
La evolución de la distribución normal

The Evolution of the Normal Distribution

Autor: Saul Stahl^a
stahl@math.ku.edu

Traducido por: Luis Másmela^b
lamasmela@udistrital.edu.co

Traducido por: William Rincón^c
williamrincon@usantotomas.edu.co

Resumen

La estadística es la más extensa de todas las aplicaciones de la matemática y en el centro de la estadística se encuentra la distribución normal, conocida por millones de personas como la curva en forma de campana o la curva acampanada. En este artículo se expone la evolución que la más importante distribución de probabilidad ha sufrido en varios siglos.

Palabras clave: distribución normal, probabilidad.

Abstract

Statistics is the most widely applied of all mathematical disciplines and at the center of statistics lies the normal distribution, known to millions of people as the bell curve, or the bell-shaped curve. In this paper, the evolution of the most important probability distribution during several centuries is presented.

Key words: normal distribution, probability.

1. Introducción

La estadística es la más extensa de todas las aplicaciones de la matemática y en el centro de la estadística se encuentra la distribución normal, conocida por millones de personas como la curva en forma de campana o la curva acampanada. Esta es realmente una familia biparamétrica de curvas que se obtienen a través de la ecuación

^aDepartment of Mathematics. University of Kansas.

^bProfesor. Universidad Distrital.

^cProfesor. Universidad Santo Tomás.

$$y = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

Varias de estas curvas aparecen en la Figura 1. No solamente es conocida como la curva acampanada para muchos, también son conocidos sus principales usos como la descripción general de formas de gráficos de datos. Esta distribución, claro esta, tiene otros usos y juega un rol importante en las ciencias sociales como también en las ciencias naturales. Como es el caso de muchos conceptos matemáticos importantes, el origen de esta curva logra ser una historia que es a la vez instructiva y divertida. Realmente, se dan dos historias aquí: la invención de la curva como una herramienta para calcular probabilidades y el reconocimiento de su utilidad en la descripción de conjuntos de datos.

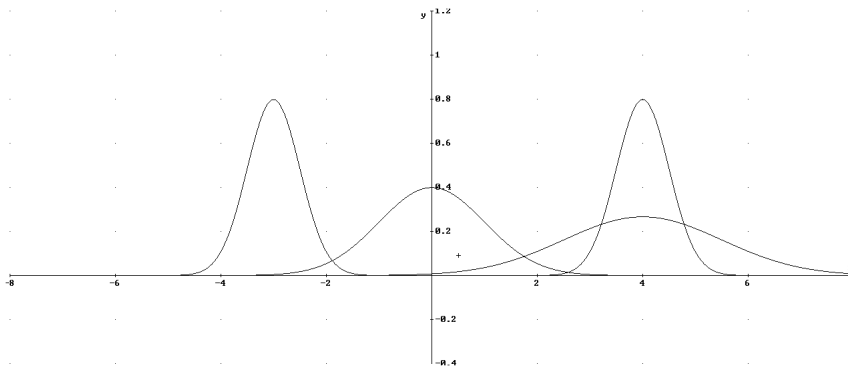


Figura 1: Curvas acampanadas.

2. Una herramienta para la aproximación.

Los orígenes de la teoría matemática de la probabilidad son justamente atribuidos a la famosa correspondencia entre Fermat y Pascal, la cual fue motivada en 1654 por las preguntas del apostador Chevalier de Méré. Entre los varios tipos de problemas ellos consideraron la distribución binomial, que en la actualidad es descrita por sumas del tipo

$$\sum_{k=i}^j \binom{n}{k} p^k (1-p)^{n-k}. \quad (2)$$

Estas sumas denotan la probabilidad de que se den entre i y j éxitos en n ensayos con probabilidad de éxito p . Tal ensayo - llamado ensayo de Bernoulli - es el más elemental de los experimentos aleatorios. Éste tiene dos resultados posibles, usualmente denominados éxito y fracaso. El k -ésimo término en (2) es la probabilidad que k de los n ensayos sean éxitos.

Sobre los ejemplos binomiales Fermat y Pascal trabajaron con pequeños valores de n , ellos no trataron con el desafío que presenta el cálculo para la evaluación de sumas generales de este tipo. Sin embargo, cálculos más complicados no demoraron en llegar.

Por ejemplo, en 1712, el matemático alemán Gravesande probó la hipótesis de que nacimientos masculinos y femeninos son igualmente probables, en contra de la información de la época en Londres, correspondiente a los 82 años que abarca el periodo de 1629 a 1710. Él notó que el número relativo de nacimientos masculinos varía de $765/15.448 = 0.5027$ en 1703 a $4748/8855 = 0.5362$ en 1661. Gravesande multiplica estos cocientes por 11429, el número promedio de nacimientos que estos 82 años abarca, obteniendo los límites nominales de 5745 a 6128 sobre el número de nacimientos masculinos en cada año. Por consiguiente, la probabilidad del exceso de nacimientos masculinos observados debido a la aleatoriedad es de solamente la potencia 82 de

$$P\left(5745 \leq x \leq 6182 \mid p = \frac{1}{2}\right) = \sum_{x=5745}^{6182} \binom{11429}{x} \left(\frac{1}{2}\right)^{11429} \\ \approx \frac{3849150}{13196800} \approx 0.292$$

(Hald explica los detalles de esta aproximación racional) Gravesande hace uso de la fórmula recursiva

$$\binom{n}{x+1} = \binom{n}{x} \frac{n-x}{x+1}$$

Sugerida por Newton para similares propósitos. Pero aún así, la de Gravesande es claramente una onerosa tarea, ya que la probabilidad de esta diferencia en tasas de nacimientos basadas en 82 años es la extremada pequeña cifra de 0.292^{82} . Gravesande concluye que la mayor tasa de nacimientos masculinos se debe a la intervención divina.

Pocos años después Jacob Bernoulli encuentra estimadores para las sumas binomiales del tipo (2). Esos estimadores, sin embargo, no utilizaban la función exponencial e^x .

