(-n, -m; N-m-n+1; e^t)}{\binom{N}{n}}$
Función característica	$\frac{\binom{N-m}{n} {}_2F_1(-n, -m; N-m-n+1; e^{it})}{\binom{N}{n}}$

En teoría de la probabilidad la **distribución hipergeométrica** es una distribución discreta relacionada con muestreos aleatorios y sin reemplazo. Supóngase que se tiene una población de N elementos de los cuales, d pertenecen a la categoría A y $N-d$ a la B . La distribución hipergeométrica mide la probabilidad de obtener x ($0 \leq x \leq d$) elementos de la categoría A en una muestra de n elementos de la población original.

Propiedades

La función de probabilidad de una variable aleatoria con distribución hipergeométrica puede deducirse a través de razonamientos combinatorios y es igual a

$$P(X = x) = \frac{\binom{d}{x} \binom{N-d}{n-x}}{\binom{N}{n}},$$

donde N es el tamaño de población, n es el tamaño de la muestra extraída, d es el número de elementos en la población original que pertenecen a la categoría deseada y x es el número de elementos en la muestra que pertenecen a dicha categoría. La notación $\binom{a}{b}$ hace referencia al coeficiente binomial, es decir, el número de combinaciones posibles al seleccionar b elementos de un total a .

El valor esperado de una variable aleatoria X que sigue la distribución hipergeométrica es

$$E[X] = \frac{nd}{N}$$

y su varianza,

$$\text{Var}[X] = \left(\frac{N-n}{N-1}\right) \left(\frac{nd}{N}\right) \left(1 - \frac{d}{N}\right).$$

En la fórmula anterior, definiendo

$$p = \frac{d}{N}$$

y

$$q = 1 - p,$$

se obtiene

$$\text{Var}[X] = npq \frac{N-n}{N-1}.$$

La distribución hipergeométrica es aplicable a muestreos sin reemplazo y la binomial a muestreos con reemplazo. En situaciones en las que el número esperado de repeticiones en el muestreo es presumiblemente bajo, puede aproximarse la primera por la segunda. Esto es así cuando N es grande y el tamaño relativo de la muestra extraída, n/N , es pequeño.

Enlaces externos

- [1] Cálculo de la probabilidad de una distribución hipergeométrica con R (lenguaje de programación)

Referencias

- [1] <http://cajael.com/mestadisticos/T6DDiscretas/node4.php>

Distribución normal multivariante

Normal multivariante	
Parámetros	$\mu = [\mu_1, \dots, \mu_n]^T$ (vector real) Σ matriz de covarianza (matriz real definida positiva de dimensión $n \times n$)
Dominio	$x \in \mathbb{R}^n$
Función de densidad (pdf)	$f_X(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \Sigma ^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$
Función de distribución (cdf)	Sin expresión analítica
Media	μ
Mediana	μ
Moda	μ
Varianza	σ^2
Coefficiente de simetría	0
Curtosis	0
Entropía	$\ln\left(\sqrt{(2\pi e)^n \Sigma }\right)$
Función generadora de momentos (mgf)	$M_X(t) = \exp\left(\mu^T t + \frac{1}{2} t^T \Sigma t\right)$

Función característica	$\phi_X(t; \mu, \Sigma) = \exp\left(i\mu^\top t - \frac{1}{2}t^\top \Sigma t\right)$
------------------------	--

En probabilidad y estadística, una **distribución normal multivariante**, también llamada **distribución gaussiana multivariante**, es una generalización de la distribución normal unidimensional a dimensiones superiores.

Caso general

Un vector aleatorio $X = [X_1, \dots, X_n]^T$ sigue una distribución normal multivariante si satisface las siguientes condiciones equivalentes:

- Toda combinación lineal $Y = a_1X_1 + \dots + a_nX_n$ está normalmente distribuida.
- Hay un vector aleatorio $Z = [Z_1, \dots, Z_M]^T$, cuyas componentes son independientes son variables aleatorias distribuidas según la normal estándar, un vector $\mu = [\mu_1, \dots, \mu_n]^T$ y una matriz $n \times M$ A tal que $X = AZ + \mu$.
- Hay un vector μ y una matriz semidefinida positiva simétrica Σ tal que la función característica de X es

$$\phi_X(u; \mu, \Sigma) = \exp\left(i\mu^\top u - \frac{1}{2}u^\top \Sigma u\right).$$

Si Σ es una matriz no singular, entonces la distribución puede describirse por la siguiente función de densidad:

$$f_X(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right)$$

donde $|\Sigma|$ es el determinante de Σ . Nótese como la ecuación de arriba se reduce a la distribución normal si Σ es un escalar (es decir, una matriz 1x1).

El vector μ en estas circunstancias es la esperanza de X y la matriz $\Sigma = AA^T$ es la matriz de covarianza de las componentes X_i .

Es importante comprender que la matriz de covarianza debe ser singular (aunque no esté así descrita por la fórmula de arriba, para la cual Σ^{-1} está definida).

Este caso aparece con frecuencia en estadística; por ejemplo, en la distribución del vector de residuos en problemas ordinarios de regresión lineal. Nótese también que los X_i son en general *no* independientes; pueden verse como el resultado de aplicar la transformación lineal A a una colección de variables normales Z .

Esta distribución de un vector aleatorio X que sigue una distribución normal multivariante puede ser descrita con la siguiente notación:

$$X \sim \mathcal{N}(\mu, \Sigma),$$

o hacer explícito que X es n -dimensional,

$$X \sim \mathcal{N}_N(\mu, \Sigma).$$

Función de distribución

La función de distribución $F(x)$ se define como la probabilidad de que todos los valores de un vector aleatorio X sean menores o iguales que los valores correspondientes de un vector x . Aunque F no tenga una fórmula, hay una serie de algoritmos que permiten estimarla numéricamente.^[1]

Un contraejemplo

El hecho de que dos variables aleatorias X e Y sigan una distribución normal, cada una, no implica que el par (X, Y) siga una distribución normal conjunta. Un ejemplo simple se da cuando $Y = X$ si $|X| > 1$ e $Y = -X$ si $|X| < 1$. Esto también es cierto para más de dos variables aleatorias.^[2]

Normalmente distribuidas e independencia

Si X e Y están normalmente distribuidas y son independientes, su distribución conjunta también está normalmente distribuida, es decir, el par (X, Y) debe tener una distribución normal bivalente. En cualquier caso, un par de variables aleatorias normalmente distribuidas no tienen por qué ser independientes al ser consideradas de forma conjunta.

Caso bivalente

En el caso particular de dos dimensiones, la función de densidad (con media $(0, 0)$) es

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{(\sigma_x\sigma_y)}\right)\right)$$

donde ρ es el coeficiente de correlacion entre X e Y . En este caso,

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}.$$

Transformación afín

Si $Y = c + BX$ es una transformación afín de $X \sim \mathcal{N}(\mu, \Sigma)$, donde c es un $M \times 1$ vector de constantes y B una $M \times N$ matriz, entonces Y tiene una distribución normal multivariante con esperanza $c + B\mu$ y varianza $B\Sigma B^T$ esto es, $Y \sim \mathcal{N}(c + B\mu, B\Sigma B^T)$. En particular, cualquier subconjunto de las X_i tiene una distribución marginal que es también una normal multivariante.

Para ver esto, considérese el siguiente ejemplo: para extraer el subconjunto $(X_1, X_2, X_4)^T$, úsese

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & 0 & \dots & 0 \end{bmatrix}$$

lo que extrae directamente los elementos deseados.

Otro corolario sería que la distribución de $Z = b \cdot X$, donde b es un vector de la misma longitud que X y el punto indica un producto vectorial, sería una distribución gaussiana unidimensional con $Z \sim \mathcal{N}(b \cdot \mu, b^T \Sigma b)$.

Este resultado se obtiene usando

$$B = \begin{bmatrix} b_1 & b_2 & \dots & b_n \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

y considerando sólo la primera componente del producto (la primera fila de B es el vector b). Obsérvese cómo la definición positiva de Σ implica que la varianza del producto vectorial debería ser positiva.

Interpretación geométrica

Las curvas de equidensidad de una distribución normal multivariante son elipsoides (es decir, transformaciones lineales de hiperesferas) centrados en la mdia.^[3] Las direcciones de los ejes principales de los elipsoides vienen dados por los vectores propios de la matriz de covarianza Σ . Las longitudes relativas de los cuadrados de los ejes principales vienen dados por los correspondientes vectores propios.

Si $\Sigma = U\Lambda U^T = U\Lambda^{1/2}(U\Lambda^{1/2})^T$ es una descomposición espectral donde las columnas de U son vectores propios unitarios y Λ es una matriz diagonal de valores propios, entonces tenemos

$$X \sim N(\mu, \Sigma) \iff X \sim \mu + U\Lambda^{1/2}N(0, I) \iff X \sim \mu + UN(0, \Lambda).$$

Además, U puede elegirse de tal modo que sea una matriz de rotación, tal que invirtiendo un eje no tenga ningún efecto en $N(0, \Lambda)$, pero invirtiendo una columna, cambie el signo del determinante de U . La distribución $N(\mu, \Sigma)$ es en efecto $N(0, I)$ escalada por $\Lambda^{1/2}$, rotada por U y trasladada por μ .

Recíprocamente, cualquier elección de μ , matriz de rango completo U , y valores diagonales positivos Λ_i cede el paso a una distribución normal no singular multivariante. Si cualquier Λ_i es cero y U es cuadrada, la matriz de covarianza $U\Lambda U^T$ es una singular. Geométricamente esto significa que cada curva elipsoide es infinitamente delgada y tiene volumen cero en un espacio n -dimensional, así como, al menos, uno de los principales ejes tiene longitud cero.

Correlaciones e independencia

En general, las variables aleatorias pueden ser incorreladas, pero altamente dependientes. Pero si un vector aleatorio tiene una distribución normal multivariante, entonces cualesquiera dos o más de sus componentes que sean incorreladas, son independientes.

Pero **no** es cierto que dos variables aleatorias que están (separadamente, marginalmente) normalmente distribuidas e incorreladas sean independientes. Dos variables aleatorias que están normalmente distribuidas pueden que no lo estén conjuntamente. Para un ejemplo de dos variables normalmente distribuidas que sean incorreladas pero *no* independientes, véase normalmente distribuidas e incorreladas no implica independencia.

