

Una aproximación histórica a la evolución de la curva normal

Luis Alejandro Másmela Caíta*

*Universidad Distrital Francisco José de Caldas
Universidad Autónoma de Colombia*

Jenny Carolina Serrato Rincón**

*Escuela Colombiana de Carreras Industriales
Universidad de La Salle*

Resumen

Este documento se basa en el artículo de Saul Stahl [7] titulado "The evolution of the normal distribution". Se hace una breve revisión histórica de una de las más importantes distribuciones de probabilidad: la distribución normal. También conocida como distribución acampanada o distribución gaussiana, utilizada en diferentes campos de la Ciencia. Fue nombrada así en honor del matemático alemán Carl Friedrich Gauss, quien fue el primero que presentó una prueba formal para la distribución de errores. Este documento presenta algunas intervenciones históricas de diferentes matemáticos que intentaron, de diferentes maneras, lograr encontrar una forma funcional para la distribución. La evolución histórica se presenta enfatizando sobre la distribución en dos aspectos: el primero como una distribución de los errores en el trabajo experimental y, por otro lado, como una herramienta para el cálculo en problemas de probabilidad.

Palabras y Frases Claves: Distribución Normal, Distribución acampanada, Campana de Gauss, Método de Mínimos Cuadrados, Media.

Abstract

This paper is based on the Saul Stahl [7] paper's called "The evolution of the normal distribution". A brief historical review of one of the most important probability distributions is done. The normal distribution, know as the bell shaped distribution or the gaussian distribution as well, is used in different scientific studies. Called in honor of the German mathematician Carl Friedrich Gauss who was the first who presented a formal proof for the errors distribution. This paper show some historical interventions of several mathematicians that tried, in different ways, of achieved a functional form for this distribution. The historical evolution is presented using two focus: The first as one distribution of the errors in the experimental work and, by other side, as a tool for made the computation in probability problems.

Key works: Normal Distribution, Bell curve, Bell Distribution, Least Squares Method, Mean.

* lamasmelac@unal.edu.co

** jcserrator@bt.unal.edu.co

1. Introducción

En el presente documento, basado en el artículo "The evolution of the normal distribution" del profesor Saul Stahl [7], se realiza una revisión cronológica de una de las distribuciones más importantes en el campo de la estadística, "la distribución normal", más conocida como la distribución de Gauss; utilizada por casi todos los profesionales en diferentes campos, salud, ciencias sociales, física y actuaria, entre otras. Es también conocida por su forma gráfica como la distribución en forma de campana o curva acampanada.

Se propone una aproximación a la evolución de la curva normal desde una perspectiva de tipo cronológico, ahondando en dos enfoques básicos que constituyen la génesis de lo que hoy conocemos como la distribución de probabilidad normal, en primer lugar el enfoque a partir de la teoría de los errores, iniciado por los antiguos astrónomos y que lleva al planteamiento de una distribución de probabilidad que modela el comportamiento de tales errores, y por otro lado como una herramienta que permite aproximar el cálculo de ciertos tipos de probabilidades.

2. Enfoque histórico de la evolución de la curva normal

A través de la historia, diferentes matemáticos, de manera paralela, buscaron la forma de una curva que modelara la distribución de los errores cometidos, al intentar realizar distintos tipos de mediciones. Son los astrónomos los que inicialmente se enfrentan al problema de, cómo tratar los errores cometidos al realizar mediciones para encontrar valores "exactos" de cierta cantidad. Incluso, el problema para algunos de ellos se basó en no aceptar un conjunto de mediciones para estimar una cantidad determinada. Para otros, en cómo obtener un solo valor a partir de un conjunto de datos. Algunos de ellos promediaron, otros obtuvieron la mediana de los datos, otros más agruparon los datos y utilizaron a la vez la media y la mediana.

Se hace referencia a una época en la cual calcular medias o medianas era mal visto, puesto que se trataba de procesos que no tenían un soporte científico, además, los elementos de medición empleados para contrastar empíricamente las predicciones, fueron bastante rudimentarios.

Fue el italiano Galileo Galilei, hacia 1632, el primero que argumentó por escrito que debía hacerse un tratamiento sistemático y científico a la medición de los errores. Y es él quien resumió de manera informal en cinco puntos su análisis:

1. Existe un único valor numérico, el cual da la verdadera distancia de la estrella al centro de la tierra.
2. Todas las observaciones que se tomen respecto a esta medida están cargadas de errores debido al observador, a los instrumentos utilizados y a las condiciones observacionales.
3. Los datos obtenidos a partir de las observaciones se distribuyen de manera simétrica entorno al valor verdadero; o de manera equivalente, los errores cometidos se distribuyen de manera simétrica entorno de cero.
4. Errores pequeños ocurren de manera más frecuente que errores grandes.