De Moivre comienza su búsqueda para tales aproximaciones en 1721. En 1733, él prueba que

$$\binom{n}{\frac{n}{2} + d} \left(\frac{1}{2}\right)^n \approx \frac{2}{\sqrt{2\pi n}} e^{-2d^2/n} \quad (3)$$

y

$$\sum_{|x-n/2| \leq d} \binom{n}{x} \left(\frac{1}{2}\right)^n \approx \frac{4}{\sqrt{2\pi}} \int_0^{d/\sqrt{n}} e^{-2y^2} dy. \quad (4)$$

De Moivre también afirma que (4) puede ser generalizada en un contexto similar asimétrico, con x variando de $n/2$ a $d + n/2$. Esto se obtiene fácilmente, con la precisión de la aproximación buscada por Moivre en su prueba.

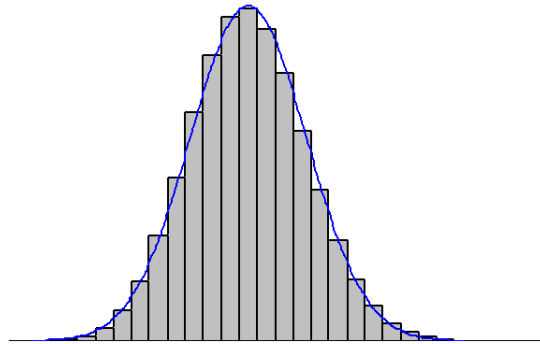


Figura 2: Una aproximación desde la probabilidad binomial.

La Figura 2 muestra cómo las probabilidades binomiales asociadas con 50 repeticiones de un ensayo de Bernoulli con probabilidad $p = 0.3$ de éxito son aproximadas por una curva exponencial. El descubrimiento de De Moivre es un acontecimiento modelo en todos los cursos introductorios de estadística, que es llamada aproximación de la binomial por la normal, y es representado como:

$$\sum_i^j \binom{p}{k} p^k (1-p)^{n-k} \approx N\left(\frac{j - np}{\sqrt{np(1-p)}}\right) - N\left(\frac{i - np}{\sqrt{np(1-p)}}\right)$$

donde

$$N(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx.$$

Ya que esta integral es fácilmente evaluada por métodos numéricos, y descrita en tablas, esto proporciona efectivamente una aproximación muy práctica para probabilidades binomiales acumuladas.

3. La búsqueda para una curva de errores

La astronomía fue la primera ciencia llamada a calcular medidas exactas. Por consiguiente, esta fue también la primera ciencia en preocuparse por la medida del error y por encarar la pregunta de cómo proceder en la presencia de varias observaciones diferentes de la misma cantidad. En el siglo II D.C., Hipparcus parece favorecer la línea intermedia; Ptolomeo, en el siglo II A.C., cuando enfrenta algunas discrepancias en cuanto a la duración de un año, decide trabajar con la observación que mejor se ajuste a su teoría. Hacia finales del siglo XVI, Ticho Brahe incorpora la repetición de medidas en la metodología de la astronomía. Curiosamente, él fallo al especificar cómo esas repetidas observaciones deben convertirse en un único número. Consecuentemente, los astrónomos crean su propio, a

menudo *ad hoc*, método para extraer una media, o “dato representativo” debido a sus observaciones. Algunas veces ellos promedian en algunas ocasiones usan la mediana, en otras agrupan sus datos y recurren a ambos, promedios y medianas. Hay veces en las que explican sus procedimientos, pero frecuentemente no lo hacen. Hay que considerar, por ejemplo, el siguiente pasaje, que proviene de Kepler y reporta observaciones hechas, en realidad, por el mismo Brahe:

En 1600 Enero 13/23 a las $11^h 50^m$ la ascensión exacta de Marte fue:

	o	'	”
usando el brillante paso de Geminis	134	23	39
usando cor Leonis	134	27	37
usando Pollus	134	23	18
a las $12^h 17^m$, usando la tercera en el ala de Virgo	134	29	48

	o	'	”
La media, tratando las observaciones imparcialmente	134	24	33

La selección de Kepler de datos representativos es desconcertante. Note que:

Promedio: $134^{\circ} 26' 5.5''$
 Mediana: $134^{\circ} 25' 38''$

y es dificultoso creer que un astrónomo, quien registra ángulos al segundo más cercano, pueda fallar al notar una discrepancia de más de un minuto. El consenso es que la media elegida puede no haber sido el resultado de un error, pero debe haber sido obtenida mediante algún cálculo. La literatura contiene al menos dos intentos para reconstruir estos cálculos, pero estos autores no encuentran ninguno convincente, ya que ambas explicaciones son *ad hoc*; aquellas no evidencian ser utilizadas en otros momentos, y ambos resultan en estimados que difieren de las de Kepler por al menos cinco segundos.

Para la latitud, ellos registran sus cálculos de datos representativos, los astrónomos de aquel tiempo parecen haber usado procedimientos improvisados que toman ambos promedios y medianas como sus componentes. La controversia entre la mediana versus el promedio dura varios siglos y ahora tiene que ser resuelta en favor de la última, particularmente en círculos científicos. Como será visto en los pasajes más adelante, esta decisión tiene una fuerte relación con la distribución normal.

El primer científico en notar y escribir que la medida de los errores es merecedora de un tratamiento sistemático y científico fue Galileo en su famoso *Dialogue Concerning the two Chief System of the World—Ptolemaic and Copernican*, publicado en 1632. Su análisis informal de las propiedades de errores aleatorios inherentes a las observaciones de fenómenos celestes es resumido por Stigler, en cinco puntos:

1. Hay solamente un número, el cual da la distancia de la estrella al centro de la tierra, la verdadera distancia.

2. Todas las observaciones están cargadas de errores, debido al observador, los instrumentos y otras condiciones observacionales.
3. Las observaciones son distribuidas simétricamente alrededor del valor verdadero; esto es, los errores están distribuidos simétricamente alrededor de cero.
4. Errores pequeños ocurren más frecuentemente que errores grandes.
5. La distancia calculada es una función de las observaciones angulares directas, tal que, pequeños ajustes de las observaciones pueden resultar en un largo ajuste de la distancia.

