

PROBABILIDAD Y DISTRIBUCIÓN

Probabilidad y Distribución

Contenidos

Artículos

Probabilidad	1
Distribución normal	6
Distribución binomial	23
Distribución de Poisson	26
Proceso de Poisson	30
Regresión de Poisson	32

Referencias

Fuentes y contribuyentes del artículo	34
Fuentes de imagen, Licencias y contribuyentes	35

Licencias de artículos

Licencia	36
----------	----

Probabilidad

La **probabilidad** es un método mediante el cual se obtiene la frecuencia de un suceso determinado mediante la realización de un experimento aleatorio, del que se conocen todos los resultados posibles, bajo condiciones *suficientemente* estables. La teoría de la probabilidad se usa extensamente en áreas como la estadística, la física, la matemática, las ciencias y la filosofía para sacar conclusiones sobre la probabilidad discreta de sucesos potenciales y la mecánica subyacente discreta de sistemas complejos.

Historia

Las probabilidades constituyen una rama de las matemáticas que se ocupa de medir o determinar cuantitativamente la posibilidad de que un suceso o experimento produzca un determinado resultado.

El diccionario de la Real Academia Española define «azar» como una casualidad, un caso fortuito, y afirma que la expresión «al azar» significa «sin orden». La idea de Probabilidad está íntimamente ligada a la idea de azar y nos ayuda a comprender nuestras posibilidades de ganar un juego de azar o analizar las encuestas. Pierre-Simon Laplace afirmó: "Es notable que una ciencia que comenzó con consideraciones sobre juegos de azar haya llegado a ser el objeto más importante del conocimiento humano". Comprender y estudiar el azar es indispensable, porque la probabilidad es un soporte necesario para tomar decisiones en cualquier ámbito.

Según Amanda Dure, "Antes de la mitad del siglo XVII, el término 'probable' (en latín *probable*) significaba *aprobable*, y se aplicaba en ese sentido, unívocamente, a la opinión y a la acción. Una acción u opinión probable era una que las personas sensatas emprenderían o mantendrían, en las circunstancias."^[1]

Aparte de algunas consideraciones elementales hechas por Girolamo Cardano en el siglo XVI, la doctrina de las probabilidades data de la correspondencia de Pierre de Fermat y Blaise Pascal (1654). Christiaan Huygens (1657) le dio el tratamiento científico conocido más temprano al concepto. *Ars Conjectandi* (póstumo, 1713) de Jakob Bernoulli y *Doctrine of Chances* (1718) de Abraham de Moivre trataron el tema como una rama de las matemáticas. Véase *El surgimiento de la probabilidad (The Emergence of Probability)* de Ian Hacking para una historia de los inicios del desarrollo del propio concepto de probabilidad matemática.

La teoría de errores puede trazarse atrás en el tiempo hasta *Opera Miscellanea* (póstumo, 1722) de Roger Cotes, pero una memoria preparada por Thomas Simpson en 1755 (impresa en 1756) aplicó por primera vez la teoría para la discusión de errores de observación. La reimpresión (1757) de esta memoria expone los axiomas de que los errores positivos y negativos son igualmente probables, y que hay ciertos límites asignables dentro de los cuales se supone que caen todos los errores; se discuten los errores continuos y se da una curva de la probabilidad.

Pierre-Simon Laplace (1774) hizo el primer intento para deducir una regla para la combinación de observaciones a partir de los principios de la teoría de las probabilidades. Representó la ley de la probabilidad de error con una curva $y = \phi(x)$, siendo x cualquier error e y su probabilidad, y expuso tres propiedades de esta curva:

1. es simétrica al eje y ;
2. el eje x es una asíntota, siendo la probabilidad del error ∞ igual a 0;
3. la superficie cerrada es 1, haciendo cierta la existencia de un error.

Dedujo una fórmula para la media de tres observaciones. También obtuvo (1781) una fórmula para la ley de facilidad de error (un término debido a Lagrange, 1774), pero una que llevaba a ecuaciones inmanejables. Daniel Bernoulli (1778) introdujo el principio del máximo producto de las probabilidades de un sistema de errores concurrentes.

El método de mínimos cuadrados se debe a Adrien-Marie Legendre (1805), que lo introdujo en su *Nouvelles méthodes pour la détermination des orbites des comètes (Nuevos métodos para la determinación de las órbitas de los cometas)*. Ignorando la contribución de Legendre, un escritor irlandés estadounidense, Robert Adrain, editor de "The Analyst" (1808), dedujo por primera vez la ley de facilidad de error,

$$\phi(x) = ce^{-h^2x^2}$$

siendo c y h constantes que dependen de la precisión de la observación. Expuso dos demostraciones, siendo la segunda esencialmente la misma de John Herschel (1850). Gauss expuso la primera demostración que parece que se conoció en Europa (la tercera después de la de Adrain) en 1809. Demostraciones adicionales se expusieron por Laplace (1810, 1812), Gauss (1823), James Ivory (1825, 1826), Hagen (1837), Friedrich Bessel (1838), W. F. Donkin (1844, 1856) y Morgan Crofton (1870). Otros personajes que contribuyeron fueron Ellis (1844), De Morgan (1864), Glaisher (1872) y Giovanni Schiaparelli (1875). La fórmula de Peters (1856) para r , el error probable de una única observación, es bien conocida.

En el siglo XIX, los autores de la teoría general incluían a Laplace, Sylvestre Lacroix (1816), Littrow (1833), Adolphe Quetelet (1853), Richard Dedekind (1860), Helmert (1872), Hermann Laurent (1873), Liagre, Didion, y Karl Pearson. Augustus De Morgan y George Boole mejoraron la exposición de la teoría.

En 1930 Andréi Kolmogorov desarrolló la base axiomática de la probabilidad utilizando teoría de la medida.

En la parte geométrica (véase geometría integral) los colaboradores de *The Educational Times* fueron influyentes (Miller, Crofton, McColl, Wolstenholme, Watson y Artemas Martin).

Teoría

La probabilidad constituye un importante parámetro en la determinación de las diversas casualidades obtenidas tras una serie de eventos esperados dentro de un rango estadístico.

Existen diversas formas como método abstracto, como la teoría Dempster-Shafer y la teoría de la relatividad numérica, esta última con un alto grado de aceptación si se toma en cuenta que disminuye considerablemente las posibilidades hasta un nivel mínimo ya que somete a todas las antiguas reglas a una simple ley de relatividad.^[cita requerida]

La probabilidad de un evento se denota con la letra p y se expresa en términos de una fracción y no en porcentajes, por lo que el valor de p cae entre 0 y 1. Por otra parte, la probabilidad de que un evento "no ocurra" equivale a 1 menos el valor de p y se denota con la letra q

$$P(Q) = 1 - P(E)$$

Los tres métodos para calcular las probabilidades son la regla de la adición, la regla de la multiplicación y la distribución binomial.