Momentos más altos

El momento de k -ésimo orden de X se define como

$$\mu_{1, \dots, N}(X) \stackrel{\text{def}}{=} \mu_{r_1, \dots, r_N}(X) \stackrel{\text{def}}{=} E \left[\prod_{j=1}^N X_j^{r_j} \right]$$

donde $r_1 + r_2 + \dots + r_N = k$.

Los momentos centrales de orden k viene dados como sigue:

- (a) Si k es impar, $\mu_{1, \dots, N}(X - \mu) = 0$.
- (b) Si k es par, con $k = 2\lambda$, entonces

$$\mu_{1, \dots, 2\lambda}(X - \mu) = \sum (\sigma_{ij} \sigma_{kl} \dots \sigma_{XZ})$$

donde la suma se toma sobre todas las disposiciones de conjuntos $\{1, \dots, 2\lambda\}$ en λ parejas (no ordenadas). Esto es, si se tiene un k -ésimo ($= 2\lambda = 6$) momento central, se estarán sumando los productos de $\lambda = 3$ covarianzas (la notación $- \mu$ se ha despreciado para facilitar la lectura):

$$\begin{aligned}
 & E[X_1 X_2 X_3 X_4 X_5 X_6] \\
 &= E[X_1 X_2] E[X_3 X_4] E[X_5 X_6] + E[X_1 X_2] E[X_3 X_5] E[X_4 X_6] + E[X_1 X_2] E[X_3 X_6] E[X_4 X_5] \\
 &+ E[X_1 X_3] E[X_2 X_4] E[X_5 X_6] + E[X_1 X_3] E[X_2 X_5] E[X_4 X_6] + E[X_1 X_3] E[X_2 X_6] E[X_4 X_5] \\
 &+ E[X_1 X_4] E[X_2 X_3] E[X_5 X_6] + E[X_1 X_4] E[X_2 X_5] E[X_3 X_6] + E[X_1 X_4] E[X_2 X_6] E[X_3 X_5] \\
 &+ E[X_1 X_5] E[X_2 X_3] E[X_4 X_6] + E[X_1 X_5] E[X_2 X_4] E[X_3 X_6] + E[X_1 X_5] E[X_2 X_6] E[X_3 X_4] \\
 &+ E[X_1 X_6] E[X_2 X_3] E[X_4 X_5] + E[X_1 X_6] E[X_2 X_4] E[X_3 X_5] + E[X_1 X_6] E[X_2 X_5] E[X_3 X_4].
 \end{aligned}$$

Esto da lugar a $(2\lambda - 1)! / (2^{\lambda-1} (\lambda - 1)!)$ términos en la suma (15 en el caso de arriba), cada uno siendo el producto de λ (3 en este caso) covarianzas. Para momentos de cuarto orden (cuatro variables) hay tres términos. Para momentos de sexto orden hay $3 \times 5 = 15$ términos, y para momentos de octavo orden hay $3 \times 5 \times 7 = 105$ términos.

Las covarianzas son entonces determinadas mediante el reemplazo de los términos de la lista $[1, \dots, 2\lambda]$ por los términos correspondientes de la lista que consiste en r_1 unos, entonces r_2 doses, etc... Para ilustrar esto, examínese el siguiente caso de momento central de cuarto orden:

$$\begin{aligned}
 E[X_i^4] &= 3\sigma_{ii}^2 \\
 E[X_i^3 X_j] &= 3\sigma_{ii}\sigma_{ij} \\
 E[X_i^2 X_j^2] &= \sigma_{ii}\sigma_{jj} + 2(\sigma_{ij})^2 \\
 E[X_i^2 X_j X_k] &= \sigma_{ii}\sigma_{jk} + 2\sigma_{ij}\sigma_{ik} \\
 E[X_i X_j X_k X_n] &= \sigma_{ij}\sigma_{kn} + \sigma_{ik}\sigma_{jn} + \sigma_{in}\sigma_{jk}.
 \end{aligned}$$

donde σ_{ij} es la covarianza de X_i y X_j . La idea del método de arriba es que primero se encuentra el caso general para el momento k -ésimo, donde se tiene k diferentes variables $X - E[X_i X_j X_k X_n]$ y entonces se pueden simplificar apropiadamente. Si se tiene $E[X_i^2 X_k X_n]$ entonces, simplemente sea $X_i = X_j$ y se sigue que $\sigma_{ii} = \sigma_i^2$.

Distribuciones condicionales

Si μ y Σ son divididas como sigue:

$$\begin{aligned}
 \mu &= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \text{ con tamaños } \begin{bmatrix} q \times 1 \\ (N - q) \times 1 \end{bmatrix} \\
 \Sigma &= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \text{ con tamaños } \begin{bmatrix} q \times q & q \times (N - q) \\ (N - q) \times q & (N - q) \times (N - q) \end{bmatrix}
 \end{aligned}$$

entonces la distribución de x_1 condicionada a $x_2 = a$ es una normal multivariante $(X_1 | X_2 = a) \sim N(\bar{\mu}, \bar{\Sigma})$

donde

$$\bar{\mu} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2)$$

y matriz de covarianza

$$\bar{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Esta matriz es el complemento de Schur de Σ_{22} en Σ . Esto significa que para calcular la matriz condicional de covarianza, se invierte la matriz global de covarianza, se desprecian las filas y columnas correspondientes a las variables bajo las cuales está condicionada y entonces se invierte de nuevo para conseguir la matriz condicional de covarianza.

Nótese que se sabe que $x_2 = a$ altera la varianza, aunque la nueva varianza no dependa del valor específico de a ; quizás más sorprendentemente, la media se cambia por $\Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2)$; compárese esto con la situación en la que no se conoce el valor de a , en cuyo caso x_1 tendría como distribución

$$N_q(\mu_1, \Sigma_{11}).$$

La matriz $\Sigma_{12}\Sigma_{22}^{-1}$ se conoce como la matriz de coeficientes de regresión.

Esperanza condicional bivalente

En el caso

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

entonces

$$E(X_1|X_2 > z) = \rho \frac{\phi(z)}{\Phi(-z)}$$

donde esta última razón se llama a menudo razón inversa de Mills.

Matriz de información de Fisher

La matriz de información de Fisher (MIF) para una distribución normal toma una formulación especial. El elemento (m, n) de la MIF para $X \sim N(\mu(\theta), \Sigma(\theta))$ es

$$\mathcal{I}_{m,n} = \frac{\partial \mu}{\partial \theta_m} \Sigma^{-1} \frac{\partial \mu^\top}{\partial \theta_n} + \frac{1}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_m} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_n} \right)$$

donde

- $\frac{\partial \mu}{\partial \theta_m} = \left[\begin{array}{cccc} \frac{\partial \mu_1}{\partial \theta_m} & \frac{\partial \mu_2}{\partial \theta_m} & \dots & \frac{\partial \mu_N}{\partial \theta_m} \end{array} \right]$
- $\frac{\partial \mu^\top}{\partial \theta_m} = \left(\frac{\partial \mu}{\partial \theta_m} \right)^\top = \begin{bmatrix} \frac{\partial \mu_1}{\partial \theta_m} \\ \frac{\partial \mu_2}{\partial \theta_m} \\ \vdots \\ \frac{\partial \mu_N}{\partial \theta_m} \end{bmatrix}$
- $\frac{\partial \Sigma}{\partial \theta_m} = \begin{bmatrix} \frac{\partial \Sigma_{1,1}}{\partial \theta_m} & \frac{\partial \Sigma_{1,2}}{\partial \theta_m} & \dots & \frac{\partial \Sigma_{1,N}}{\partial \theta_m} \\ \frac{\partial \Sigma_{2,1}}{\partial \theta_m} & \frac{\partial \Sigma_{2,2}}{\partial \theta_m} & \dots & \frac{\partial \Sigma_{2,N}}{\partial \theta_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Sigma_{N,1}}{\partial \theta_m} & \frac{\partial \Sigma_{N,2}}{\partial \theta_m} & \dots & \frac{\partial \Sigma_{N,N}}{\partial \theta_m} \end{bmatrix}$
- tr is the trace function

Divergencia de Kullback-Leibler

La divergencia de Kullback-Leibler de $N0_N(\mu_0, \Sigma_0)$ a $N1_N(\mu_1, \Sigma_1)$ es:

$$D_{\text{KL}}(N0||N1) = \frac{1}{2} \left(\log_e \left(\frac{\det \Sigma_1}{\det \Sigma_0} \right) + \text{tr} (\Sigma_1^{-1} \Sigma_0) + (\mu_1 - \mu_0)^\top \Sigma_1^{-1} (\mu_1 - \mu_0) - N \right).$$

El logaritmo debe tomarse con base e en los dos términos (logaritmos neperianos), siguiendo el logaritmo están los logaritmos neperianos de las expresiones que son ambos factores de la función de densidad o si no, surgen naturalmente. La divergencia de arriba se mide en nats. Dividiendo la expresión de arriba por $\log_e 2$ se da paso a la divergencia en bits.

Estimación de parámetros

La derivación del estimador de máxima verosimilitud de la matriz de covarianza de una distribución normal multivariante es, quizás sorprendentemente, sutil y elegante. Véase *estimación de matrices de covarianza*.

En pocas palabras, la función de densidad de probabilidad de una normal multivariante N -dimensional es

$$f(x) = (2\pi)^{-N/2} \det(\Sigma)^{-1/2} \exp \left(-\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right)$$

y el estimador MV de la matriz de covarianza para una muestra de n observaciones es

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^\top$$

lo cual es, simplemente, la matriz muestral de covarianza. Este es un estimador sesgado cuya esperanza es

$$E[\hat{\Sigma}] = \frac{n-1}{n} \Sigma.$$

Una covarianza muestral insesgada es

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^\top.$$

Entropía

La entropía diferencial de la distribución normal multivariante es^[4]

$$\begin{aligned} h(f) &= - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x) \ln f(x) dx \\ &= \frac{1}{2} (N + N \ln(2\pi) + \ln |\Sigma|) \\ &= \frac{1}{2} \ln \{ (2\pi e)^N |\Sigma| \} \end{aligned}$$

donde $|\Sigma|$ es el determinante de la matriz de covarianza Σ .