Desafortunadamente, Galileo no direccionó la cuestión de cómo se estima la verdadera distancia. Él, sin embargo, afirmó que "...es más probable obtener observaciones con errores pequeños que con errores grandes..." Es por tanto razonable creer que el valor verdadero más probable es aquel que minimiza la suma de las desviaciones de los valores observados (Que Galileo creó en la sencilla suma de las desviaciones soportado por sus cálculos en pp.307-308 del Dialogo.) en otras palabras, al enfrentarse con los valores observados x_1, x_2, \dots, x_n , Galileo querrá probablemente creer que el valor verdadero más probable es el valor de x que minimiza la función

$$f(x) = \sum_{i=1}^n |x - x_i| \quad (5)$$

Como se sabe, este mínimo es bien conocido como la mediana de x_1, x_2, \dots, x_n y no su promedio, un hecho que Galileo consideró probablemente interesante.

Esto se demuestra fácilmente a partir de un argumento inductivo, el cual se basa en las observaciones reindexando los valores con el propósito de que $x_1 < x_2 < \dots < x_n$, entonces

$$\sum_{i=1}^n |x - x_i| = \sum_{i=2}^{n-1} |x - x_i| + (x_n - x_1) \quad \text{si } x \in [x_1, x_n],$$

mientras que

$$\sum_{i=1}^n |x - x_i| > \sum_{i=2}^{n-1} |x - x_i| + (x_n - x_1) \quad \text{si } x \notin [x_1, x_n].$$

El concepto del promedio supone cientos de años para ser lo que universalmente es ahora y su lenta evolución es muy interesante. Alrededor de 1660, encontramos a Robert Boyle, más tarde presidente de la Royal Society, argumentando elocuentemente en contra de la idea de repetidos experimentos:

"...experimentos deben ser estimados por su valor, no por su número; ...un simple experimento... puede también ser tratado completamente... Como una de esos grandes y

orientadas perlas... puede un valor un valor muy grande de esas pequeñas... perlas, que son adquiridas por onzas...”

En un artículo que fue publicado póstumamente en 1722, Roger Cotes hace la siguiente sugerencia:

Sea p el sitio de algún objeto definido por observaciones q, r, s , luego de subsecuentes observaciones. Hay también pesos P, Q, R, S inversamente proporcionales al desplazamiento el cual puede surgir de los errores en las observaciones simples, las cuales están dadas con límites de error; y los pesos P, Q, R, S se conciben como colocados en p, q, r, s y se busca su centro de gravedad Z , el punto Z es el más probable sitio del objeto, y puede ser sin peligro el verdadero sitio.

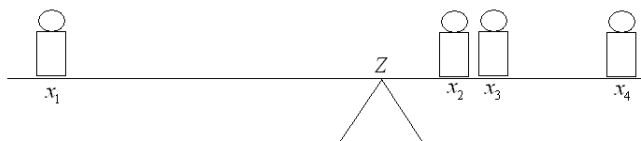


Figura 3: Una explicación balanceada del promedio.

Cotes aparentemente visualiza las observaciones como fichas x_1, x_2, \dots, x_n que son equivalentes a (p, q, r, s, \dots) con respectivos pesos físicos w_1, w_2, \dots, w_n que son equivalentes a $(P, Q, R, S; \dots)$ alineados sobre una abscisa horizontal. La Figura 3 muestra un caso donde $n = 4$ y todas las fichas tienen igual peso. Cuando Cotes hizo esto, fue natural para él sugerir que el centro de gravedad Z de este sistema representa todas las observaciones. Después de todo, también en física, para un cuerpo entero se asume que la masa se concentra en su centro de gravedad y así puede decirse que la totalidad del cuerpo está representada por un simple punto. Cotes propone centros de gravedad que concuerdan con los pesos promedio basados en la siguiente argumentación: por la definición de centro de gravedad, si el pivote es desplazado a la abscisa Z este se balanceará y por la ley de Arquímedes de palancas,

$$\sum_{i=1}^n w_i(Z - x_i) = 0$$

o

$$Z = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}. \quad (6)$$

Claro está que, cuando los pesos w_i son iguales, Z llega a ser el clásico promedio

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

Esto sugiere una temprana aparición del método de mínimos cuadrados. En este contexto, el método propone la representación de los datos x_1, x_2, \dots, x_n por el valor de x que minimiza la función

$$g(x) = \sum_{i=1}^n w_i(x - x_i)^2 \quad (7)$$

Haciendo la derivada con respecto a x es claro que Z de (6) provee este mínimo. Notese que la mediana minimiza la función $f(x)$ de (5) mientras que el promedio ponderado minimiza la función $g(x)$ de (7). Es curioso que cada uno de los dos más importantes datos representativos pueden ser identificados como los minimizadores de una no obvia, aunque completamente natural, función. Es frustrante lo poco que se conoce acerca de esta observación.

El artículo de Thomas Simpson de 1956 es de interés aquí por dos razones. Primero viene su abierto párrafo:

“Es bien conocido su señoría, que el método practicado por astrónomos, en orden a disminuir los errores aparecidos por las imperfecciones de los instrumentos, y de los órganos de los sentidos, tomando la Media de varias observaciones, tienen que no ser válida generalmente, pero algunas personas, de considerable importancia, son de la opinión, e incluso mantienen públicamente, que una observación simple, tomada con debido cuidado, es deducida como la media de un gran numero.”