Regla de la adición

La regla de la adición o regla de la suma establece que la probabilidad de ocurrencia de cualquier evento en particular es igual a la suma de las probabilidades individuales, si es que los eventos son mutuamente excluyentes, es decir, que dos no pueden ocurrir al mismo tiempo.

$P(A \text{ o } B) = P(A) \cup P(B) = P(A) + P(B)$ si A y B son mutuamente excluyente. $P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B)$ si A y B son no excluyentes.

Siendo: $P(A)$ = probabilidad de ocurrencia del evento A. $P(B)$ = probabilidad de ocurrencia del evento B. $P(A \text{ y } B)$ = probabilidad de ocurrencia simultánea de los eventos A y B.

Regla de la multiplicación

La regla de la multiplicación establece que la probabilidad de ocurrencia de dos o más eventos estadísticamente independientes es igual al producto de sus probabilidades individuales.

$P(A \text{ y } B) = P(A \cap B) = P(A)P(B)$ si A y B son independientes. $P(A \text{ y } B) = P(A \cap B) = P(A)P(B|A)$ si A y B son dependientes

Regla de Laplace

La **regla de Laplace** establece que:

- La probabilidad de ocurrencia de un suceso *imposible* es 0.
- La probabilidad de ocurrencia de un suceso *seguro* es 1, es decir, $P(A) = 1$.

Para aplicar la regla de Laplace es necesario que los experimentos den lugar a sucesos equiprobables, es decir, que todos tengan o posean la misma probabilidad.

- La probabilidad de que ocurra un suceso se calcula así:

$$P(A) = \text{N}^\circ \text{ de casos favorables} / \text{N}^\circ \text{ de resultados posibles}$$

Esto significa que: la probabilidad del evento A es igual al cociente del número de casos favorables (los casos dónde sucede A) sobre el total de casos posibles.

Distribución binomial

La probabilidad de ocurrencia de una combinación específica de eventos independientes y mutuamente excluyentes se determina con la distribución binomial, que es aquella donde hay solo dos posibilidades, tales como masculino/femenino o si/no.

1. Hay dos resultados posibles mutuamente excluyentes en cada ensayo u observación.
2. La serie de ensayos u observaciones constituyen eventos independientes.
3. La probabilidad de éxito permanece constante de ensayo a ensayo, es decir el proceso es estacionario.

Para aplicar esta distribución al cálculo de la probabilidad de obtener un número dado de éxitos en una serie de experimentos en un proceso de Bernoulli, se requieren tres valores: el número designado de éxitos (m), el número de ensayos y observaciones (n); y la probabilidad de éxito en cada ensayo (p).

Entonces la probabilidad de que ocurran m éxitos en un experimento de n ensayos es:

$$P(x = m) = (nC_m)(P^m)(1-P)^{n-m}$$

Siendo: nC_m el número total de combinaciones posibles de m elementos en un conjunto de n elementos.

$$\text{En otras palabras } P(x = m) = [n!/(m!(n-m)!)](p^m)(1-p)^{n-m}$$

Ejemplo. La probabilidad de que un alumno apruebe la asignatura Cálculo de Probabilidades es de 0,15. Si en un semestre intensivo se inscriben 15 alumnos ¿Cuál es la probabilidad de que aprueben 10 de ellos?

$P(x = 10) = 15C_{10}(0,15)^{10}(0,85)^5 = 10!/(10!(15-10)!)(0,15)^{10}(0,85)^5 = 7,68 * 10^{-6}$ Generalmente existe un interés en la probabilidad acumulada de "m o más" éxitos o "m o menos" éxitos en n ensayos. En tal caso debemos tomar en cuenta que:

$$P(x < m) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + \dots + P(x = m - 1)$$

$$P(x > m) = P(x = m + 1) + P(x = m + 2) + P(x = m + 3) + \dots + P(x = n)$$

$$P(x \leq m) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + \dots + P(x = m)$$

$$P(x \geq m) = P(x = m) + P(x = m + 1) + P(x = m + 2) + \dots + P(x = n)$$

Supongamos que del ejemplo anterior se desea saber la probabilidad de que aprueben:

a.- al menos 5

b.- más de 12

a.- la probabilidad de que aprueben al menos 5 es:

$P(x \geq 5)$ es decir, que:

$$1 - P(x < 5) = 1 - [P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4)] =$$

$$1 - [0,0874 + 0,2312 + 0,2856 + 0,2184 + 0,1156] = 0,0618$$

Nota: Al menos, a lo menos y por lo menos son locuciones adverbiales sinónimas.

Ejemplo: La entrada al cine por lo menos tendrá un costo de 10 soles (como mínimo podría costar 10 soles o más).

b.- la probabilidad de que aprueben más de 12 es $P(x > 12)$ es decir, que:

$$P(x > 12) = P(x = 13) + P(x = 14) + P(x = 15)$$

$$P(x > 12) = 1,47 * 10^{-9} + 3,722 * 10^{-11} + 4,38 * 10^{-13} = 1,507 * 10^{-9}$$

La esperanza matemática en una distribución binomial puede expresarse como:

$$E(x) = np = 15(0,15) = 2,25$$

Y la varianza del número esperado de éxitos se puede calcular directamente:

$$\text{Var}(x) = np(1-p) = 15(0,15)(1-0,15) = 1,9125$$

Estadísticas y probabilidades, con sus diferentes diagramaciones como: diagrama de barras, diagrama de línea, y diagrama de círculos que se aplican de acuerdo al tipo de estadísticas y probabilidades matemáticas.

Aplicaciones

Dos aplicaciones principales de la teoría de la probabilidad en el día a día son en el análisis de riesgo y en el comercio de los mercados de materias primas. Los gobiernos normalmente aplican métodos probabilísticos en regulación ambiental donde se les llama "análisis de vías de dispersión", y a menudo miden el bienestar usando métodos que son estocásticos por naturaleza, y escogen qué proyectos emprender basándose en análisis estadísticos de su probable efecto en la población como un conjunto. No es correcto decir que la estadística está incluida en el propio modelado, ya que típicamente los análisis de riesgo son para una única vez y por lo tanto requieren más modelos de probabilidad fundamentales, por ej. "la probabilidad de otro 11-S". Una ley de números pequeños tiende a aplicarse a todas aquellas elecciones y percepciones del efecto de estas elecciones, lo que hace de las medidas probabilísticas un tema político.