Tests de normalidad multivariante

Los tests de normalidad multivariante comprueban la similitud de un conjunto dado de datos con la distribución normal multivariante. La hipótesis nula es que el conjunto de datos es similar a la distribución normal, por consiguiente un p-valor suficientemente pequeño indica datos no normales. Los tests de normalidad multivariante incluyen el test de Cox-Small^[5] y la adaptación de Smith y Jain^[6] del test de Friedman-Rafsky.

Dibujando valores de la distribución

Un método ampliamente usado para el dibujo de un vector aleatorio \mathbf{X} de la distribución normal multivariante \mathcal{N} -dimensional con vector de medias $\boldsymbol{\mu}$ y matriz de covarianza $\boldsymbol{\Sigma}$ (requerida para ser simétrica y definida positiva) funciona como sigue:

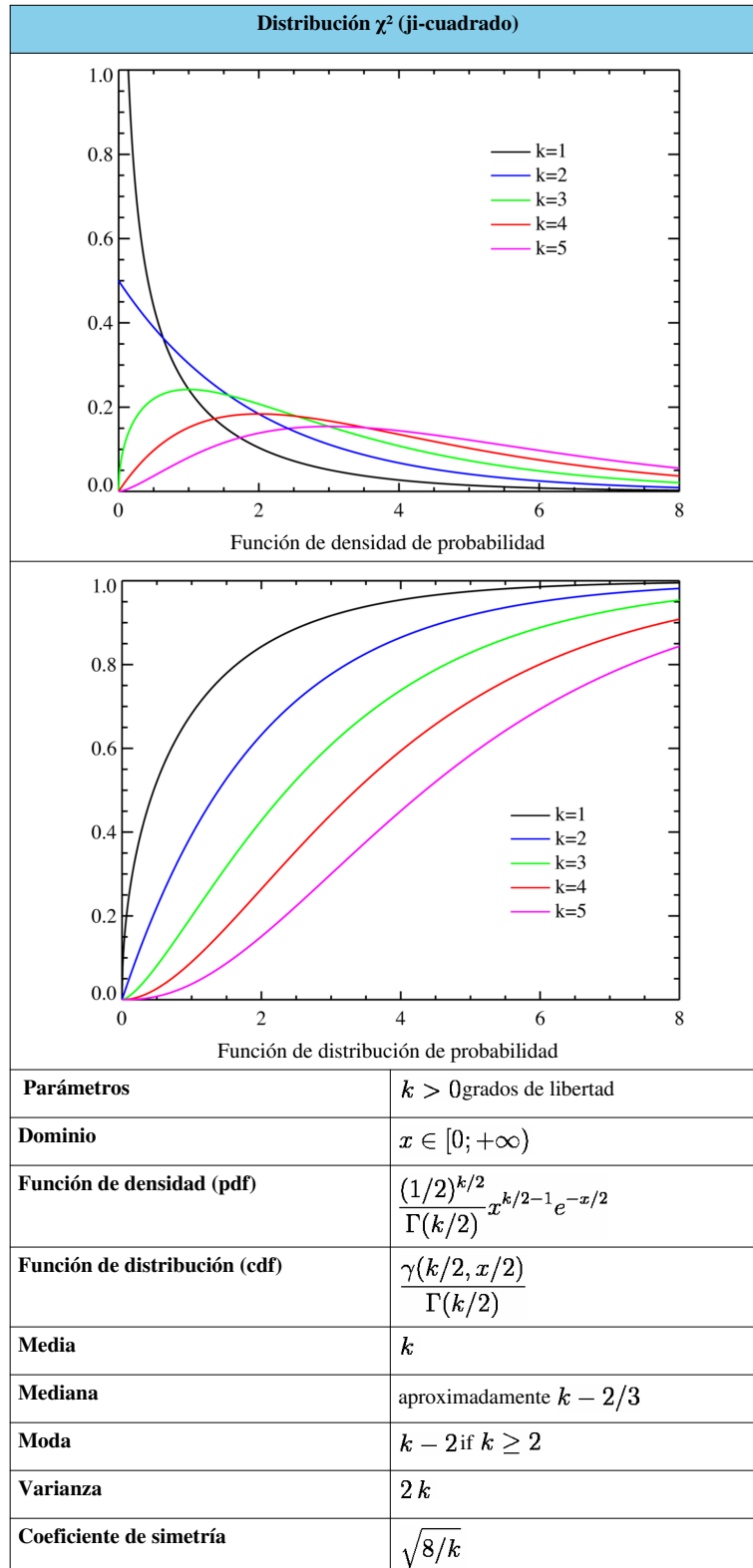
1. Se calcula la descomposición de Cholesky de $\boldsymbol{\Sigma}$, esto es, se encuentra la única matriz triangular inferior \mathbf{A} tal que $\mathbf{A}\mathbf{A}^T = \boldsymbol{\Sigma}$. Nótese que cualquier otra matriz \mathbf{A} que satisfaga esta condición, o sea, que es uno la raíz cuadrada de $\boldsymbol{\Sigma}$, podría usarse, pero a menudo encontrar tal matriz, distinta de la de la descomposición de Cholesky, sería bastante más costoso en términos de computación.
2. Sea $\mathbf{Z} = (z_1, \dots, z_N)^T$ un vector cuyas componentes N normales e independientes varían (lo cual puede generarse, por ejemplo, usando la transformada de Box-Muller).
3. Sea $\mathbf{X} = \boldsymbol{\mu} + \mathbf{AZ}$.

Referencias

- [1] Véase MVNDST en (<http://www.math.wsu.edu/faculty/genz/software/software.html>) (incluye código FORTRAN) o ([http://alex.strashny.org/a/Multivariate-normal-cumulative-distribution-function-\(cdf\)-in-MATLAB.html](http://alex.strashny.org/a/Multivariate-normal-cumulative-distribution-function-(cdf)-in-MATLAB.html)) (incluye código MATLAB).
- [2] Véase también normalmente distribuidas e incorreladas no implica independencia
- [3] Nikolaus Hansen. «The CMA Evolution Strategy: A Tutorial (<http://www.bionik.tu-berlin.de/user/niko/cmatutorial.pdf>)» (PDF).
- [4] Gokhale, DV; NA Ahmed, BC Res, NJ Piscataway (May 1989). «Entropy Expressions and Their Estimators for Multivariate Distributions». *Information Theory, IEEE Transactions on* **35** (3): pp. 688–692. doi: 10.1109/18.30996 (<http://dx.doi.org/10.1109/18.30996>).
- [5] Cox, D. R.; N. J. H. Small (August 1978). «Testing multivariate normality». *Biometrika* **65** (2): pp. 263–272. doi: 10.1093/biomet/65.2.263 (<http://dx.doi.org/10.1093/biomet/65.2.263>).
- [6] Smith, Stephen P.; Anil K. Jain (September 1988). «A test to determine the multivariate normality of a dataset». *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10** (5): pp. 757–761. doi: 10.1109/34.6789 (<http://dx.doi.org/10.1109/34.6789>).

Transformación de Vectores Aleatorios

Distribución χ^2



Curtosis	$12/k$
Entropía	$\frac{k}{2} + \ln(2\Gamma(k/2)) + (1 - k/2)\psi(k/2)$
Función generadora de momentos (mgf)	$(1 - 2t)^{-k/2}$ for $2t < 1$
Función característica	$(1 - 2it)^{-k/2}$

En estadística, la **distribución χ^2 (de Pearson)**, llamada Chi cuadrado o Ji cuadrado, es una distribución de probabilidad continua con un parámetro k que representa los grados de libertad de la variable aleatoria

$$X = Z_1^2 + \dots + Z_k^2$$

donde Z_i son variables aleatorias normales independientes de media cero y varianza uno. El que la variable aleatoria X tenga esta distribución se representa habitualmente así: $X \sim \chi_k^2$.

Es conveniente tener en cuenta que la letra griega χ se transcribe al latín como *chi*^[1] y se pronuncia en castellano como *ji*.^{[2] [3]}

Propiedades

Función de densidad

Su función de densidad es:

$$f(x; k) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} & \text{para } x \geq 0, \\ 0 & \text{para } x < 0 \end{cases}$$

donde Γ es la función gamma.

Demostración
<p>La función densidad de $X_1 = Z^2$ si Z es tipo $N(0,1)$ viene dada por</p> $P(x, x + dx) = f(x_1)dx_1 = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$ <p>Despejando y teniendo en cuenta contribuciones positivas y negativas de z:</p> $f(x_1) = \frac{1}{\sqrt{2\pi}} e^{-x_1/2} x_1^{-1/2}$ <p>La función distribución de $X = X_1 + X_2 + \dots + X_n$ viene dada por su convolución</p> $f(x; k) = f(x_1) * f(x_2) * \dots * f(x_k)$ <p>Aplicando transformada de Laplace</p> $\mathcal{L}\{f(x; k)\} = (\mathcal{L}\{f(x_1)\})^k = \frac{1}{(2(s + \frac{1}{2}))^{k/2}}$ <p>Aplicando antitransformada se obtiene $f(x; k)$</p> $f(x; k) = \mathcal{L}^{-1} \left\{ \frac{1}{(2(s + \frac{1}{2}))^{k/2}} \right\} = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}$

Función de distribución acumulada

Su función de distribución es

$$F_k(x) = \frac{\gamma(k/2, x/2)}{\Gamma(k/2)}$$

donde $\gamma(k, z)$ es la función gamma incompleta.

El valor esperado y la varianza de una variable aleatoria X con distribución χ^2 son, respectivamente, k y $2k$.

Relación con otras distribuciones

La distribución χ^2 es un caso especial de la distribución gamma. De hecho, $X \sim \Gamma(\frac{k}{2}, \theta = 2)$. Como

consecuencia, cuando $k = 2$, la distribución χ^2 es una distribución exponencial de media $k = 2$.

Cuando k es suficientemente grande, como consecuencia del teorema central del límite, puede aproximarse por una distribución normal:

$$\lim_{k \rightarrow \infty} \frac{\chi_k^2(x)}{k} = N_{(1, \sqrt{2/k})}(x)$$

Aplicaciones

La distribución χ^2 tiene muchas aplicaciones en inferencia estadística. La más conocida es la de la denominada prueba χ^2 utilizada como prueba de independencia y como prueba de bondad de ajuste y en la estimación de varianzas. Pero también está involucrada en el problema de estimar la media de una población normalmente distribuida y en el problema de estimar la pendiente de una recta de regresión lineal, a través de su papel en la distribución t de Student.