Así, incluso más tarde, a mediados del siglo XVIII persistieron dudas acerca del valor de repeticiones de un experimento. Más importante, sin embargo, fue la experimentación de Simpson con *curvas de errores* que especifican densidades de probabilidad del modelo de distribución de errores aleatorios. En las dos proposiciones, Simpson calculó la probabilidad de que el error en la media de varias observaciones no exceda una cuota dada cuando los errores individuales toman valores

$$-v, \dots, -3, -2, -1, 0, 1, 2, 3, \dots, v$$

Con probabilidades que son proporcionales a cualesquiera de dos

$$r^{-v}, \dots, r^{-3}, r^{-2}, r^{-1}, r^0, r^1, r^2, r^3, \dots, r^v$$

o

$$r^{-v}, 2r^{1-v}, 3r^{2-v}, \dots, (v+1)r^0, \dots, 3r^{v-2}, 2r^{v-1}, r^v.$$

La selección de curvas de error de Simpson puede ser extraña, pero ellas son con toda seguridad dictadas por el estado del arte de la probabilidad en su tiempo. Para $r = 1$ (el caso simple), esas dos distribuciones producen las dos curvas en la parte superior de la Figura 3. Un año después, Simpson, mientras que, efectivamente crea la noción de una distribución del error continua, da con similares problemas en el contexto de curvas del error descrita en la parte inferior de la Figura 3.

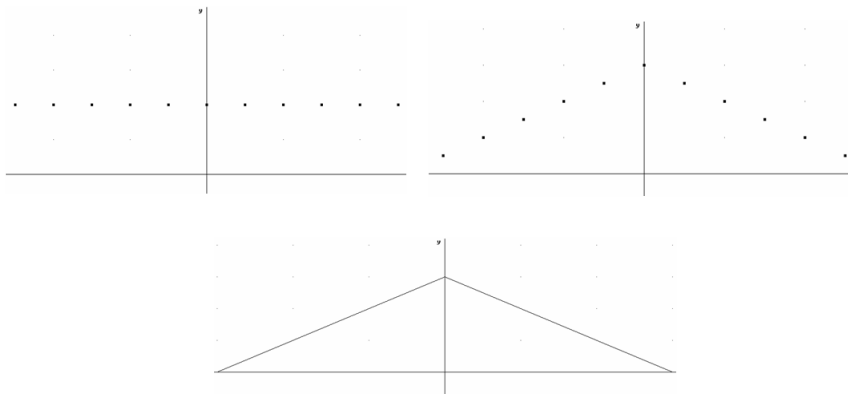


Figura 4: Curvas de error de Simpson.

En 1744, Laplace propone la primera de estas curvas de error. Denotando esta función por $\phi(x)$, él estipula que esta debe ser simétrica en x y monótona decreciente para $x > 0$. Más aún, él propone que:

“...como no tenemos razón de suponer una diferente ley para las ordenadas que para sus diferencias, de esto se sigue que, debemos, sujetos a las reglas de probabilidad, suponer el cociente de dos diferencias consecutivas infinitamente pequeñas ser igual a aquellas de las correspondientes ordenadas. De este modo tenemos: ”

$$\frac{d\phi(x+dx)}{d\phi(x)} = \frac{\phi(x+dx)}{\phi(x)}.$$

Por lo tanto $\frac{d\phi(x)}{dx} = -m\phi(x)$.

...De esta forma

$$\phi(x) = \frac{m}{2}e^{-m|x|}.$$

El argumento de Laplace puede ser parafraseado como sigue. Aparte de ser simétrica y decreciente (para $x > 0$), no conocemos nada acerca de $\phi(x)$ y $\phi'(x)$. Por tanto presumiblemente por *Occam's razor*, debe ser asumido que ellos son proporcionales, la simple suposición de igualdad a $\phi(x) = \frac{m}{2}e^{-m|x|}$, la cual es imposible. La resultante ecuación diferencial se resuelve fácilmente y la curva de error extractada es presentada en la Figura 3. Laplace no fue en alguna forma perturbado por esta curva no diferenciable en $x = 0$. Estamos a punto de ver que es perfectamente posible considerar singularidades más drásticas.

Laplace debió tener conciencia del defecto de su raciocinio, tres cortos años después él propuso una curva alternativa. Sea a el supremo de todos los posibles errores (en el contexto de un experimento específico) y sea n un entero positivo. Escoja

n puntos aleatorios dentro de un intervalo unitario, así, dividiendo entre $n + 1$ espacios. Ordenando los espacios como:

$$d_1 > d_2 > \dots > d_{n+1}, \quad d_1 + d_2 + \dots + d_{n+1} = 1.$$

Sea \bar{d}_i el valor esperado de d_i . Dibuje los puntos $(i/n, \bar{d}_i)$, $i = 1, 2, \dots, n + 1$ y hagamos n infinitamente grande. El límite de las configuraciones es una curva que es proporcional a $\ln(a/x)$ sobre $(0, a]$. La segunda curva candidata de Laplace cumple con las condiciones de simetría y del requerimiento que la probabilidad total sea 1 (Figura 3):

$$y = \frac{1}{2a} \ln \left(\frac{a}{|x|} \right) \quad -a \leq x \leq a.$$

Esta curva, con su singularidad infinita en 0 y dominio finito (un revés de las propiedades de la curva del error de la Figura 3 y la curva en forma de campana) constituye un paso hacia atrás en el proceso de evolución; sugiere la sospecha de que Laplace fue seducido por la considerable complejidad de la matemática de la derivada de la curva. Hasta el punto de que él parece llamar la atención al comentar la excesiva complejidad de la curva y sugiere que el análisis del error usando esta curva debe ser llevado solamente en *muy delicadas* investigaciones, tal como el tránsito de Venus al otro lado del sol.

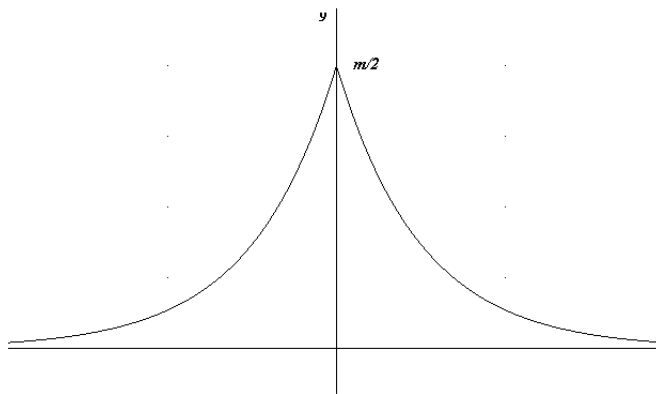


Figura 5: Primera curva de error de Laplace.