Un buen ejemplo es el efecto de la probabilidad percibida de cualquier conflicto generalizado sobre los precios del petróleo en Oriente Medio - que producen un efecto dominó en la economía en conjunto. Un cálculo por un mercado de materias primas en que la guerra es más probable en contra de menos probable probablemente envía los precios hacia arriba o hacia abajo e indica a otros comerciantes esa opinión. Por consiguiente, las probabilidades no se calculan independientemente y tampoco son necesariamente muy racionales. La teoría de las finanzas conductuales surgió para describir el efecto de este pensamiento de grupo en el precio, en la política, y en la paz y en los conflictos.

Se puede decir razonablemente que el descubrimiento de métodos rigurosos para calcular y combinar los cálculos de probabilidad ha tenido un profundo efecto en la sociedad moderna. Por consiguiente, puede ser de alguna importancia para la mayoría de los ciudadanos entender cómo se calculan los pronósticos y las probabilidades, y cómo contribuyen a la reputación y a las decisiones, especialmente en una democracia.

Otra aplicación significativa de la teoría de la probabilidad en el día a día es en la fiabilidad. Muchos bienes de consumo, como los automóviles y la electrónica de consumo, utilizan la teoría de la fiabilidad en el diseño del producto para reducir la probabilidad de avería. La probabilidad de avería también está estrechamente relacionada con la garantía del producto.

Se puede decir que no existe una cosa llamada probabilidad. También se puede decir que la probabilidad es la medida de nuestro grado de incertidumbre, o esto es, el grado de nuestra ignorancia dada una situación. Por consiguiente, puede haber una probabilidad de 1 entre 52 de que la primera carta en un baraja sea la *J* de diamantes. Sin embargo, si uno mira la primera carta y la reemplaza, entonces la probabilidad es o bien 100% ó 0%, y la elección correcta puede ser hecha con precisión por el que ve la carta. La física moderna proporciona ejemplos importantes de situaciones determinísticas donde sólo la descripción probabilística es factible debido a información incompleta y la complejidad de un sistema así como ejemplos de fenómenos realmente aleatorios.

En un universo determinista, basado en los conceptos newtonianos, no hay probabilidad si se conocen todas las condiciones. En el caso de una ruleta, si la fuerza de la mano y el periodo de esta fuerza es conocido, entonces el número donde la bola parará será seguro. Naturalmente, esto también supone el conocimiento de la inercia y la fricción de la ruleta, el peso, lisura y redondez de la bola, las variaciones en la velocidad de la mano durante el movimiento y así sucesivamente. Una descripción probabilística puede entonces ser más práctica que la mecánica newtoniana para analizar el modelo de las salidas de lanzamientos repetidos de la ruleta. Los físicos se encuentran con la misma situación en la teoría cinética de los gases, donde el sistema determinístico *en principio*, es tan complejo (con el número de moléculas típicamente del orden de magnitud de la constante de Avogadro $6 \cdot 10^{23}$) que sólo la descripción estadística de sus propiedades es viable.

La mecánica cuántica, debido al principio de indeterminación de Heisenberg, sólo puede ser descrita actualmente a través de distribuciones de probabilidad, lo que le da una gran importancia a las descripciones probabilísticas. Algunos científicos hablan de la expulsión del paraíso.^[*cita requerida*] Otros no se conforman con la pérdida del determinismo. Albert Einstein comentó estupendamente en una carta a Max Born: *Jedenfalls bin ich überzeugt, daß der Alte nicht würfelt. (Estoy convencido de que Dios no tira el dado)*. No obstante hoy en día no existe un medio mejor para describir la física cuántica si no es a través de la teoría de la probabilidad. Mucha gente hoy en día confunde el hecho de que la mecánica cuántica se describe a través de distribuciones de probabilidad con la suposición de que es por ello un proceso aleatorio, cuando la mecánica cuántica es probabilística no por el hecho de que siga procesos aleatorios sino por el hecho de no poder determinar con precisión sus parámetros fundamentales, lo que imposibilita la creación de un sistema de ecuaciones determinista.

Investigación biomédica


La mayoría de las investigaciones biomédicas utilizan muestras de probabilidad, es decir, aquellas que el investigador pueda especificar la probabilidad de cualquier elemento en la población que investiga. Las muestras de probabilidad permiten usar estadísticas inferenciales, aquellas que permiten hacer inferencias a partir de datos. Por otra parte, las muestras no probabilísticas solo permiten usarse estadísticas descriptivas, aquellas que solo permiten describir, organizar y resumir datos. Se utilizan cuatro tipos de muestras probabilísticas: muestras aleatorias simples, muestras aleatorias estratificadas, muestra por conglomerados y muestras sistemáticas.

Referencias

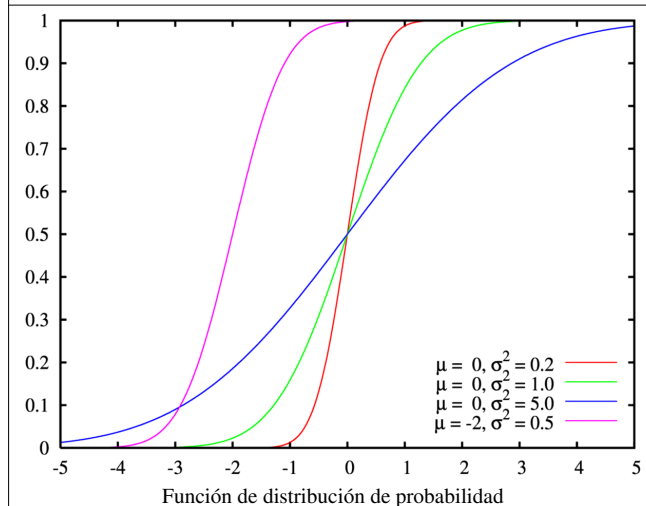
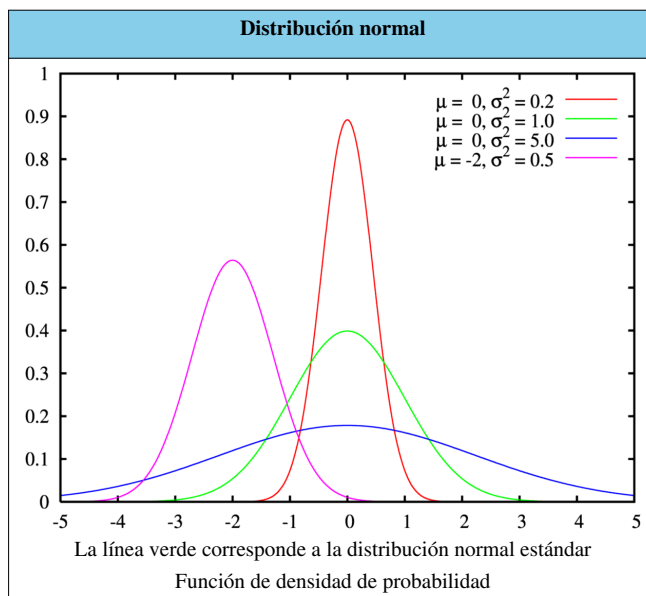
[1] Jeffrey, R.C., *Probability and the Art of Judgment*, Cambridge University Press. (1992). pp. 54-55. ISBN 0-521-39459-7