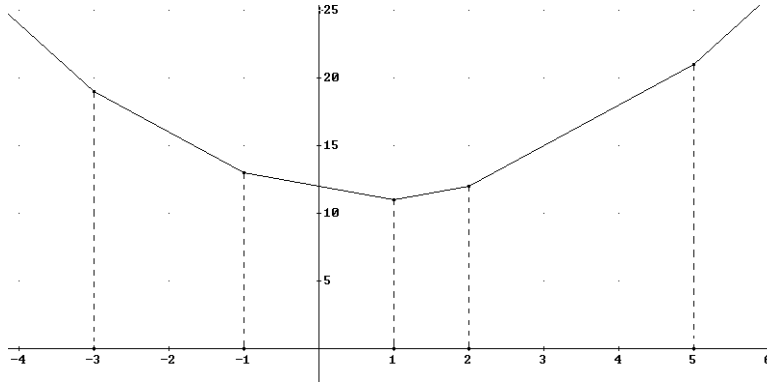


Figura 1: Mediana para un número impar de valores.

5. La distancia calculada se obtiene como función de las observaciones angulares, de modo que, un ajuste pequeño en las observaciones puede resultar en una variación grande en la distancia calculada.

Desafortunadamente Galileo no dio indicios del proceso de cómo encontrar la verdadera distancia, pero se puede esperar que cuando escribió: “errores pequeños ocurren de manera más frecuente que errores grandes”, Galileo haya minimizado la suma de desviaciones al valor buscado, que no son más que el valor absoluto de la diferencia entre la observación y el valor de interés. Se puede mostrar que el valor que minimiza esta suma es la mediana.

Denotando por x el valor buscado y x_1, x_2, \dots, x_n las observaciones, la función que calcula las desviaciones al valor buscado está dada por

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

Para un conjunto con un número impar de datos, suponga $x_1 = 5, x_2 = -3, x_3 = 2, x_4 = 1$ y $x_5 = -1$, la función de desviaciones al valor buscado toma la forma

$$f(x) = |5 - x| + |-3 - x| + |2 - x| + |1 - x| + |-1 - x|$$

y su gráfica se ilustra en la Figura 1.

Se observa que los vértices de la gráfica en la Figura 1, corresponden a los valores de las observaciones y que ésta las ordena de menor a mayor, esto es, $x_{(1)} = -3, x_{(2)} = -1, x_{(3)} = 1, x_{(4)} = 2$ y $x_{(5)} = 5$. La gráfica alcanza su mínimo en la observación del medio, $x_{(3)} = 1$.

Para un número par de datos, digamos $x_1 = 5, x_2 = -3, x_3 = 1$ y $x_4 = -1$ la función toma la forma

$$f(x) = |5 - x| + |-3 - x| + |1 - x| + |-1 - x|$$

y su gráfico se da en la Figura 2.

De igual forma la función ordena los datos de menor a mayor, $x_{(1)} = -3, x_{(2)} = -1, x_{(3)} = 1, x_{(4)} = 5$. La gráfica, en la Figura 2 alcanza su mínimo en cualquier punto entre $x_{(2)} = -1$ y $x_{(3)} = 1$,

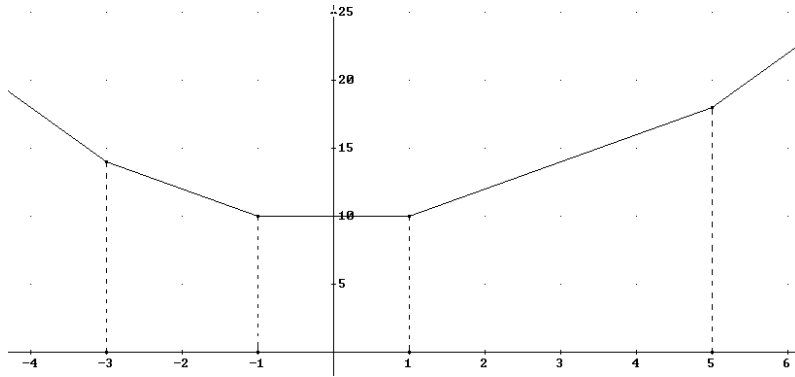


Figura 2: Mediana para un número par de valores.

por lo general se toma el punto medio. Se puede mostrar que el valor que minimiza la suma en (1) es la mediana del conjunto de datos x_1, x_2, \dots, x_n , que se define como:

$$\tilde{x} = \begin{cases} x_{(n/2+0,5)} & \text{si } n \text{ es impar} \\ \frac{1}{2} [x_{(n/2)} + x_{(n/2+0,5)}] & \text{si } n \text{ es par} \end{cases}$$

en donde $x_{(i)}$ denota el i -ésimo dato ordenado.