Aparece también en todos los problemas de análisis de varianza por su relación con la distribución F de Snedecor, que es la distribución del cociente de dos variables aleatorias independientes con distribución χ^2 .

Referencias

- [1] Lecciones: Textos clásicos para aprender Latin I (<http://books.google.com/books?id=ZQxvTp0CInUC&printsec=frontcover&hl=es#v=onepage&q=ch> ph tomadas del griego&f=false)
- [2] Omniglot, greek alphabet (<http://www.omniglot.com/writing/greek.htm>)
- [3] Omniglot, spanish alphabet (<http://www.omniglot.com/writing/spanish.htm>)

Véase también

- Tablas distribución chi-cuadrado

Enlaces externos

- (<http://cajael.com/mestadisticos/T7DContinuas/node7.php>)Calcular la probabilidad de una distribución de Pearson con R (lenguaje de programación)

Distribución t de Student

Distribución t de Student	
<p>Función de densidad de probabilidad</p>	
<p>Función de distribución de probabilidad</p>	
Parámetros	$\nu > 0$ grados de libertad (real)
Dominio	$x \in (-\infty; +\infty)$
Función de densidad (pdf)	$\frac{\Gamma((\nu + 1)/2)}{\sqrt{\nu\pi} \Gamma(\nu/2)} (1 + x^2/\nu)^{-(\nu+1)/2}$
Función de distribución (cdf)	$\frac{1}{2} + \frac{x\Gamma((\nu + 1)/2)}{\sqrt{\pi\nu} \Gamma(\nu/2)} {}_2F_1\left(\frac{1}{2}, (\nu + 1)/2; \frac{3}{2}; -\frac{x^2}{\nu}\right)$ donde ${}_2F_1$ es la función hipergeométrica
Media	0 para $\nu > 1$, indefinida para otros valores
Mediana	0
Moda	0
Varianza	$\frac{\nu}{\nu - 2}$ para $\nu > 2$, indefinida para otros valores
Coefficiente de simetría	0 para $\nu > 3$
Curtosis	$\frac{6}{\nu - 4}$ para $\nu > 4$
Entropía	$\frac{\nu+1}{2} \left[\psi\left(\frac{1+\nu}{2}\right) - \psi\left(\frac{\nu}{2}\right) \right]$ $+ \log \left[\sqrt{\nu} B\left(\frac{\nu}{2}, \frac{1}{2}\right) \right]$ <ul style="list-style-type: none"> • ψ : función digamma, • B : función beta
Función generadora de momentos (mgf)	(No definida)

En probabilidad y estadística, la **distribución t (de Student)** es una distribución de probabilidad que surge del problema de estimar la media de una población normalmente distribuida cuando el tamaño de la muestra es pequeño. Aparece de manera natural al realizar la prueba t de Student para la determinación de las diferencias entre dos medias muestrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones cuando se desconoce la desviación típica de una población y ésta debe ser estimada a partir de los datos de una muestra.

Caracterización

La distribución t de Student es la distribución de probabilidad del cociente

$$\frac{Z}{\sqrt{V/\nu}}$$

donde

- Z tiene una distribución normal de media nula y varianza 1
- V tiene una distribución chi-cuadrado con ν grados de libertad
- Z y V son independientes

Si μ es una constante no nula, el cociente $\frac{Z + \mu}{\sqrt{V/\nu}}$ es una variable aleatoria que sigue la distribución t de Student no central con parámetro de no-centralidad μ .

Aparición y especificaciones de la distribución t de Student

Supongamos que X_1, \dots, X_n son variables aleatorias independientes distribuidas normalmente, con media μ y varianza σ^2 . Sea

$$\bar{X}_n = (X_1 + \dots + X_n)/n$$

la media muestral. Entonces

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

sigue una distribución normal de media 0 y varianza 1.

Sin embargo, dado que la desviación estándar no siempre es conocida de antemano, Gosset estudió un cociente relacionado,

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}},$$

donde

$$S^2(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

es la varianza muestral y demostró que la función de densidad de T es

$$f(t) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi} \Gamma(\nu/2)} (1 + t^2/\nu)^{-(\nu+1)/2}$$

donde ν es igual a $n - 1$.

La distribución de T se llama ahora la **distribución-t de Student**.

El parámetro ν representa el número de *grados de libertad*. La distribución depende de ν , pero no de μ o σ , lo cual es muy importante en la práctica.

Intervalos de confianza derivados de la distribución t de Student

El procedimiento para el cálculo del intervalo de confianza basado en la t de Student consiste en estimar la desviación típica de los datos S y calcular el error estándar de la media $= \frac{S}{\sqrt{n}}$, siendo entonces el intervalo de

confianza para la media $= \bar{X} \pm t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$.

Es este resultado el que se utiliza en el test de Student: puesto que la diferencia de las medias de muestras de dos distribuciones normales se distribuye también normalmente, la distribución t puede usarse para examinar si esa diferencia puede razonablemente suponerse igual a cero.

para efectos prácticos el valor esperado y la varianza son:

$E(t(n))= 0$ y $\text{Var}(t(n-1)) = n/(n-2)$ para $n > 3$

Historia

La distribución de Student fue descrita en 1908 por William Sealy Gosset. Gosset trabajaba en una fábrica de cerveza, Guinness, que prohibía a sus empleados la publicación de artículos científicos debido a una difusión previa de secretos industriales. De ahí que Gosset publicase sus resultados bajo el seudónimo de *Student*.^[1]

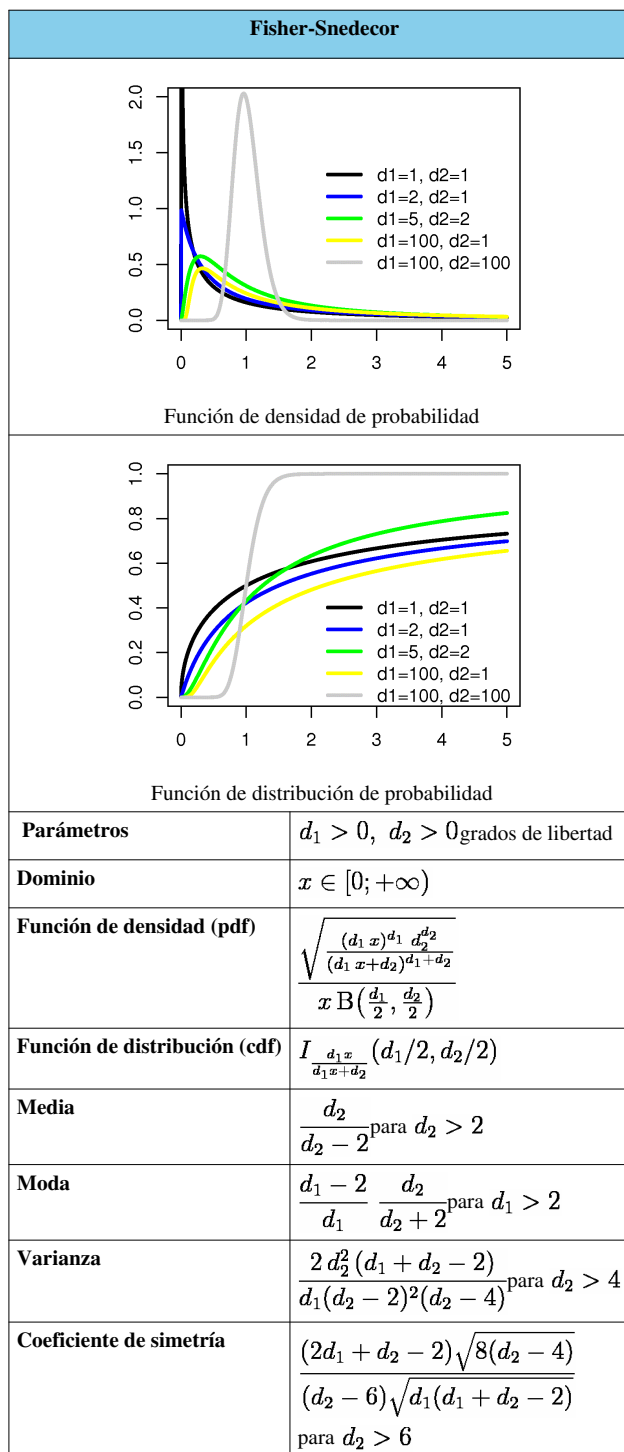
Referencias

[1] Walpole, Roland; Myers, Raymond y Ye, Keying (2002). *Probability and Statistics for Engineers and Scientists*. Pearson Education.

Enlaces externos

- Tabla de distribución de T de Student (<http://tablas-estadisticas.blogspot.com/2010/06/t-de-student.html>)
- Prueba t de Student en la UPTC de Colombia (<http://virtual.uptc.edu.co/ova/estadistica/docs/libros/tstudent.pdf>)
- Tabla distribución t de Student
- Distribución t-Student: Puntos porcentuales para probabilidad superior (http://www.vaxasoftware.com/doc_edu/mat.html)
- (<http://cajael.com/mestadisticos/T7DContinuas/node11.php>) Calcular la probabilidad de una distribución t-Student con R (lenguaje de programación)

Distribución F



Usada en teoría de probabilidad y estadística, la **distribución F** es una distribución de probabilidad continua. También se la conoce como **distribución F de Snedecor** (por George Snedecor) o como **distribución F de Fisher-Snedecor**.

Una variable aleatoria de distribución *F* se construye como el siguiente cociente:

$$F = \frac{U_1/d_1}{U_2/d_2}$$

donde

- U_1 y U_2 siguen una distribución chi-cuadrado con d_1 y d_2 grados de libertad respectivamente, y
- U_1 y U_2 son estadísticamente independientes.

La distribución F aparece frecuentemente como la *distribución nula* de una prueba estadística, especialmente en el análisis de varianza. Véase el test F .

La función de densidad de una $F(d_1, d_2)$ viene dada por

$$g(x) = \frac{1}{B(d_1/2, d_2/2)} \left(\frac{d_1 x}{d_1 x + d_2} \right)^{d_1/2} \left(1 - \frac{d_1 x}{d_1 x + d_2} \right)^{d_2/2} x^{-1}$$

para todo número real $x \geq 0$, donde d_1 y d_2 son enteros positivos, y B es la función beta.

La función de distribución es

$$G(x) = I_{\frac{d_1 x}{d_1 x + d_2}}(d_1/2, d_2/2)$$

donde I es la función beta incompleta regularizada.