Enlaces externos

Wikilibros

-  Wikilibros alberga un libro o manual sobre **Probabilidades**.
- Edwin Thompson Jaynes. *Probability Theory: The Logic of Science*. Preprint: Washington University, (1996). — HTML (<http://omega.albany.edu:8008/JaynesBook.html>) y PDF (<http://bayes.wustl.edu/etj/prob/book.pdf>) (en inglés)
- Un problema de probabilidades: (<http://matematicasyfutbol.blogspot.com.es/2013/08/adinovos-quien-ganara-la-champions.html>)

Distribución normal



Parámetros	$\mu \in \mathbb{R} \quad \sigma > 0$
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
Coefficiente 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 aproximada en fenómenos reales.

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

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.

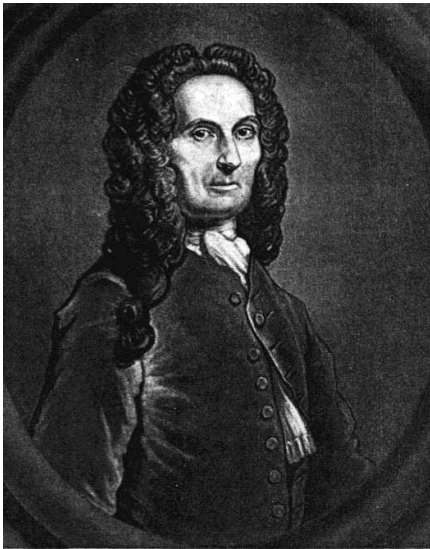
Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son:

- 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 reimpreso 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. 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

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.^{[4][5]} 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 .

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.

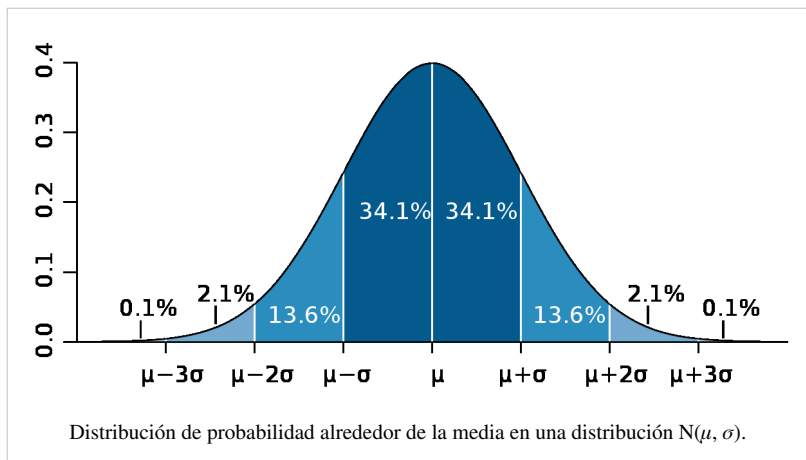
Para una distribución normal, la función característica es

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

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.
 - La divergencia de Kullback-Leibler, $DKL(X||Y) = \frac{1}{2} \left(\log \left(\frac{\sigma_Y^2}{\sigma_X^2} \right) + \frac{\sigma_X^2}{\sigma_Y^2} + \frac{(\mu_Y - \mu_X)^2}{\sigma_Y^2} - 1 \right)$.
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 $\mu = \sigma^2 = \lambda$.

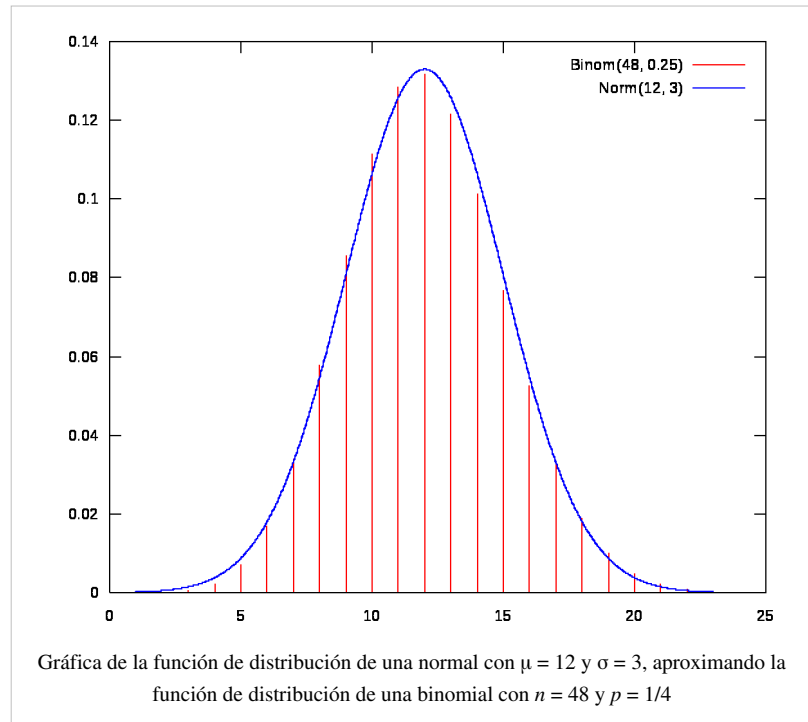
La exactitud de estas aproximaciones depende del propósito para el que se necesiten y de la tasa de convergencia a la distribución normal. Se da el caso típico de que tales aproximaciones son menos precisas en las colas de la distribución. El Teorema de Berry-Esséen proporciona un límite superior general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

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 los 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 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 suposición natural lleva a 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 en torno 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.