Posteriormente, en 1777, Daniel Bernoulli escribió:

“Los astrónomos son una clase de hombres de la más escrupulosa sagacidad; entonces, es a ellos a los que yo elijo para exponer las dudas que tengo algunas veces en el estudio acerca de la universal regla aceptada para tratar varias suaves discrepantes observaciones del mismo evento. Por esas reglas las observaciones son sumadas entre sí y la suma dividida por el número de observaciones: el cociente es entonces aceptado como el verdadero valor de la cantidad requerida, hasta que mejor y más información es obtenida. De esta forma, si las varias observaciones pueden ser consideradas teniendo, como es, los mismos pesos, el

centro de gravedad es aceptado como la verdadera posición del objeto bajo investigación. Estas reglas están de acuerdo con las usadas en la teoría de la probabilidad cuando todos los errores de las observaciones son considerados igualmente probables.

Pero, ¿es posible tener que las varias observaciones tiene los mismos pesos o momentos para todos y cada uno de los errores? ¿Son los errores de algún fácil grado como otros de muchos minutos? ¿Hay por donde se quiera ver la misma probabilidad? Tales afirmaciones pueden ser absurdas, la cual es indudablemente la razón por la que astrónomos prefieren rechazar completamente observaciones las cuales ellos han juzgado ser también desviada de la verdad, mientras que al conservar el resto y, efectivamente, asignando a él la misma confiabilidad. ”

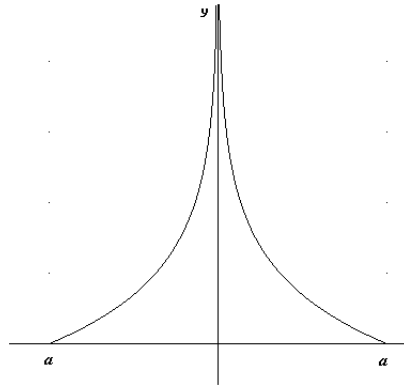


Figura 6: Segunda curva de error de Laplace.

Es interesante notar que Bernoulli admite que el hacer promedios es aceptado universalmente. Para la escurridiza curva del error, él toma esto para garantizar que se puede tener un dominio finito y explícito acerca de la tangente horizontal como el punto máximo y casi vertical próximo a los límites del dominio. Él sugiere la semielipse como tal curva, la cual, siguiendo un argumento lógico, entonces se reemplaza con un semicírculo.

El próximo desarrollo importante tiene sus raíces en un evento celeste que ocurre en enero 1 de 1801. En aquel día, el astrónomo italiano Giuseppe Piazzi divisa un cuerpo celeste, y que sospecha fuertemente es un nuevo planeta. Él anuncia su descubrimiento y lo llama Ceres. Desafortunadamente, seis semanas después, suficientes observaciones tienen que ser tomadas para hacer posible una exacta determinación de su órbita, así es como aciertan que es efectivamente un planeta; Ceres desaparece detrás del sol y no se espera que aparezca por casi un año. El interés sobre este posible nuevo planeta se extendió y astrónomos en toda Europa presentan ellos mismos, por conjetura a partir de cálculos, la localización donde Ceres es más probable que aparezca. El joven Gauss, quien es renombrado por ser un extraordinario matemático, propone un área del cielo para la búsqueda que fue totalmente diferente de aquella sugerida por otros astrónomos, y su resultado es correcto. Un artículo en el *Magazine* (1999) describe la historia en detalle.

Gauss explica que él uso el criterio de mínimos cuadrados para localizar la órbita que mejor se ajuste a las observaciones. Este criterio fue justificado por la teoría del error basada en las tres afirmaciones siguientes:

1. Pequeños errores son más probables que grandes errores.
2. Para cualquier número real ϵ las probabilidades de errores de magnitud ϵ y $-\epsilon$ son iguales.
3. En la presencia de varias medidas de la misma cantidad, el más probable valor de la cantidad medida es su promedio.

Sobre la base de esas afirmaciones él concluye que la densidad de probabilidad para el error (esto es, la curva del error) es

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

Donde h es una constante positiva que es pensada por Gauss como la "precisión del proceso de medida". Él reconoce esta como la curva de campana determinada por $\mu = 0$ y $\sigma = 1/\sqrt{2}h$.

El ingenioso desarrollo de Gauss de esta curva del error hace uso solamente de algunos argumentos básicos de probabilidad y hechos estándar de cálculo. Como este escrito puede caer en manos de un estudiante universitario de matemáticas con un curso en cálculo basado en estadística, la prueba es presentada a continuación con solo modificaciones menores.

4. La prueba

Sea p el verdadero valor (pero desconocido) de la medida de la cantidad, sean n observaciones independientes dadas por los estimadores M_1, M_2, \dots, M_n , y sea $\phi(x)$ la función de densidad de probabilidad del error aleatorio. Gauss toma esto para garantizar que la función es diferenciable. La afirmación 1 anterior implica que $\phi(x)$ tiene un máximo en $x = 0$ mientras que la afirmación 2 significa que $\phi(-x) = \phi(x)$. Si definimos

$$f(x) = \frac{\phi'(x)}{\phi(x)}$$

entonces

$$f(-x) = -f(x).$$

Notese que $M_i - p$ denota el error de la i -ésima medida y por consiguiente, ya que esas medidas (y errores) se asumen estocásticamente independiente, esto se sigue que

$$\Omega = \phi(M_1 - p)\phi(M_2 - p) \cdots \phi(M_n - p)$$

Es la densidad conjunta de los n errores. Gauss interpreta la afirmación 3 como se dice, en terminología moderna, que

$$\bar{M} = \frac{M_1 + M_2 + \cdots + M_n}{n}$$

Es el estimador máximo verosímil de p . En otras palabras, dadas las medidas M_1, M_2, \dots, M_n , escoger $p = \bar{M}$ maximiza el valor de Ω . Ya que

$$0 = \left. \frac{\partial \Omega}{\partial p} \right|_{p=\bar{M}} = - \sum_{i=1}^n \left[\phi'(M_i - \bar{M}) \prod_{j \neq i} \phi(M_j - \bar{M}) \right] = -\Omega \sum_{i=1}^n \frac{\phi'(M_i - \bar{M})}{\phi(M_i - \bar{M})}.$$