Distribuciones relacionadas

- $Y \sim \chi_{\nu_1}^2$ es una distribución ji-cuadrada cuando $Y = \lim_{\nu_2 \rightarrow \infty} \nu_1 X$ para $X \sim F(\nu_1, \nu_2)$.

Enlaces externos

- Tabla de valores críticos de una distribución F ^[1]
- Prueba de significación mediante la distribución F ^[2]
- Distribution Calculator ^[3] Calcula las probabilidades y valores críticos para las distribuciones normal, t , ji-cuadrada y F
- [4] Calcular la probabilidad de una distribución F-Snedecor con R (lenguaje de programación)

Referencias

- [1] <http://www.itl.nist.gov/div898/handbook/eda/section3/eda3673.htm>
 [2] <http://home.clara.net/sisa/signhlp.htm>
 [3] http://www.vias.org/simulations/simusoftware_distcalc.html
 [4] <http://cajael.com/mestadisticos/T7DCContinuas/node17.php>

Problema Central del Límite

Desigualdad de Chebyshev

En probabilidad, la **desigualdad de Chebyshev** (habitualmente también escrito como "Tchebycheff") es un resultado que ofrece una cota inferior a la probabilidad de que el valor de una variable aleatoria con varianza finita esté a una cierta distancia de su esperanza matemática. La desigualdad recibe su nombre del matemático ruso Pafnuti Chebyshev.

Formulación

Si X es una variable aleatoria de media μ y varianza finita σ^2 , entonces, para todo número real $k > 0$,

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}.$$

Sólo en caso de que $k > 1$ la desigualdad proporcionan una cota no trivial.

Ejemplos

Para ilustrar este resultado, supongamos que los artículos de Wikipedia tienen una extensión media de 1000 caracteres y una desviación típica de 200 caracteres. De la desigualdad de Chebyshev, usando $k = 2$, se deduce que al menos el 75% de los artículos tendrán una extensión comprendida entre 600 y 1400 caracteres.

Otra consecuencia del teorema es que para cada distribución de media μ y desviación típica finita σ , al menos la mitad de sus valores se concentrarán en el intervalo $(\mu - \sqrt{2}\sigma, \mu + \sqrt{2}\sigma)$.

Demostración

Para demostrar la desigualdad se parte de la variable aleatoria auxiliar Y definida así:

$$Y = \begin{cases} 0 & \text{si } |X - \mu| \leq k\sigma \\ 1 & \text{si } |X - \mu| > k\sigma \end{cases}$$

Entonces, trivialmente,

$$k\sigma Y \leq |X - \mu|$$

y por lo tanto,

$$k^2\sigma^2 Y^2 \leq (X - \mu)^2.$$

Tomando esperanzas en ambos miembros se obtiene

$$k^2\sigma^2 E[Y^2] \leq E[(X - \mu)^2] = \sigma^2,$$

por lo que

$$E[Y^2] \leq \frac{1}{k^2}.$$

Pero, a su vez, dado que Y sólo puede ser 0 o 1,

$$E[Y^2] = P(Y = 1) = P(|X - \mu| > k\sigma),$$

lo que prueba el resultado.

Discusión

La cota proporcionadas por la desigualdad de Chebyshev, en general, no se pueden mejorar. De hecho es posible construir una variable aleatoria cuya cota de Chebyshev coincida con las probabilidad real. Sin embargo, en casos concretos el teorema proporciona cotas poco precisas.

El teorema puede ser útil a pesar de las cotas imprecisas porque se aplica a una amplia gama de variables que incluye las que están muy alejadas de la distribución normal, y porque las cotas son fáciles de calcular. El teorema se emplea para demostrar la ley débil de los grandes números.

Véase también

- Desigualdad de Márkov.

Desigualdad de Márkov

En teoría de la probabilidad, **la desigualdad de Márkov** proporciona una cota superior para la probabilidad de que una función no negativa de una variable aleatoria sea mayor o igual que una constante positiva. Su nombre le viene del matemático ruso Andréi Márkov.

La desigualdad de Márkov relaciona las probabilidades con la esperanza matemática y proporciona cotas útiles -aunque habitualmente poco ajustadas- para la función de distribución de una variable aleatoria.

Teorema

La desigualdad de Márkov afirma que si X es una variable aleatoria cualquiera y $a > 0$, entonces

$$\Pr(|X| \geq a) \leq \frac{\mathbf{E}(|X|)}{a}.$$

Prueba

Para cualquier suceso A , sea I_A la variable aleatoria indicatriz de A , esto es, $I_A = 1$ si ocurre A y es 0 en el caso contrario. Entonces

$$aI_{(|X| \geq a)} \leq |X|.$$

Por lo tanto

$$\mathbf{E}(aI_{(|X| \geq a)}) \leq \mathbf{E}(|X|).$$

Ahora, nótese que el lado izquierdo de esta desigualdad coincide con

$$a\mathbf{E}(I_{(|X| \geq a)}) = a\Pr(|X| \geq a).$$

Por lo tanto tenemos

$$a\Pr(|X| \geq a) \leq \mathbf{E}(|X|)$$

y como $a > 0$, se pueden dividir ambos lados entre a .

Prueba alternativa

Una prueba más formal, relacionada con el análisis real, es la siguiente:

$$\Pr(|X| \geq a) = \int_a^\infty f(x)dx \leq \int_a^\infty \frac{|x|}{a} f(x)dx \leq \frac{1}{a} \int_{-\infty}^\infty |x|f(x)dx = \frac{E(|X|)}{a}$$

En la introducción de $\frac{|x|}{a}$, nótese que ya que estamos considerando la variable aleatoria sólo en sus valores iguales

o mayores a a , $|X| \geq a$ y, por tanto, $\frac{|X|}{a} \geq 1$, con lo que al multiplicar $f(x)dx$ por algo mayor a uno será

igual o mayor. La segunda desigualdad viene de añadir la suma $\int_{-\infty}^a |x|f(x)dx$, que siempre será positiva ya que se integra algo positivo como es el valor absoluto (por : $f(x)$ que es positiva).

- La desigualdad de Márkov se emplea para probar la Desigualdad de Chebyshev.

Ley de los grandes números

En la teoría de la probabilidad, bajo el término genérico de **ley de los grandes números** se engloban varios teoremas que describen el comportamiento del promedio de una sucesión de variables aleatorias conforme aumenta su número de ensayos.

Estos teoremas prescriben condiciones suficientes para garantizar que dicho promedio converge (en los sentidos explicados abajo) al promedio de las esperanzas de las variables aleatorias involucradas. Las distintas formulaciones de la ley de los grandes números (y sus condiciones asociadas) especifican la convergencia de formas distintas.

Las leyes de los grandes números explican por qué el promedio de una muestra al azar de una población de gran tamaño tenderá a estar cerca de la media de la población completa.

Cuando las variables aleatorias tienen una varianza finita, el teorema central del límite extiende nuestro entendimiento de la convergencia de su promedio describiendo la distribución de diferencias estandarizadas entre la suma de variables aleatorias y el valor esperado de esta suma: sin importar la distribución subyacente de las variables aleatorias, esta diferencia estandarizada converge a una variable aleatoria normal estándar.

La frase "ley de los grandes números" es también usada ocasionalmente para referirse al principio de que la probabilidad de que cualquier evento posible (incluso uno improbable) ocurra *al menos una vez* en una serie, incrementa con el número de eventos en la serie. Por ejemplo, la probabilidad de que un individuo gane la lotería es bastante baja; sin embargo, la probabilidad de que *alguien* gane la lotería es bastante alta, suponiendo que suficientes personas compren boletos de lotería.

Ley débil

La **ley débil de los grandes números** establece que si X_1, X_2, X_3, \dots es una sucesión infinita de variables aleatorias independientes que tienen el mismo valor esperado μ y varianza σ^2 , entonces el promedio

$$\bar{X}_n = (X_1 + \dots + X_n)/n$$

converge en probabilidad a μ . En otras palabras, para cualquier número positivo ε se tiene

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1.$$

Ley fuerte

La **ley fuerte de los grandes números** establece que si X_1, X_2, X_3, \dots es una sucesión infinita de variables aleatorias independientes e idénticamente distribuidas que cumplen $E(|X_i|) < \infty$ y tienen el valor esperado μ , entonces

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1,$$

es decir, el promedio de las variables aleatorias converge a μ casi seguramente (en un conjunto de probabilidad 1).

Esta ley justifica la interpretación intuitiva de que el valor esperado de una variable aleatoria como el "promedio a largo plazo al hacer un muestreo repetitivo".

Véase también

- Teorema central del límite
- Teorema de Bernoulli
- Falacia del jugador
- Andréi Kolmogórov

Referencias

- David Pollard, *A user's guide to measure theoretic probability*, Cambridge University Press (2003).

Teorema del límite central

El **teorema del límite central** o **teorema central del límite** indica que, en condiciones muy generales, si S_n es la suma de n variables aleatorias independientes, entonces la función de distribución de S_n «se aproxima bien» a una distribución normal (también llamada *distribución gaussiana*, *curva de Gauss* o *campana de Gauss*). Así pues, el teorema asegura que esto ocurre cuando la suma de estas variables aleatorias e independientes es lo suficientemente grande.^{[1] [2]}

Definición

Sea $\mathcal{N}(\mu, \sigma^2)$ la función de densidad de la distribución normal definida como^[1]

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

con una media μ y una varianza σ^2 . El caso en el que su función de densidad es $\mathcal{N}(0, 1)$, a la distribución se le conoce como normal estándar.

Se define S_n como la suma de n variables aleatorias, independientes, idénticamente distribuidas, y con una media μ y varianza σ^2 finitas ($\sigma^2 \neq 0$):

$$S_n = X_1 + \dots + X_n$$

de manera que, la media de S_n es $n \cdot \mu$ y la varianza $n \cdot \sigma^2$, dado que son variables aleatorias independientes. Con tal de hacer más fácil la comprensión del teorema y su posterior uso, se hace una estandarización de S_n como

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

para que la media de la nueva variable sea igual a 0 y la desviación estándar sea igual a 1. Así, las variables Z_n convergerán en distribución a la distribución normal estándar $N(0,1)$, cuando n tienda a infinito. Como consecuencia, si $\Phi(z)$ es la función de distribución de $N(0,1)$, para cada número real z :

$$\lim_{n \rightarrow \infty} \Pr(Z_n \leq z) = \Phi(z)$$

donde $\Pr(\cdot)$ indica probabilidad y \lim se refiere a límite matemático.