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.

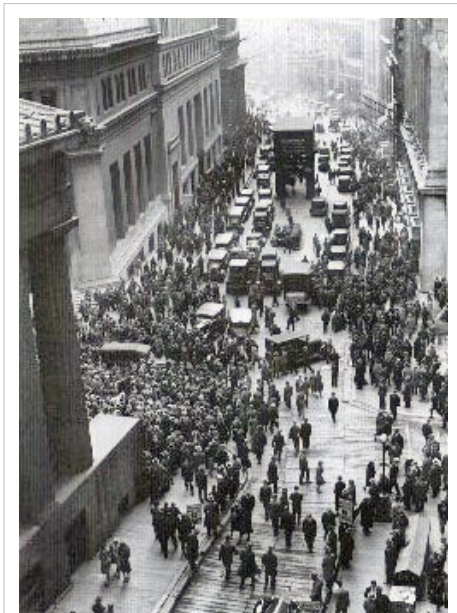
La asunción de que el tamaño lineal de los especímenes biológicos es normal (más que lognormal) nos lleva a una distribución no normal del peso (puesto que el peso o el volumen es proporcional al cuadrado o el cubo de la longitud y las distribuciones gaussianas sólo mantienen las transformaciones lineales). 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 isotropa) 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 sùmense 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.^[6]

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 [7]):

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

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. 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.

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] La función Q (<http://cnx.org/content/m11537/latest/>)
- [5] <http://www.eng.tau.ac.il/~jo/academic/Q.pdf>
- [6] Johnson NL, Kotz S, Balakrishnan N. (1995) *Continuous Univariate Distributions Volume 2*, Wiley. Equation(26.48)
- [7] http://www.math.sfu.ca/~cbm/aands/page_932.htm

Enlaces externos

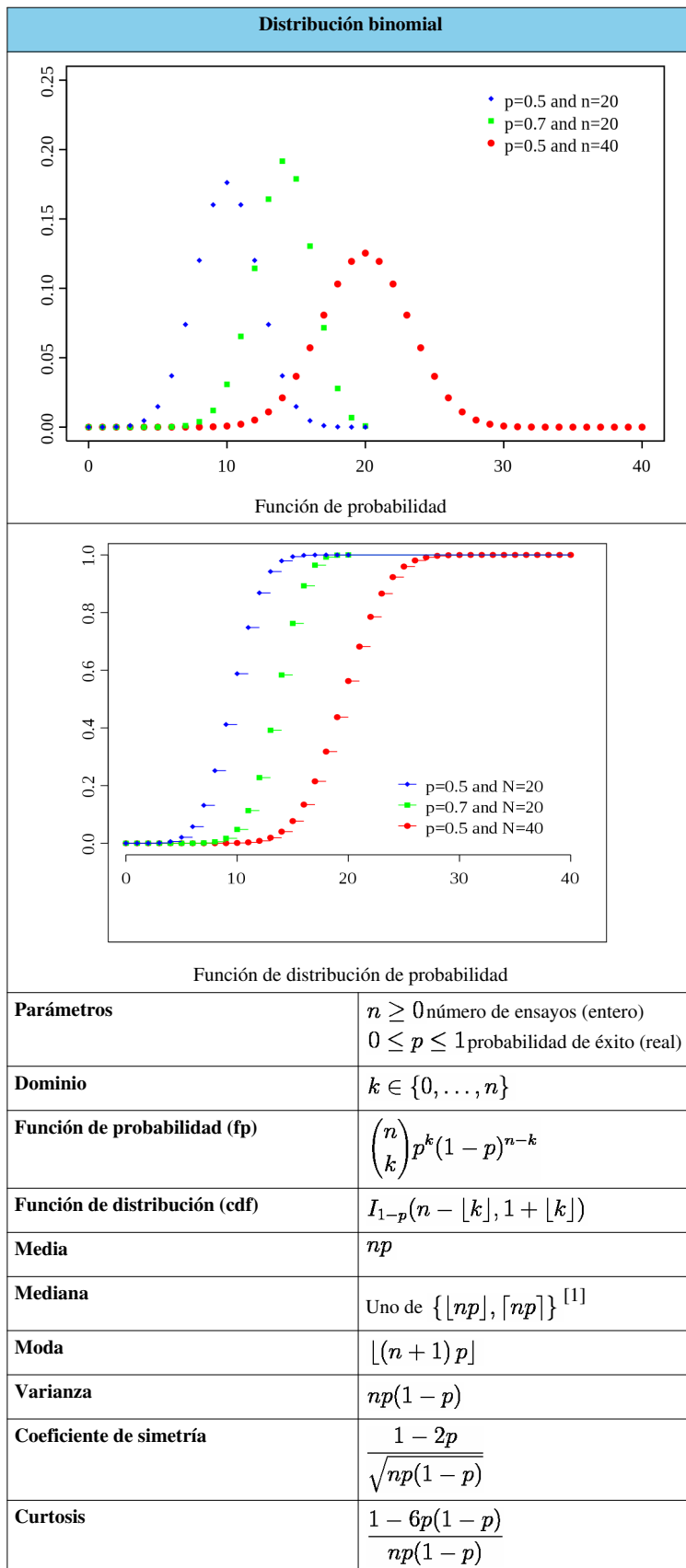
- Áreas bajo la curva normal (<http://www.digitalreview.com.ar/distribucionnormal/>) Tabla conteniendo los valores de la función normal
- Calculadora de probabilidades en una distribución Normal (<http://www.ugr.es/~jsalinas/normal.htm>). Permite hacer cálculos directos e inversos.
- Demostración de la distribución normal (<http://www.foro.resuelveproblemas.com/Matematicas-La-distribuci3n-normal>)
- Tabla de la distribución normal (http://www.vaxasoftware.com/doc_edu/mat/dnormal.pdf) Tabla de la distribución normal en formato PDF

Ajuste por software de una distribución de probabilidad a un conjunto 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.elektro-energetika.cz/calculations/no.php?language=espanol>)
- Calcular la probabilidad de una distribución normal (<http://cajael.com/mestadisticos/T7DContinuas/node3.php>) con R (lenguaje de programación)
- Implementación Real en C del algoritmo Expectation Maximization (<https://github.com/juandavm/em4gmm>) (EM) para estimar Gaussian Mixture Models (<https://github.com/juandavm/em4gmm>) (GMMs).