Se atribuye la aparición de las probabilidades al intercambio de correspondencia, hacia 1654, entre los franceses Pierre Fermat y Blaise Pascal, motivados por las peticiones que realizó el apostador Chevalieré de Mere. Uno de los primeros trabajos conocidos y realizados por Fermat y Pascal se trató de cálculos de sumas de probabilidades binomiales del tipo (2). Básicamente trabajaron sobre el cálculo de la probabilidad que se dieran entre i y j éxitos en n repeticiones independientes de cierto experimento. Este tipo de experimento denominado proceso de Bernoulli, se refiere al más simple de los experimentos aleatorios. Ellos trabajaron sobre el cálculo de este tipo de sumatorias, específicamente para valores pequeños de n .

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

Para 1712, con base en estas sumas binomiales, Willem 'sGravesande, filósofo y matemático holandés, probó en Londres, que la proporción de nacimientos masculinos y femeninos eran iguales, en el periodo de 82 años comprendido entre 1629 y 1710. La prueba se realizó basándose en la información recolectada en estos años, y se calcularon probabilidades que involucran aproximadamente 15.000 nacimientos por año. Las expresiones que surgieron para el cálculo requerido, involucró sumas binomiales con grandes valores de n , los cuales eran complejos para la época. Él observó que el número de nacimientos para el periodo de interés oscilaban entre $7765/15448 = 0,5027$ en 1703 y $4748/8855 = 0,5362$ en 1661. 'sGravesande multiplica estos cocientes por 11429, el número promedio de nacimientos que estos 82 años abarca. Obteniendo los límites de 5745 y 6128 sobre el número de nacimientos masculinos en cada año. Bajo el supuesto que la proporción de nacimientos

sea igual, es decir que $p = 1/2$, la probabilidad del exceso de nacimientos masculinos observados debido a la aleatoriedad es de solamente la potencia 82 de

$$P\left(5745 \leq x \leq 6182 \mid p = \frac{1}{2}\right) = \sum_{x=5745}^{6182} \binom{11429}{x} \left(\frac{1}{2}\right)^{11429} \\ \approx \frac{3849150}{13196800} \approx 0,292.$$

'sGravesande para el cálculo de estas sumatorias 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 esta se convierte en una tarea tediosa. Ya que la probabilidad de esta diferencia en tasas de nacimientos, basado en 82 años es la extremadamente pequeña cifra de $0,292^{82}$, 'sGravesande concluye que la mayor tasa de nacimientos masculino se debe a la *intervención divina*.

Para este tipo de sumas binomiales, en donde el valor del parámetro n es grande y su cálculo resulta un trabajo arduo, el francés Abraham De Moivre en 1721 se puso en la tarea de encontrar métodos de aproximación para ellas. Hacia 1733 él encontró un par de estas aproximaciones, ver (3) y (4), las cuales involucran la función exponencial natural. Éstas hacen parte de cualquier curso introductorio de probabilidad, proceso conocido como aproximación de la binomial por la normal, o como el teorema de Moivre en honor a este gran matemático.

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

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

La aproximación de (3), debida a Moivre, se ilustra en la Figura 3, allí se toman valores de $n = 10$ y $n = 100$. Los puntos ilustran los valores de la parte izquierda de la expresión en (3), mientras que la curva los valores de la función de la derecha y que involucra la función exponencial. Se puede observar que no todos los puntos caen sobre la curva ya que se trata de una aproximación.

La Figura 4, muestra formas funcionales de distribuciones binomiales con valor de parámetro $p = 0,25$ y valores del parámetro $n = 50$, $n = 100$, $n = 150$ y $n = 200$. Distribuciones de esta forma, en donde el valor de n se hace grande, permiten ser aproximadas cada vez mejor a través de la distribución normal, resultado conocido como teorema de Moivre. La expresión en (4) muestra un caso particular para el teorema de Moivre o teorema de aproximación de la binomial por la normal, para el caso particular de $p = 1/2$.

Como se mencionó al principio, desde la época de Galileo, surgió el problema, de cómo estimar una cantidad desconocida. También de cómo proceder ante la toma de varios datos a favor de estimar dicha cantidad. El concepto de media o promedio requirió mucho tiempo de evolución y aceptación para ser lo que ahora es. Históricamente, hubo oposición a obtener de un conjunto de datos un valor representativo, incluso importantes figuras rechazaron la idea de repetidos experimentos, para obtener la medida de una magnitud de interés.