Enunciado formal

De manera formal, normalizada y compacta el enunciado del teorema es:^[3]

Teorema del límite central: Sea X_1, X_2, \dots, X_n un conjunto de variables aleatorias, independientes e idénticamente distribuidas con media μ y varianza σ^2 distinta de cero. Sea

$$S_n = X_1 + \dots + X_n$$

Entonces

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z \right) = \Phi(z).$$

Es muy común encontrarlo con la variable estandarizada Z_n en función de la media muestral \bar{X}_n ,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}},$$

puesto que son equivalentes, así como encontrarlo en versiones no normalizadas como puede ser:^[4] ^[5]

Teorema (del límite central): Sea X_1, X_2, \dots, X_n un conjunto de variables aleatoria, independientes e idénticamente distribuidas de una distribución con media μ y varianza $\sigma^2 \neq 0$. Entonces, si n es suficientemente grande, la variable aleatoria

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

tiene aproximadamente una distribución normal con $\mu_{\bar{X}} = \mu$ y $\sigma_{\bar{X}}^2 = \sigma^2/n$.

Nota: es importante remarcar que este teorema no dice nada acerca de la distribución de X_i , excepto la existencia de media y varianza.^[4]

Propiedades

- El teorema del límite central garantiza una distribución normal cuando n es suficientemente grande.
- Existen diferentes versiones del teorema, en función de las condiciones utilizadas para asegurar la convergencia. Una de las más simples establece que es suficiente que las variables que se suman sean independientes, idénticamente distribuidas, con valor esperado y varianza finitas.
- La aproximación entre las dos distribuciones es, en general, mayor en el centro de las mismas que en sus extremos o colas, motivo por el cual se prefiere el nombre "teorema del límite central" ("central" califica al límite, más que al teorema).
- Este teorema, perteneciente a la teoría de la probabilidad, encuentra aplicación en muchos campos relacionados, tales como la inferencia estadística o la teoría de renovación.


Véase también

- Ley de los grandes números
- Distribución normal
- Teorema de De Moivre-Laplace

Referencias

- [1] Filmus, Yuval (Enero/Febrero 2010) (en inglés). *Two Proofs of the Central Limit Theorem* (<http://www.cs.toronto.edu/~yuvalf/CLT.pdf>). pp. 1-3. . Consultado el 13-12-2010.
 - [2] Grinstead, Charles M.; Snell, J. Laurie (1997). «9. Central Limit Theorem (http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/Chapter9.pdf)» (en inglés, PDF). *Introduction to Probability* (<http://books.google.es/books?id=14oq4uWGckwC>) (2 edición). AMS Bookstore. pp. 325-360. ISBN 0821807498. . Consultado el 15/04/2009.
 - [3] Charles Stanton. «Central limit theorem (<http://www.math.csusb.edu/faculty/stanton/probstat/clt.html>)» (en inglés). *Probability and Statistics Demos* (<http://www.math.csusb.edu/faculty/stanton/>). Consultado el 13 de diciembre de 2010.
 - [4] Wasserman, Larry. «5. Convergence of Random Variables» (en inglés). *All of Statistics*. Springer. p. 77. ISBN 0-387-40272-1.
 - [5] *Weisstein, Eric W., «Central Limit Theorem (<http://mathworld.wolfram.com/CentralLimitTheorem.html>)» (en inglés), *MathWorld*, Wolfram Research, .
- Blaiotta, Jimena; Delieutraz, Pablo (30 de julio de 2004). «Teorema central del límite (https://www.u-cursos.cl/ingenieria/2009/2/MA3401/1/material_docente/bajar?id_material=260765)» (en castellano) (PDF). Consultado el 15 de diciembre de 2010.
 - Behar Gutiérrez, Roberto; Grima Cintas, Pere (2004) (en castellano). *55 respuestas a dudas típicas de Estadística*. Madrid: Ediciones Díaz de Santos, S.A. pp. 187-189. ISBN 84-7978-643-4.

Enlaces externos

-  Wikimedia Commons alberga contenido multimedia sobre **Teorema del límite central**. Commons

Fuentes y contribuyentes del artículo

Axiomas de probabilidad *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50384074> *Contribuyentes:* AstroNomo, Bonnot, Cgb, Diegusjaimes, Dodo, Egaida, Emijrp, Fenicio, Fibonacci, Flobo, GermanX, Hlnodovic, Hu12, JakobVoss, Joseangelmadrid, Joseaperez, Juan Mayordomo, Jurgens, Maldoror, Matdrodes, Moriel, Netito777, PoLuX124, Rdaneel, Sabbut, Ttpiippo, Wewe, Xenoforme, 67 ediciones anónimas

Sigma-álgebra *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=48936141> *Contribuyentes:* Bunder, CSTAR, Equi, Fibonacci, Grizzly Sigma, HUB, Jose32, Ricardogpn, Richard8933, Sardur, Wewe, 17 ediciones anónimas

Dominio de definición *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=48552107> *Contribuyentes:* DRaco Heroicus, Davius, Elwikipedista, Farisori, Gastonguridi, Gimlinu, Götz, HUB, Ingenioso Hidalgo, Javier Arturo Ollarves Rojas, Matdrodes, Moriel, Noprado, Rehernan, Renacimiento, Resped, Spirit-Black-Wikipedista, Tano4595, 47 ediciones anónimas

Función de densidad de probabilidad *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50713438> *Contribuyentes:* José, Af3, Alexv86, Carter17, Cgb, Davius, Diegusjaimes, Dodo, Farisori, GermanX, Humbefa, Jmcalderon, JorgeGG, Juan Mayordomo, Lobillo, PabloCastellano, Paintman, Pilaf, Tano4595, Ybenitez, 33 ediciones anónimas

Variable aleatoria *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50530930> *Contribuyentes:* Alex 15090, Allfrouros, AlvaroEstadisticaCEU, Camilo, Carlos Manuel Nina, CayoMarcio, Cogliatti Juan i, Dany yun, Davidmosen, Davius, Diegusjaimes, Emijrp, Estadística 2009, Evaromero, Ezarate, Farisori, Gafotas, GermanX, Greek, Guilloip, Gökhan, Hiperfelix, Huhsunqu, Humbefa, Humberto, Icvav, Isha, JViejo, JakobVoss, Joseangelmadrid, Joseaperez, Juan Mayordomo, Juancdg, Laura Fiorucci, Luisedu90, Matdrodes, Metróno, Neodop, Numbo3, Pino, PoLuX124, Rastrojo, Samcienfuegos, Tartaglia, Wewe, Ybenitez, 108 ediciones anónimas

Esperanza matemática *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=49691666> *Contribuyentes:* -antonio-, Alchaemist, AlfonsoERomero, Alvaro9, Antony1204, Diegusjaimes, Ernesto Graf, Farisori, Fenicio, GermanX, JakobVoss, Jmcalderon, JoseA, Joseaperez, Juan Manuel, Juan Mayordomo, Jynus, MaSt, Matdrodes, Netito777, Nogueiras, Proximo.xv, Rdaneel, Sabbut, Taragui, Tartaglia, Template namespace initialisation script, 60 ediciones anónimas

Media aritmética *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50684493> *Contribuyentes:* Alakasm, AlfonsoERomero, Alopezruiz, Alvaro qc, Andeba, Açipni-Lovrij, Baiji, Benedicto, BetoCG, BlackBeast, Cdlfd, Cobalttempest, Correogsk, Cyberdelic, David0811, Davius, Diego 5397, Diegusjaimes, Dreitmen, Eduardosalg, Egaida, Elwikipedista, Er Komandante, Gothmog, Gökhan, Götz, Halfdrag, JMCC1, JakobVoss, Jarisleif, Jgalgarra, Jkbw, JorgeGG, Josamaga, Joseaperez, Juan José Moral, Juan Mayordomo, Jugones55, KnightRider, L30nc1t0, Lwk, Madalberta, Magister Mathematicae, Maldoror, Mar del Sur, Marcelo Forets, Maria Patricia Gonzalez, Matdrodes, Mel 23, Melocoton, Mgamezm, Moriel, Mxtintin, Netito777, ONDIA, PACO, Pablo.cl, Paracelsiux, Pinar, PoLuX124, Ricardogpn, Rod 93, Sabbut, Saloca, Sanbec, Snakeyes, Sorareader, Superzerocool, Tamorian, Taragui, Tartaglia, Template namespace initialisation script, Tirithel, Tortillovsky, Yavidaxiu, Yeza, Youssefsan, Yrithinnd, Zill, Zulucho, 201 ediciones anónimas

Mediana (estadística) *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50680484> *Contribuyentes:* AlfonsoERomero, Almendro, Anatoli1024, Andreaemperu, Camilo, Chidalgo, Cobalttempest, CommonsDelinker, Cratón, DJ Nietzsche, Dark0101, Dav7mx, Delphidius, Diegusjaimes, Eduardosalg, Emijrp, Escarlati, Euratom, Ezzeje, GNM, Gabrielsvb, HUB, Hameryko, Hprmedina, Javierito92, Javitotengo, Jkbw, Joseaperez, Jtico, Juan Mayordomo, Leonpolanco, Libertad y Saber, Locos epraix, Mafores, Manwë, Mecamático, Pablo Alcaiyaga, Pason, PoLuX124, Raco, Rafiko77, Rbonvall, Retama, RobertoAlva, Rubenherrera, Savh, Sergio Andres Segovia, Super braulio, Tartaglia, Ticqueretaro, Tirithel, Uruk, Vaskop, Vecellio, X7iBiT, Zulucho, 227 ediciones anónimas

Moda (estadística) *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50696799> *Contribuyentes:* Angel GN, BetoCG, BlackBeast, BuenaGente, Camilo, Crescent Moon, DJ Nietzsche, DeosamoX, Diegusjaimes, Dorico, Dossier2, Dracktrane, Dronkit, Ekran, Evaristokbza, Gugsus, HUB, Isha, Israelvalenciac, JMPerez, Javal, Jkbw, JorgeGG, Joseaperez, Juan Mayordomo, Libertad y Saber, Maldoror, Netito777, PoLuX124, Porao, Purpu, Rafiko77, Rastrojo, Rubenherrera, Sabbut, Savh, Super braulio, Tartaglia, Technopat, Tirithel, Wikiléptico, Yairhardy9, 140 ediciones anónimas