Distribución binomial

00



Entropía	$\frac{1}{2} \ln(2\pi n e p(1-p)) + O\left(\frac{1}{n}\right)$
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 de Bernoulli independientes entre sí, 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 tres obtenidos: entonces $X \sim B(10, 1/6)$
- Se lanza una moneda dos veces y se cuenta el número X de caras obtenidas: entonces $X \sim B(2, 1/2)$

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 (con 6 caras) 50 veces y queremos conocer 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

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

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

Relaciones con otras variables aleatorias

Si n tiende a infinito y p es tal que el 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 $p=0.5$ y 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

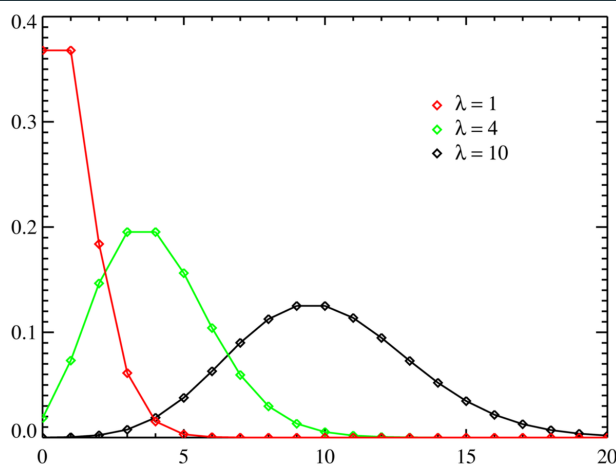
- [1] Hamza, K. (1995). The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions. *Statist. Probab. Lett.* 23 21–25.

Enlaces externos

- Calculadora Distribución binomial (<http://www.elektro-energetika.cz/calculations/bi.php?language=espanol>)
- (<http://cajael.com/mestadisticos/T6DDiscretas/node2.php>) Cálculo de la probabilidad de una distribución binomial con R (lenguaje de programación)

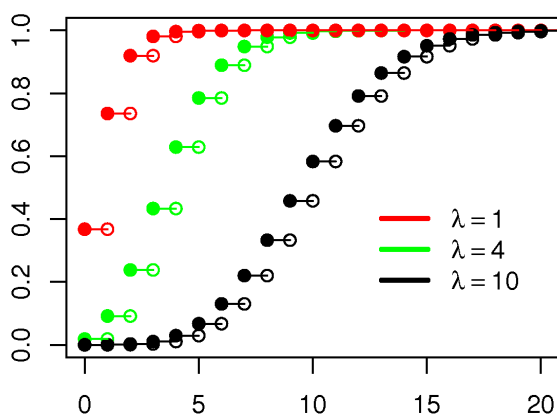
Distribución de Poisson

Distribución de Poisson



El eje horizontal es el índice k . La función solamente está definida en valores enteros de k . Las líneas que conectan los puntos son solo guías para el ojo y no indican continuidad.

Función de probabilidad



El eje horizontal es el índice k .
Función de distribución de probabilidad

Parámetros	$\lambda \in (0, \infty)$
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	$\lceil \lambda \rceil - 1$
Varianza	λ
Coficiente 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 o densidad 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

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

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

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

Intervalo de confianza

Un criterio fácil y rápido para calcular un intervalo de confianza aproximada de λ es propuesto por Guerriero (2012). Dada una serie de eventos k (al menos el 15 - 20) en un periodo de tiempo T , los límites del intervalo de confianza para la frecuencia vienen dadas por:

$$F_{low} = \left(1 - \frac{1.96}{\sqrt{k-1}}\right) \frac{k}{T}$$

$$F_{upp} = \left(1 + \frac{1.96}{\sqrt{k-1}}\right) \frac{k}{T}$$

entonces los límites del parámetro λ están dadas por: $\lambda_{low} = F_{low}T$; $\lambda_{upp} = F_{upp}T$.

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.

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

References

- [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.elektro-energetika.cz/calculations/po.php?language=espanol>
- [5] <http://cajael.com/mestadisticos/T6DDiscretas/node7.php>

Proceso de Poisson

En estadística y simulación un **Proceso de Poisson** (también conocido como "**Ley de los sucesos raros**") llamado así por el matemático Siméon Denis Poisson (1781–1840) es un proceso estocástico de tiempo continuo que consiste en "contar" eventos raros (de ahí el nombre "ley de los eventos raros") que ocurren a lo largo del tiempo.

Definición.

Un proceso Poisson con intensidad (o tasa) $\lambda \geq 0$ es un proceso de contar en tiempo continuo $\{N_t, t \geq 0\}$, donde N_t es una colección de variables aleatorias con las siguientes propiedades:

1. $N(0) = 0$.
2. Si $s \leq t$ entonces $N_s \leq N_t$.
3. Para todo $n > 0$ y $0 < t_1 < t_2 < \dots < t_n$, las variables aleatorias $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$, son independientes
4. Para toda $h > 0$ y $t \in \mathbb{R}^+$, N_h y $N_{t-h} - N_t$ tienen la misma distribución.
5. $P\{N(h) = 1\} = \lambda h + o(h)$.
6. $P\{N(h) \geq 2\} = o(h)$.

Donde $o(h)$ es una función tal que:

$$\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$$

Interpretación intuitiva

N_t es el número de eventos que se han producido desde el instante cero hasta el instante t . Como en cualquier proceso estocástico, en el instante cero es una variable aleatoria; pero, después del instante t es un dato.

Propiedades

A partir de la definición es posible demostrar que:

- Las variables aleatorias N_t tienen distribución Poisson con parámetro λt
- Si T_k denota el tiempo transcurrido desde el (k-1)-ésimo evento hasta el k-ésimo, entonces T_k es una variable aleatoria con distribución exponencial y parámetro λ
- Si S_n denota el tiempo transcurrido desde el inicio del conteo hasta el n-ésimo evento, entonces S_n tiene distribución Gamma con parámetros (n, λ)

Aplicación en seguros

Una importante aplicación del proceso Poisson se encuentra en la probabilidad de ruina de una compañía aseguradora. El problema fue tratado formalmente por Filip Lundberg en su tesis doctoral en 1903. Posteriormente Crámer desarrolla las ideas de Lundberg y da lugar a lo que hoy se conoce como el Proceso de Ruina o Modelo de Crámer-Lundberg.