De esto se sigue que

$$f(M_1 - \bar{M}) + f(M_2 - \bar{M}) + \cdots + f(M_n - \bar{M}) = 0. \quad (8)$$

Recordando que las medidas M_i pueden asumir valores arbitrarios y en particular, si M y N son números reales arbitrarios podemos usar

$$M_1 = M, \quad M_2 = M_3 = \cdots = M_n = M - nN$$

Para tal conjunto de medidas $\bar{M} = M - (n-1)N$.

Sustituyendo en (8) producimos

$$f[(n-1)N] + (n-1)f(-N) = 0 \quad \text{o} \quad f[(n-1)N] = (n-1)f(N).$$

Es un ejercicio bien conocido que bajo esta condición de homogeneidad, combinado con la continuidad de f , implica que $f(x) = kx$ para algún real k . Esto produce la ecuación diferencial

$$\frac{\phi'(x)}{\phi(x)} = kx.$$

Integrando con respecto a x produce

$$\ln \phi(x) = \frac{k}{2}x^2 + c \quad \text{o} \quad \phi(x) = Ae^{kx^2/2}.$$

En su orden, para que $\phi(x)$ asuma un máximo en $x = 0$, k debe ser negativo y así podemos establecer $k/2 = -h^2$. Finalmente, ya que

$$\int_{-\infty}^{\infty} e^{-h^2x^2} dx = \frac{\sqrt{\pi}}{h}$$

De donde se sigue que

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

Con lo que la prueba se completa. \square

La Figura 4 muestra un histograma de algunas medidas de la ascensión corregida de Marte junto con una aproximación de una curva exponencial. Este ajuste es ciertamente sorprendente.

Se noto anteriormente que el promedio es, en efecto, un estimador de mínimos cuadrados de los datos. Ésto significa que Gauss usó una estimación particular de mínimos cuadrados para justificar su teoría de errores, la cual en cambio fue usada para justificar el criterio general de mínimos cuadrados. Hay un elemento de Bootstrapping en este razonamiento que deja más adelante estadísticos insatisfechos y puede tener un efecto similar sobre el mismo Gauss. Él retorna al tema dos veces, doce y trece años después, para explicar su curva de error por medio de diferentes cadenas de razonamientos.

Actualmente, una muy posible explicación está explícita en el Teorema del Límite Central publicada por Laplace en 1810. La traducción de Laplace parafraseada expresa:

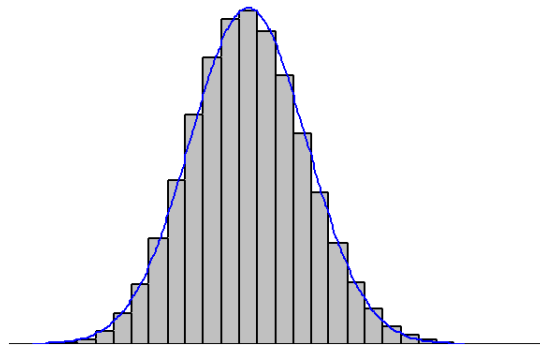


Figura 7: Medidas distribuidas normalmente.

...si es asumido que para cada observación los errores positivos y negativos son igualmente probables, la probabilidad que la media del error de n observaciones serán contenidos entre los límites $\pm rh/n$, siendo igual a

$$\frac{2}{\sqrt{\pi}} \sqrt{\frac{k}{2k'}} \int e^{-\frac{k}{2k'} r^2} dr$$

donde h es el intervalo dentro del cual los errores de cada observación pueden caer. Si la probabilidad del error $\pm x$ es designada por $\phi(x/h)$, entonces k es la integral $\int dx \cdot \phi(x/h)$

evaluada en $x = -\frac{1}{2}h$ o $x = \frac{1}{2}h$, y k' es la integral $\int \frac{x^2}{h^2} dx \cdot \phi(x/h)$ evaluada en el mismo intervalo.

Holgadamente hablando, el teorema de Laplace establece que si la curva del error de una observación simple es simétrica, entonces la curva del error de la suma de varias observaciones es efectivamente aproximada por una de las curvas Gaussianas de (1). Ya que si tomamos el paso adicional de imaginar que el error implicado en una observación adicional es el total de un gran número de “elementos” o errores “atómicos”, entonces el teorema predice que el error aleatorio que ocurre en aquella observación individual es efectivamente controlada por la curva de Moivre y Gauss (1).

Esta afirmación, promulgada por Hagen y Bessel, llega a ser conocida como *hipótesis de error elemental*. Un estudio secundario es publicado por Daniel Bernoulli en 1780, aunque con un más estrecho alcance. Asumiendo un error fijo de $\pm\alpha$ para cada oscilación de un reloj de péndulo, Bernoulli concluye que el error acumulado durante un día, será, en terminología moderna, aproximadamente distribuido de forma normal.

Este puede ser el momento de recapitular el nacimiento del promedio para el prominente disfrute del estimador selecto. Los tratamientos de Kepler de sus observaciones muestran que alrededor de 1600 hay todavía un procedimiento no estándar para resumir múltiples observaciones. Alrededor de 1600 Boyle todavía objeta la idea de *combinar* varias medidas en una sola. Medio siglo después, Cotes propone el promedio como el mejor estimador. Un artículo de Simpson de 1956 indica que los oponentes del proceso de promediar, que aparentan una minoría, tiene una calma no dada antes. El artículo de Bernoulli de 1777 admite que la costumbre de promediar tiende a hacerse universal. Finalmente, en algún momento en la primera década del siglo XIX, Gauss asume la optimalidad del promedio como un axioma para el propósito de determinar la distribución de la medida de los errores.