Varianza *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50022483> *Contribuyentes:* Adrien, Alakasm, Aremar00, Carloschida, Cgb, Ctrl Z, Diegusjaimes, Er Komandante, Evaromero, FbPort, Fcojperz, Fenicio, Francisco Albani, GermanX, HUB, Hprmedina, Humbefa, Ilmer Condor, JakobVoss, Jgibaja, Jkbw, Jmcalderon, Joseaperez, Juan Manuel, Juan Mayordomo, Katerin jimena, Laurang, Madalberta, ManuP, Matdrodes, Mel 23, Muro de Aguas, Mxcatania, Phirosiberia, Ravave, Retama, Roberto Fiadone, RoyFocker, Savh, Stoni, Template namespace initialisation script, Tirithel, Triku, Uncertain, Ungoliant, Xenoforme, Žiedas, 155 ediciones anónimas

Asimetría estadística *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50441514> *Contribuyentes:* Dhidalgo, Eva R M, GermanX, Hprmedina, Jgalgarra, Juan Mayordomo, Vitamine, Yuikmy, 16 ediciones anónimas

Desviación estándar *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50474624> *Contribuyentes:* Alakasm, Alcmos, Alhen, Anacardo, AngieGM, Açipni-Lovrij, Bamanishia, Bazookao, Bryansalazar, Chesnok, Cobalttempest, DamianFinol, DarkMars, DaveFX, Davidge, Diegusjaimes, Doggui, Dominican, Dvdcrojas, EL AGUSTIN, CLARO, Eduardosalg, Erescepemi, Fanwandler, Fiquei, Fz, Gaius iulius caesar, Halfdrag, Hampcky, Isha, JAGT, Ja.esparza, Jarisleif, Javierito92, Jfajfaj, Jkbw, Jorge C.Al, JorgeGG, Josamaga, Joseaperez, Juan Manuel, Juan Mayordomo, Jynus, Leonpolanco, Limbo@MX, Losmaspintas, Luis lencina, Madalberta, Maldoror, Manuelt15, Martincarr, Matdrodes, Mecamático, Moriel, Muro de Aguas, Nemo, Netito777, Paintman, Periku, PoLuX124, Rafiko77, Raulsch, Rellu, Retama, Richy, Roberpl, Salvor Hardin, Sauron, Sebrev, Srbnana, Stoni, Taichi, Taragui, Tartaglia, Tirithel, Tomateje, Tostadora, Triku, Typhoon, Uncronopio, Vitamine, Xenoforme, Yeza, Zaskie, 286 ediciones anónimas

Covarianza *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50638666> *Contribuyentes:* Ascánder, Cgb, Diegusjaimes, Eargones, Elwikipedista, Gmagno, Icvav, JRuizJim, JakobVoss, Jgoye, Jkbw, Joseaperez, Juan José Moral, Juan Mayordomo, Pablosreitano, Paintman, Pakebal, Pertile, Retama, Shalafy, Srengel, Tano4595, Tirithel, Triku, Yeza, 58 ediciones anónimas

Correlación *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50650908> *Contribuyentes:* Alhen, Bucho, Camilo, Egozcue, El Quinche, Grillitus, Humberto, Juan Mayordomo, Laurang, Matdrodes, Mxcatania, Redeyes, Tirithel, 41 ediciones anónimas

Regresión lineal *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50112225> *Contribuyentes:* AldanaN, Alefisisco, Alfonso Aguilar, Alvaro qc, Antón Francho, Banfield, Carro e, Chomolungma, ConPermiso, Daniel De Leon Martinez, Davius, Dhcp, Diego D E, El Quinche, Elwikipedista, Fbport, Felipealvarez, Gaeddal, Gafotas, Gecime, GermanX, Ggenellina, HanPritcher, Hlnodovic, Hucknall, Icchu, Ingenioso Hidalgo, Jcoronelf, JorgeGG, Joseaperez, Juan Mayordomo, Juan renombrado, Kved, Laura Fiorucci, Madmaxsr, Marvelshine, Matdrodes, Mcapdevila, Pacomegia, Raulsch, Riviera, Rjgalindo, Rouxfederico, Sergio Andres Segovia, Snakeeater, Super braulio, Tano4595, Tirithel, TorQue Astur, Vitamine, Yeza, 131 ediciones anónimas

Función generadora de momentos *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=47699010> *Contribuyentes:* Ricardogpn, Tartaglia, 2 ediciones anónimas

Distribución de probabilidad *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50300712> *Contribuyentes:* Af3, AlfonsoERomero, Artorius, Carlosfran2ceu, Cflm001, Cgb, Ctrl Z, Davius, Diegusjaimes, Dnu72, Dodo, Enen, Evra83, Farisori, Germanrinconrey, Immersia, Interwiki, J.delanoy, JakobVoss, JorgeGG, Joseaperez, Juan Mayordomo, Juanwrs, Juliabis, Kved, Lucien leGrey, Maldoror, Manuelt15, Manwë, Maquita, Matdrodes, Miss Manzana, Moriel, Neozonik, PabloAlv, PabloCastellano, Paulrc, Peregrino, Petrus, Sabbut, Sebrev, Tano4595, Umarth8, Vitamine, Zahualli, 109 ediciones anónimas

Distribución uniforme continua *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50112978> *Contribuyentes:* *TikiTac*, AlfonsoERomero, Califasusoso, Cgb, Farisori, Humbefa, Juan Mayordomo, Kurt86, Lobillo, Moriel, Rastrojo, Ricardogpn, Ronald2308, Uruk, 18 ediciones anónimas

Distribución gamma *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50519438> *Contribuyentes:* Af3, JakobVoss, Javier Jelovcan, JorgeGG, Joseaperez, Phirosiberia, Sfindino, Tano4595, 26 ediciones anónimas

Distribución exponencial *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50520016> *Contribuyentes:* Af3, Alberto Salguero, Comosea, ConPermiso, Elbarak, Farisori, GermanX, JakobVoss, Javier Jelovcan, Jerowiki, Joseaperez, Kved, Mafucl, Matdrodes, Moriel, Nakp, Peejayem, Phirosiberia, Roberto Pablo CASTILLO, Rufflos, Schummy, Taichi, Template namespace initialisation script, Wesisnay, Wilmer32, 40 ediciones anónimas

Distribución beta *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50517296> *Contribuyentes:* Af3, Gökhan, Humbefa, JakobVoss, JorgeGG, Joseaperez, Juan Mayordomo, Moriel, Stoni, Tano4595, 4 ediciones anónimas

Distribución normal *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50607700> *Contribuyentes:* A ver, Af3, Airunp, Alexv86, AlfonsoERomero, Antur, AstroNomo, B1mbo, BlackBeast, BuenaGente, Carlos.Gracia-Lázaro, Cgb, Chesnok, Christianfge, ConPermiso, Dhcp, Diegusjaimes, Dodo, Doloco, Edmenb, Eduardosalg, Er Komandante, Euratom, Farisori, Fsd141, Germanrinconrey, Gperjim, Guanucoluis, Gökhan, HiTe, Humbefa, Jerowiki, Jkbw, JoeLoui, Jorge c2010, JorgeGG, JoseA, Joseaperez, Joxemai, Juan Carlos Tapia, Juan Manuel, Juan

Mayordomo, LP, Leonpolanco, Matdroses, Moonkey, Omary22 24, Oscar ., Pasmargo, Paulrc, Rafiko77, Ricardogpn, Roche, Rufflos, SPZ, Savh, Sergio Andres Segovia, Srbanana, Taichi, Tano4595, Tartaglia, Tirithel, Tomatej, Vivero, Xenoforme, 161 ediciones anónimas

Distribución binomial *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50722150> *Contribuyentes:* .Sergio, Akma72, Alex economist, Amanuense, Babbage, Bentzia, Camilo, Cgb, Danielyapahl, Darizabalo, Diegusjaimes, Dreitmen, Farisori, Fvmete, GermanX, Grillitus, JAGT, Jkbw, Juan Mayordomo, Juan carvacho, Juliwolfgang, Kved, Magister Mathematicae, Mahadeva, Matdroses, Mpeinadopa, Murphy era un optimista, Paintman, PoLuX124, Porao, Raulshc, Ricardogpn, Soteke, Supercyberedgar, Tano4595, Tartaglia, Tostadora, Vaskop, Walterotta, Yogobah, 125 ediciones anónimas

Distribución multinomial *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=49858004> *Contribuyentes:* EST-Tr-1, Farisori, Superzerocool, 1 ediciones anónimas

Distribución de Poisson *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50639965> *Contribuyentes:* Aldo david, Alex economist, Amanuense, Camilo, Cgb, Ciberjopower, Diegusjaimes, Flakinho, Ictlogist, JAGT, JakobVoss, Juan Mayordomo, Julian Colina, Juliwolfgang, Kved, Magister Mathematicae, Megazilla77, Mrzeon, Paintman, Pieter, Pybalo, Rufflos, Super braulio, Tano4595, 124 ediciones anónimas

Distribución hipergeométrica *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50689436> *Contribuyentes:* Alexv86, Cgb, JakobVoss, Joseaperez, Juan Mayordomo, Tano4595, 25 ediciones anónimas

Distribución normal multivariante *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=49857855> *Contribuyentes:* Cgb, Farisori, Juan Mayordomo, KurtSchwitters, Muro de Aguas, Poco a poco, Ricardogpn, Tartaglia, 3 ediciones anónimas

Distribución χ^2 *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50689438> *Contribuyentes:* Af3, Aiax, Alefísico, AlfonsoERomero, Ampersand &, AngelHeraez, Cgb, Cristiagy, Esetores, Fgiones, HiTe, Humberto, JakobVoss, Jorge c2010, JorgeGG, Joseaperez, Juan Manuel, Juan Mayordomo, Kved, Madalberta, NudoMarinero, Penetic, Resped, Sabbut, Tano4595, Toad32767, Wissons, 37 ediciones anónimas

Distribución t de Student *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50264486> *Contribuyentes:* Af3, Canyq, Cgb, Ciberelm, Diegusjaimes, Farisori, Fiquei, Folkvanger, HiTe, Jkbw, Jtico, Juan Manuel, Juan Mayordomo, Leonpolanco, Lucien leGrey, Matdroses, Morini, Plalopez, Rufflos, Tamorlan, Tartaglia, 61 ediciones anónimas