Procesos de Poisson no homogéneos

A menudo son más realistas los modelos basados en procesos de Poisson no homogéneos, en los que la tasa de llegadas es una función del parámetro de tiempo, $\lambda(t)$. Formalmente esto significa que un Proceso de Poisson no homogéneo es un proceso de contar que satisface:

1. $N(0) = 0$
2. Los incrementos en intervalos ajenos son independientes.
3. $P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$
4. $P(N(t+h) - N(t) > 1) = o(h)$

Los tres métodos más conocidos de generación de un proceso de Poisson no homogéneo de este tipo se basan en la modificación de la escala de tiempo, en el condicionamiento y en una adaptación del método de rechazo.

Para procesos homogéneos hay una densidad media λ . Eso significa que la media de los sucesos en un intervalo de tiempo t es λt .

El tiempo entre dos sucesos de un proceso de Poisson con intensidad media λ es una variable aleatoria de distribución exponencial con parámetro λ .

Aplicaciones

Se pueden modelar muchos fenómenos como un proceso de Poisson. El número de sucesos en un intervalo de tiempo dado es una variable aleatoria de distribución de Poisson donde λ es la media de números de sucesos en este intervalo. El tiempo hasta que ocurre el suceso número k en un Proceso de Poisson de intensidad λ es una variable aleatoria con distribución gamma o (lo mismo) con distribución de Erlang con $\theta = 1/\lambda$.

Otras aplicaciones:

- La cantidad de clientes que entran a una tienda.
- El número de coches que pasan por una autopista.
- La llegada de personas a una fila de espera.
- El número de llamadas que llegan a una central telefónica.
- Partículas emitidas por un material radiactivo.

Regresión de Poisson

En estadística, la **regresión de Poisson** es un tipo de modelo lineal generalizado en el que la variable de respuesta tiene una distribución de Poisson y el logaritmo de su valor esperado puede ser modelado por una combinación lineal de parámetros desconocidos, es decir, el logaritmo es la función de enlace canónica. Se usa para modelar datos de conteo (número de veces que ocurre cierto fenómeno aleatorio) y tablas de contingencia.

Formulación matemática

La regresión de Poisson se utiliza para modelar fenómenos que pueden representarse mediante una variable aleatoria Y tal que para un valor $x \in \mathbb{R}^n$ de unas variables independientes,

$$Y_{|x} \sim \text{Poisson}(\exp(a'x + b)),$$

es decir, el valor de Y condicionado a x sigue una distribución de Poisson de parámetro $\exp(a'x + b)$ para ciertos valores $a \in \mathbb{R}^n$ y $b \in \mathbb{R}$.

En concreto, debido a las propiedades de la distribución de Poisson, el valor de la media predicha es

$$\log(\mathbb{E}(Y|x)) = a'x + b.$$

A veces, por abreviar, se escribe simplemente

$$Y_{|x} \sim \text{Poisson}(\exp(\theta'x)),$$

donde x es un vector $n+1$ -dimensional que consta de n variables independientes y una constante, usualmente 1. En este caso concreto, θ es simplemente a concatenado a b .

Si Y_i son observaciones independientes de la variable aleatoria Y , la estimación de θ suele realizarse utilizando el método de la máxima verosimilitud. Este estimador no admite una forma cerrada y debe calcularse mediante métodos numéricos. Como la superficie de probabilidad para este tipo de modelos es siempre convexa, el método de Newton u otros métodos basados en el gradiente son adecuados.^[cita requerida] No obstante, los paquetes estadísticos habituales son capaces de realizar automáticamente el ajuste de este tipo de modelos.

Aplicaciones

El modelo de Poisson es apropiado cuando la variable dependiente es un conteo, como por ejemplo, el número de llamadas que llegan a una central telefónica, que dependen de otras variables como, por ejemplo el día de la semana o la hora del día. Los sucesos tienen que ser independientes.

Al aplicar este tipo de modelos a datos reales, en algunos casos, se dan fenómenos tales como:

- **Sobredispersión:** Una peculiaridad de la distribución de Poisson es que su media es igual a su varianza. Sin embargo, en ciertos conjuntos de datos se observa una varianza superior a la esperada. El fenómeno se conoce como sobredispersión e indica que el modelo no es adecuado. Un motivo frecuente es la omisión de alguna variable relevante. En algunos casos se aconseja recurrir a la distribución binomial negativa.
- **Exceso de ceros:** Otro fenómeno que aparece en la práctica es el del exceso de ceros. Puede deberse a que existen dos fenómenos estadísticos que se entrecruzan: uno genera ceros; otro, los valores no nulos. Esto ocurre, por ejemplo, al tratar de modelar el número de cigarrillos fumados por cada uno de los integrantes de un grupo de personas: puede que algunos de ellos, simplemente, no sean fumadores.

Implementaciones

Implementaciones de este modelo existen en paquetes estadísticos tales como:

- SPSS, usando el comando GENLIN
- Matlab Statistics Toolbox: funciones "glmfit" y "glmval".^[1]
- Microsoft Excel: a través de extensiones tales como XPost^[2]
- SAS: función GENMOD
- Stata: procedimiento "poisson"
- R: la función glm()

Ejemplo de ajuste de un modelo de Poisson con R

El siguiente código muestra cómo ajustar mediante un modelo de regresión de Poisson un conjunto de datos recopilados por Dobson.^{[3][4]}

```
# Construcción de los datos
counts <- c(18, 17, 15, 20, 10, 20, 25, 13, 12)
outcome <- gl(3, 1, 9)
treatment <- gl(3, 3)

# Ajuste del modelo
glm.D93 <- glm(counts ~ outcome + treatment, family=poisson())

# Resumen del modelo
anova(glm.D93)
summary(glm.D93)
```

Bibliografía

- Cameron, A.C. and P.K. Trivedi (1998). *Regression analysis of count data*, Cambridge University Press. ISBN 0-521-63201-3
- Christensen, Ronald (1997). *Log-linear models and logistic regression*. Springer Texts in Statistics (Second edición). New York: Springer-Verlag. pp. xvi+483. Plantilla:MR. ISBN 0-387-98247-7.
- Hilbe, J.M. (2007). *Negative Binomial Regression*, Cambridge University Press. ISBN 978-0-521-85772-7

Referencias

- [1] glmfit (<http://www.mathworks.com/help/toolbox/stats/glmfit.html>)
- [2] http://www.indiana.edu/~jslsoc/files_research/xpost/xpost.pdf
- [3] Dobson, A. J. (1990) *An Introduction to Generalized Linear Models* London: Chapman and Hall.
- [4] Fitting Generalized Linear Models (http://stuff.mit.edu/afs/sipb/project/r-project/arch/i386_rhel3/lib/R/library/stats/html/glm.html), página de ayuda de la función glm() de R