5. Más allá de los errores

El primer matemático que extiende la procedencia de la distribución normal más allá de la distribución de la medida de los errores fue Adolphe Quetelet (1796-1874). Él comienza su carrera como un astrónomo pero es movido por la ciencias sociales. En consecuencia, él posee una inusual combinación de cualidades que lo hacen tener la posición correcta para ser uno de los científicos más influyentes en las observaciones de todos los tiempos.

En su libro de 1846 *Cartas dirigidas a H. R. H. el gran duque de Saxe Coburg y Gotha, sobre la Teoría de las Probabilidades como aplicación a las ciencias Morales y Políticas*, Quetelet extrae el contenido de la Tabla 1 de la *Edinburgh Medical and Surgical Journal* (1817) y contiene el patrón modelo que entiende que la variedad de medidas de pecho fue idéntica con aquella conformada por el tipo de medidas repetidas que son comunes en astronomía. En terminología moderna, Quetelet alega que las medidas de los pechos de la Tabla 1 están normalmente distribuidas.

Medida	Frecuencia
33	3
34	18
35	81
36	185
37	420
38	749
39	1073
40	1079
41	934
42	658
43	370
44	92
45	50
46	21
47	4
48	1
	5738

Tabla 1: Contorno del pecho de soldados escocéses

El lector debe sacar sus propias conclusiones recordando la cercanía del intento de ajuste de la Figura 5. La prueba χ^2 de normalidad arroja un valor de χ^2_{prueba} de 47.1, el cual es más grande que el valor crítico de $\chi^2_{10,0.05} = 18.3$, significando que por estándar modernos esos datos no pueden verse como distribuidos normalmente. (El número de valores desechados fue reducido de 16 a 10 porque seis de éstos son muy pequeños.) La discrepancia indica que la justificación de Quetelet de su afirmación de normalidad de las medidas de los pechos merece una sustancial dosis de escepticismo. Esto aparece aquí traducido:

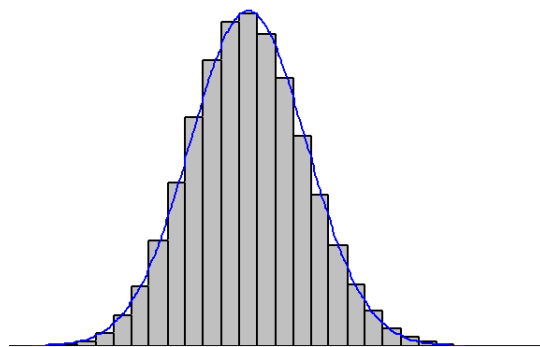


Figura 8: ¿Los datos están normalmente distribuidos?

“Yo ahora pregunto ¿será exagerado, hacer una apuesta ecuánime, que una persona poco experimentada en medir cuerpos humanos cometa un error de un a pulgada midiendo un pecho de más de 40 pulgadas de circunferencia? Bien, admitiendo este probable error, 5738 medidas hechas sobre un único individuo podrían por supuesto darse con más regularidad que en grupos, como para el orden de magnitud, que las 5738 medidas hechas sobre los soldados escocés; y si las dos series son dadas sin diseños particulares, podría ser violento establecer cuál serie fue de 5738 soldados diferentes y cuál fue obtenida de un individual con menos aciertos y más primitivos medios de apreciación.”

Este argumento es poco convincente. Podría ser una persona extraña, efectivamente, que produzca resultados que divergen por 15 para una medida de pecho de 40. Ninguna idiosincrasia o inusual condición (fatiga, por ejemplo) puede producir tal irrazonable medida, es más que probable para distorsionar el proceso de medida entero para el punto que los datos podrían desaprobado ser normales.

Es interesante notar que Quetelet fue el hombre quien acuñó la frase el *hombre promedio*. En efecto, él ve este concepto mítico como una forma ideal, la cual manifiesta formas corpóreas construidas como medidas que son rodeadas por errores:

“Si el hombre promedio fue completamente determinado, podemos... considerar él como un tipo de perfección; y toda diferencia de sus proporciones o condición, puede constituir deformidad y enfermedad; todo encuentro distinto, no solamente como respetada proporción o forma, pero como exceso en el límite observado, podría constituir una monstruosidad.”

Quetelet fue muy explícito acerca de las aplicaciones de esto, ahora poniendo en duda, principalmente a los soldados escocés. Él se toma la libertad de ver la medida de los pechos de los soldados como una repetida estimación del pecho del soldado promedio:

Yo lo sorprenderé a usted en decir que el experimento tiene que estar hecho. Si, realmente, más de un millar de copias de una estatua tienen que ser medidas, y aunque yo no acierte a la del Gladiador, estas difieren, en cualquier caso, solo ligeramente de esta: esas copias son iguales a una, así que las medidas son tomadas con todo el posible chance de error, añadiré, además, que las copias son sujetas a deformidad por un sinnúmero de causas accidentales. Uno puede esperar encontrar aquí un considerable probable error .

Finalmente, puede ser notado que la Tabla 1 contiene errores sustanciales. Los datos originales fueron divididos en tablas para once milicias locales, se clasificó cruzando por altura y medida del pecho, con totales marginales, y Quetelet identifica numerosos errores en la extracción de sus datos. El conteo actual es mostrado en la Tabla 2, donde ellos son comparados con los conteos de Quetelet.

Contorno	Actúal frecuencia	Frecuencia de Quetelet
33	3	3
34	18	18
35	81	81
36	189	185
37	409	420
38	753	749
39	1062	1073
40	1079	1079
41	934	934
42	658	658
43	370	370
44	92	92
45	50	50
46	21	21
47	4	4
48	1	1
	5738	5732

Tabla 2: Contorno del pecho de soldados escocéces.