Distribución F *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50084749> *Contribuyentes:* Ciberelm, Eselito5, FAR, Gperjim, Humberto, Juan Manuel, Juan Mayordomo, Paulrc, Tano4595, 11 ediciones anónimas

Desigualdad de Chebyshev *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50565806> *Contribuyentes:* Af3, Cektron, Cgb, Comae, Davius, Er Komandante, Fibonacci, GermanX, Hlnodovic, Joseaperez, Juan Mayordomo, Martinvernengo, Pieter, Sabbut, 18 ediciones anónimas

Desigualdad de Márkov *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=50349241> *Contribuyentes:* GermanX, Hlnodovic, Jjgibaja, Juan Mayordomo, Mister, Tamorlan, 4 ediciones anónimas

Ley de los grandes números *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=49859488> *Contribuyentes:* AlfonsoERomero, Cgb, Driver90, Enredanrestos, Erud, GermanX, Humbefa, Humberto, Jaoyola, Jerowiki, Juan Mayordomo, Matdroses, Numbo3, Pegaso2005, Pinar, ReiVaX, Tano4595, Tostadora, 44 ediciones anónimas

Teorema del límite central *Fuente:* <http://es.wikipedia.org/w/index.php?oldid=49871450> *Contribuyentes:* Alcarraz, Alrojo, Belgrano, Correogsk, Diegusjaimes, Elwikipedista, Erick1984, Farisori, JRGL, Jerowiki, Juan Mayordomo, Lmendo, Mar del Sur, Matdroses, Patelapiara, Raulshc, Tano4595, Tartaglia, Wewe, XCesar, XalD, 42 ediciones anónimas

Fuentes de imagen, Licencias y contribuyentes

Archivo:Normal distribution pdf.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Normal_distribution_pdf.png *Licencia:* GNU General Public License *Contribuyentes:* Ardonik, Gerbrant, Grendelkhan, Inductiveload, Juiced lemon, MarkSweep, Wikiwide, 10 ediciones anónimas

Archivo:Commons-logo.svg *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:Commons-logo.svg> *Licencia:* logo *Contribuyentes:* SVG version was created by User:Grunt and cleaned up by 3247, based on the earlier PNG version, created by Reidab.

Archivo:Nuvola apps edu mathematics-p.svg *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Nuvola_apps_edu_mathematics-p.svg *Licencia:* GNU Lesser General Public License *Contribuyentes:* David Vignoni (original icon); Flamurai (SVG conversion)

Archivo:MathematicalMeans.svg *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:MathematicalMeans.svg> *Licencia:* Public Domain *Contribuyentes:* Dzenanz

Archivo:davicegre3.JPG *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:Davicegre3.JPG> *Licencia:* GNU Free Documentation License *Contribuyentes:* Bobmath, Cregenwiki46

Archivo:desarrollomediana.jpg *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:Desarrollomediana.jpg> *Licencia:* Creative Commons Attribution-Sharealike 3.0,2.5,2.0,1.0 *Contribuyentes:* Rubenherrera

Archivo:Imagenmarcos1.JPG *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:Imagenmarcos1.JPG> *Licencia:* Creative Commons Attribution-Sharealike 2.5 *Contribuyentes:* Bobmath, Ekran

Archivo:Imagenmarcos2.JPG *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:Imagenmarcos2.JPG> *Licencia:* Creative Commons Attribution-Sharealike 2.5 *Contribuyentes:* Bobmath, Ekran

Archivo:Imagenmarcos3.JPG *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:Imagenmarcos3.JPG> *Licencia:* Creative Commons Attribution-Sharealike 2.5 *Contribuyentes:* Bobmath, Ekran

Archivo:Desarrollomoda.jpg *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:Desarrollomoda.jpg> *Licencia:* Creative Commons Attribution-Sharealike 3.0,2.5,2.0,1.0 *Contribuyentes:* Rubén Hernández Herrera

Archivo:SkewedDistribution.png *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:SkewedDistribution.png> *Licencia:* Creative Commons Attribution-ShareAlike 3.0 Unported *Contribuyentes:* Audrius Meskauskas

Archivo:Standard deviation diagram (decimal comma).svg *Fuente:* [http://es.wikipedia.org/w/index.php?title=Archivo:Standard_deviation_diagram_\(decimal_comma\).svg](http://es.wikipedia.org/w/index.php?title=Archivo:Standard_deviation_diagram_(decimal_comma).svg) *Licencia:* GNU Free Documentation License *Contribuyentes:* Original uploader was Nusha at sl.wikipedia

Archivo:Linear regression.svg *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Linear_regression.svg *Licencia:* Public Domain *Contribuyentes:* Sewaqu

Archivo:Standard deviation diagram.svg *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Standard_deviation_diagram.svg *Licencia:* Creative Commons Attribution 2.5 *Contribuyentes:* Mwtoews

Archivo:Binomial Distribution.svg *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Binomial_Distribution.svg *Licencia:* GNU Free Documentation License *Contribuyentes:* cflm (talk)

Archivo:Uniform distribution PDF.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Uniform_distribution_PDF.png *Licencia:* Public Domain *Contribuyentes:* EugeneZelenko, It Is Me Here, PAR

Archivo:Uniform distribution CDF.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Uniform_distribution_CDF.png *Licencia:* Public Domain *Contribuyentes:* EugeneZelenko, PAR

Image:Gamma distribution pdf.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Gamma_distribution_pdf.png *Licencia:* GNU General Public License *Contribuyentes:* Alejo2083, Autopilot, Cburnett, Ch1902, It Is Me Here, Liftarn, MarkSweep, Stannered, 1 ediciones anónimas

Archivo:Exponential distribution pdf.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Exponential_distribution_pdf.png *Licencia:* GNU General Public License *Contribuyentes:* Alejo2083, Autopilot, Cburnett, It Is Me Here, MarkSweep

Archivo:Exponential distribution cdf.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Exponential_distribution_cdf.png *Licencia:* GNU General Public License *Contribuyentes:* Alejo2083, Cburnett, MarkSweep

Image:Beta distribution pdf.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Beta_distribution_pdf.png *Licencia:* GNU General Public License *Contribuyentes:* Cburnett, It Is Me Here, Krishnavedala, LeaW, MarkSweep, WikipediaMaster, 1 ediciones anónimas

Image:Beta distribution cdf.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Beta_distribution_cdf.png *Licencia:* GNU General Public License *Contribuyentes:* Krishnavedala, MarkSweep, WikipediaMaster, 1 ediciones anónimas

Archivo:Normal distribution cdf.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Normal_distribution_cdf.png *Licencia:* GNU General Public License *Contribuyentes:* Gerbrant, Inductiveload, Juiced lemon, MarkSweep, Waldir

Archivo:Abraham de moivre.jpg *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Abraham_de_moivre.jpg *Licencia:* Public Domain *Contribuyentes:* Bjh21, Bonzo, Elcobbola, Kilom691, Saippukaappias, 竹炭魚 (Searobin)

Archivo:DisNormal01.svg *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:DisNormal01.svg> *Licencia:* Public Domain *Contribuyentes:* User:HiTe

Archivo:Normal Distribution CDF.svg *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Normal_Distribution_CDF.svg *Licencia:* Public Domain *Contribuyentes:* Inductiveload

Archivo:standard deviation diagram (decimal comma).svg *Fuente:* [http://es.wikipedia.org/w/index.php?title=Archivo:Standard_deviation_diagram_\(decimal_comma\).svg](http://es.wikipedia.org/w/index.php?title=Archivo:Standard_deviation_diagram_(decimal_comma).svg) *Licencia:* GNU Free Documentation License *Contribuyentes:* Original uploader was Nusha at sl.wikipedia

Archivo:Normal approximation to binomial.svg *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Normal_approximation_to_binomial.svg *Licencia:* GNU Free Documentation License *Contribuyentes:* User:MarkSweep

Archivo:Crowd outside nyse.jpg *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Crowd_outside_nyse.jpg *Licencia:* Public Domain *Contribuyentes:* AnRo0002, Echtner, Fnfd, Gribeco, Gryffindor, Hystrix, Infrogmaton, J 1982, Romary, SkeeziX1000, Spuk968, Yerpo, 5 ediciones anónimas

Archivo:Binomial distribution pmf.svg *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Binomial_distribution_pmf.svg *Licencia:* Public Domain *Contribuyentes:* Tayste

Archivo:Binomial distribution cdf.svg *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Binomial_distribution_cdf.svg *Licencia:* Public Domain *Contribuyentes:* Tayste

Archivo:Poisson distribution PMF.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Poisson_distribution_PMF.png *Licencia:* Public Domain *Contribuyentes:* Autopilot, EugeneZelenko, Grafite, It Is Me Here, PAR, 1 ediciones anónimas

Archivo:PoissonCDF.png *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:PoissonCDF.png> *Licencia:* GNU General Public License *Contribuyentes:* Original uploader was Pdbailey at en.wikipedia

Archivo:chi-square distributionPDF.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Chi-square_distributionPDF.png *Licencia:* Public Domain *Contribuyentes:* EugeneZelenko, It Is Me Here, PAR, WikipediaMaster

Archivo:chi-square distributionCDF.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Chi-square_distributionCDF.png *Licencia:* Public Domain *Contribuyentes:* EugeneZelenko, PAR, WikipediaMaster

Archivo:Student densite best.JPG *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Student_densite_best.JPG *Licencia:* Creative Commons Attribution-Sharealike 1.0 *Contribuyentes:* Original uploader was Thorin at fr.wikipedia

Archivo:T distributionCDF.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:T_distributionCDF.png *Licencia:* GNU Free Documentation License *Contribuyentes:* Anarkman, Juiced lemon

Archivo:F distributionPDF.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:F_distributionPDF.png *Licencia:* GNU Free Documentation License *Contribuyentes:* en:User:Pdbailey

Archivo:F distributionCDF.png *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:F_distributionCDF.png *Licencia:* GNU Free Documentation License *Contribuyentes:* en:User:Pdbailey

Licencia

Creative Commons Attribution-Share Alike 3.0 Unported
[//creativecommons.org/licenses/by-sa/3.0/](https://creativecommons.org/licenses/by-sa/3.0/)