Fuentes y contribuyentes del artículo

Probabilidad Fuente: <http://es.wikipedia.org/w/index.php?oldid=70397229> Contribuyentes: Acratta, Aldo93, Alexav8, Alvaro qc, Amanuense, Andreasmeru, Angel GN, Angelito7, Antonorsi, AqueronteBlog, Asddas, Ast Derek, Açipni-Lovrij, BlackBeast, Camilo, Capohacker13, Charly montoya, Cookie, Crashjd, Dangelin5, David0811, Denayamar, Dermot, Diegusjaimes, Dreitmen, Eduardosalg, Einundswanzig, El Sinchi, Farisori, Flakinho, GermanX, Gustronico, Götz, HUB, Harpagornis, Helmy oved, Hemerson p, Hlino, IrwinSantos, Isha, J.delanoy, Javierito92, Jkbw, JorgeGG, Juniormpalacios, Karshan, Laura Bauer, Leonpolanco, LlamaAl, Luis1970, Mafores, Magister Mathematicae, Makaka33, Manuelt15, MarcoAurelio, Mariana de El Mondongo, Matdroses, MotherForker, Mperort348, Napier, Newton-200, Nicop, OboeCrack, Paintman, Petronas, Petruss, Pino, Poco a poco, Profeshermiguad, Pólux, Quijav, Ralgis, Raulshc, RedTony, Rjgalindo, Rosarino, Rubpe19, Savh, Sebrev, Semontanés, Snakeeater, Splash04, Spyglass007, SuperBraulio13, Technopat, Thingg, Travelour, UA31, Ugly, Valentin esteveznavarro, VanKleinen, Vic Fede, Vitamine, Waka Waka, Wilfredor, 524 ediciones anónimas

Distribución normal Fuente: <http://es.wikipedia.org/w/index.php?oldid=70611614> Contribuyentes: A ver, Acratta, Af3, Airunp, Alexv86, AlfonsoERomero, Andreasmeru, Antur, Apendata, AstroNomo, Augustomarijuan, BImbo, Banfield, BlackBeast, BuenaGente, Carlos.Gracia-Lázaro, Cgb, Chesnok, Christianfgc, ConPermiso, DanielithoMoya, Davius, Dhcp, Diegusjaimes, Dodo, Doloco, Edmenb, Eduardosalg, Er Komandante, Euratom, Farisori, Fsd141, Gemini1980, Germanrinconrey, Ggenellina, Gperjim, Guanucoluis, Gökhan, HiTe, Humbefa, Jarisleif, Jerowiki, Jkbw, JoeLoui, Jorge c2010, JorgeGG, JoseA, Joseaperez, Joxemai, Juan Carlos Tapia, Juan Manuel, Juan Mayordomo, LP, Leonpolanco, Marcelo, Marianov, Marsal20, Matdroses, Miss Manzana, Moonkey, Niksfish, Omary22 24, Oscar ., Palissy, Pasmargo, Paulrc, Quien sabe puede, Rafa3040, René Vápeník, Ricardogpn, Roche, Rubpe19, Rufflos, SPZ, Savh, Sergio Andres Segovia, Srbanana, Taichi, Tano4595, Tartaglia, Technopat, Thebiguserpratwiki, Tirithel, Tomatejc, Travelour, Vivero, Xenoforme, 207 ediciones anónimas

Distribución binomial Fuente: <http://es.wikipedia.org/w/index.php?oldid=70615719> Contribuyentes: Sergio, Acratta, Akma72, Alex economist, Amanuense, Babbage, Bentzia, Camilo, Cgb, Danielyapahl, Darizabalo, Diegusjaimes, Dreitmen, Farisori, Fvmeteo, GermanX, Grillitus, JAGT, Jerowiki, Jkbw, Joffrey tgn, Juan Mayordomo, Juan carvacho, Juanjo Bazan, Juliwolfgang, Kved, Magister Mathematicae, Mahadeva, Marianov, Marsal20, Matdroses, Mpeinadopa, Murphy era un optimista, Paintman, Porao, Pólux, Raulshc, Ricardogpn, Soteke, Stm17, Supercyberedgar, Tano4595, Tartaglia, Tostadora, UA31, Vaskop, Walterotta, Yogobah, 151 ediciones anónimas

Distribución de Poisson Fuente: <http://es.wikipedia.org/w/index.php?oldid=70114900> Contribuyentes: Aldo david, Alex economist, Amanuense, Camilo, Cgb, Ciberrojopower, Diegusjaimes, Faelomx, Flakinho, Hiperfelix, Ictlogist, JAGT, JakobVoss, Jkbw, Juan Mayordomo, Julian Colina, Juliwolfgang, Kved, Laura Fiorucci, Magister Mathematicae, Megazilla77, Mrzeon, Paintman, Pieter, Pybalo, René Vápeník, Rufflos, Sadyssamar, SuperBraulio13, Tano4595, 147 ediciones anónimas

Proceso de Poisson Fuente: <http://es.wikipedia.org/w/index.php?oldid=64495754> Contribuyentes: Abece, CayoMarcio, Francisco.uribe, GermanX, JakobVoss, Juan Mayordomo, Julian Colina, Manuel Valadez Sánchez, Ruiz, Sabbut, Technopat, Ungoliant, 19 ediciones anónimas

Regresión de Poisson Fuente: <http://es.wikipedia.org/w/index.php?oldid=67454176> Contribuyentes: Cgb, Echani, Juan Mayordomo, Miguel A. Ortiz Arjona, Ronald2308, Rortiz, UAWiki, 2 ediciones anónimas

Fuentes de imagen, Licencias y contribuyentes

Archivo:Wikibooks-logo.svg *Fuente:* <http://es.wikipedia.org/w/index.php?title=Archivo:Wikibooks-logo.svg> *Licencia:* logo *Contribuyentes:* User:Bastique, User:Ramac et al.

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: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, Saippuakauppias, 竹麦魚(Searobin)

Archivo:Standard deviation diagram micro.svg *Fuente:* http://es.wikipedia.org/w/index.php?title=Archivo:Standard_deviation_diagram_micro.svg *Licencia:* Creative Commons Attribution-Sharealike 3.0 *Contribuyentes:* Ainali

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, Skeezi1000, Soerfm, Spuk968, Yerp0, 7 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 Pd Bailey at en.wikipedia

Licencia

Creative Commons Attribution-Share Alike 3.0
[//creativecommons.org/licenses/by-sa/3.0/](https://creativecommons.org/licenses/by-sa/3.0/)