El libro de Quetelet fue evaluado muy favorablemente en 1850 por el eminente y ecléctico científico británico John F. Herschel. Esta extensiva evaluación contiene las ideas más generales de los diferentes desarrollos de la curva del error de Gauss, el cual comienza con las siguientes tres afirmaciones:

1. ...la probabilidad de la ocurrencia de dos o más eventos simples independientes, es el producto de las probabilidades de estas consideradas individualmente;
2. ...el más grande error tiene más pequeña probabilidad...
3. ...errores son igualmente probables si son iguales en cantidad numérica...

El tercer postulado de Herschel es mucho más fuerte que la superficial afirmación similar de simetría de Gauss y Galileo. El último es unidimensional y es formalizado como $\phi(\epsilon) = \phi(-\epsilon)$ mientras que la forma es multidimensional y es formalizada aseverando la existencia de la función ψ tal que

$$\psi(x)\phi(y) \cdots \phi(t) = \psi(x^2 + y^2 + \cdots t^2).$$

Esencialmente, el mismo desarrollo tiene ya que ser publicado por la Americas R. Adrain en 1808, previo a la publicación del artículo de Gauss pero posterior a la localización de Ceres. En su artículo de 1860 sobre la teoría cinética de gases, el renombrado físico británico J. C. Maxwell repite el mismo argumento y usa estas palabras:

“Encontrar el número promedio de partículas cuyas velocidades se encuentran entre determinados límites, después de un gran número de colisiones entre un gran número de iguales partículas. ”

El Más notable entre sus proponentes fue el caballero inglés y escolar Sir Francis Galton, quien continuo abarcando varias décadas. Este aspecto de su carrera comienza con su libro de 1869, *Hereditary Genius*, en el cual él busca probar que los genios corren en las familias. Como él fue consciente de que la excepción a esta regla abunda, esto tenía que ser verificado con una estadística, mejor dicho, una absoluta verdad. Fue necesario una eficiente cantidad de herramientas para describir poblaciones que fueron proporcionadas por Quetelet, quien afirma sobre el amplio rango de aplicaciones de la curva del error de Gauss que Galton ha encontrado y adoptado en 1863.

Como la descripción del preciso uso de Galton hecho de la curva normal quien toma lejos de su campo, nosotros sólo discutimos sus explicaciones para la omnipresencia de la distribución normal. En sus palabras:

“Considerando la importancia de los resultados que admiten siendo desarrollado siempre que la ley de la frecuencia de errores puede mostrarse la aplicación, se darán algunas razones del porque de la aplicabilidad es más general que puede tener ser esperado de la muy artificial hipótesis sobre la cual la ley se basa. Esto será mencionado que estas son para el efecto que errores individuales de observaciones, o individuales diferencias en objetos pertenecientes al mismo grupo genérico, son completamente debido a la acción agregada de variables influenciables en diferentes combinaciones, y que esas influencias deben ser: ”

1. toda independencia en sus efectos,
2. todas iguales,
3. todas admitiendo de ser tratadas como simples alternativas “arriba del promedio” o “abajo del promedio;”
4. las usuales Tablas son calculadas sobre la adicional suposición que las variables influenciables son infinitamente numerosas.

Esto es, claro está, una informal nueva exposición del Teorema del Límite Central de Laplace. El mismo argumento tiene que ser promovido por Herschel. Galton fue completamente consciente de que las condiciones (1-4) nunca actualmente ocurren en la naturaleza y trata de mostrar que ellos son innecesarios. Su argumento, sin embargo, fue impreciso e inadecuado. Durante los pasados dos siglos el Teorema del Límite Central tiene que ser enormemente generalizado y una más fresca versión existe, conocida como el Teorema de Lindeberg, la cual hace posible distribuir con el requerimiento (2). Por tanto, la curva de DeMoivre en (1) surge como el límite de la distribución observable, incluso cuando se agregan “átomos” que poseen una variedad de distribuciones no idénticas. En general, parece ser común para estadísticos atribuir los grandes éxitos de la distribución normal a esa versión generalizada del Teorema del Límite Central. La creencia de Quetelet de que todas

las desviaciones de la media son tratadas como errores en el proceso que busca replicar una cierta idea que tiene relegada a la misma basura que contiene la phlogiston y cualquier teoría.

6. ¿Por qué el término *normal*?

Una palabra debe decirse acerca del original término *normal*. Su cualidad apropiada es atestiguar por el hecho que tres científicos independientemente inician su uso para describir la curva del error

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

Este fue el americano C. S. Peirce en 1873, el inglés Sir Francis Galton en 1879, y el alemán Wilhelm Lexis, también en 1879. Su generalizado uso es probablemente debido a la influencia del gran estadístico Karl E. Pearson, quien dice esto en 1920:

Hace muchos años [en 1893] yo llame a la curva Laplace-Gauss la curva normal, cuyo nombre, mientras esto evita la pregunta internacional de prioridad, tiene la desventaja de llevar gente a confiar que todas las otras distribuciones de frecuencia son en un sentido u otro anormal.

Al principio fue costumbre referirse a la curva del error de Gauss como la *curva Gaussiana* y Pearson, inconsciente del trabajo de Moivre, y fue también difícil reservar algún crédito al descubrimiento de Laplace. A través del tiempo él comprendió su error: la curva normal ha llegado a ser una de las más extensamente aplicadas herramientas de todas las matemáticas y nunca han podido cambiar su nombre. Es muy apropiado que la curva del error pueda basarse sobre una falsa idea.

Agradecimientos

El autor agradece a sus colegas James Church y Ben Cobb por su ayuda, a Martha Siegel por su estímulo, y a los árbitros anónimos por sus críticas constructivas. La mayoría de información contenida en este artículo vino de excelentes historias.

Los traductores agradecen al autor Saul Stahl por su amabilidad, ayuda y disposición con la Universidad Santo Tomás y al Centro de Investigaciones y Estudios Estadísticos por su apoyo en el proceso de traducción.

Referencias

Stahl, S. (2006), 'The evolution of the normal distribution', *MATHEMATICS MAGAZINE* 79(2), 96–113.