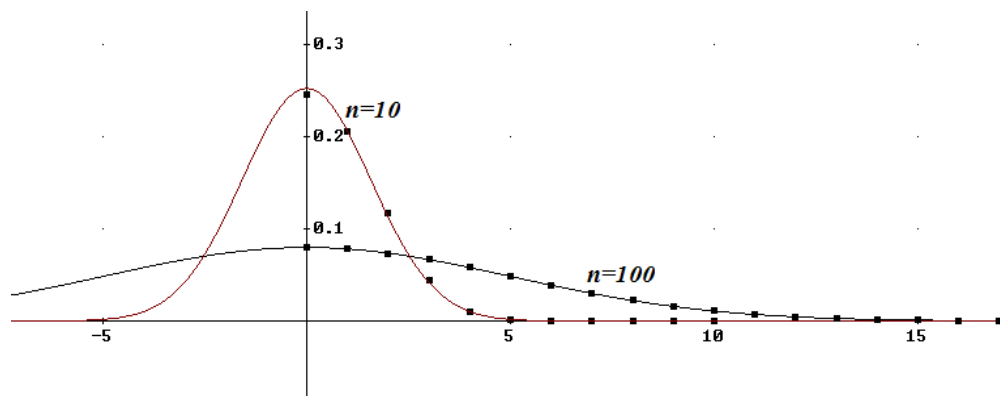


Figura 3: Aproximación de probabilidades binomiales puntuales debidas a Moivre

En 1722 Roger Cotes, matemático inglés, se refirió a las observaciones como fichas colocadas en puntos sobre una recta numérica, asociando a cada una de ellas cierto peso. Cotes afirmó que el valor representante de este sistema se puede hallar encontrando el punto sobre la recta numérica sobre el cual, al poner un pivote, dicho sistema queda balanceado. Aquel punto es conocido como el centro de gravedad y no es más que un promedio ponderado. Cuando todas las fichas tienen el mismo peso se refiere a la media o promedio.

Cotes aparentemente visualiza las observaciones como fichas x_1, x_2, \dots, x_n con respectivos pesos físicos w_1, w_2, \dots, w_n alineados sobre una abscisa. 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, en física, es común asumir que un cuerpo concentra toda su masa en su centro de gravedad y así puede decirse que la totalidad del cuerpo está representado por un simple punto. Cotes propuso centros de gravedad que concuerdan con los pesos promedio de las masas que el sistema involucra. Por la definición de centro de gravedad, si el pivote es desplazado a la abscisa Z este se balanceará

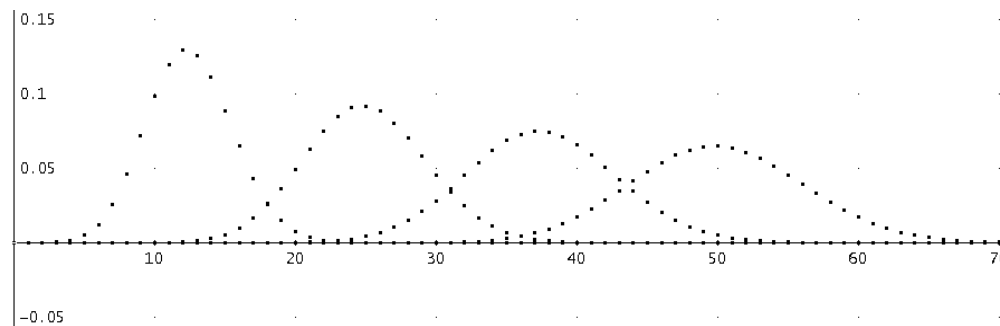


Figura 4: Teorema de Moivre para binomiales con parámetro $p = 0,25$. Ilustración para valores de $n = 50$, $n = 100$, $n = 150$ y $n = 200$.

y por la ley de Arquímedes de las Palancas,

$$\sum_{i=1}^n w_i(Z - x_i) = 0 \quad (5)$$

o

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

Claro esta, cuando los pesos w_i son iguales, Z se convierte en el promedio clásico

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

Este procedimiento mostró una temprana aparición del método de mínimos cuadrados. El conjunto de valores está representado por el valor que minimiza la suma de desviaciones al cuadrado ponderadas,

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

La función en (7) se minimiza derivando e igualando a cero, la ecuación que surge luego de este proceso resulta equivalente a la que obtuvo Cotes, ecuación (5), aplicando sobre el esquema de las fichas, la Ley de las Palancas de Arquímedes. Al igual que la mediana, la media, como medida representativa de un conjunto de datos, resulta de la minimización de una función que surge de manera natural.

Para el ejemplo del conjunto de datos par en el caso de la mediana, descrito con anterioridad, la función de la suma de desviaciones al cuadrado ponderadas para los valores $x_1 = 5, x_2 = -3, x_3 = 1$ y $x_4 = -1$, asumiendo pesos iguales toma la forma

$$g(x) = (x - 5)^2 + (-3 - x)^2 + (1 - x)^2 + (-1 - x)^2$$

que simplificada se convierte en

$$g(x) = 4x^2 - 4x + 36,$$

su gráfica que se muestra en la Figura 5, permite observar que el mínimo se alcanza en el valor de $x = 1/2$, efectivamente el promedio de los datos.

Pese a que para el siglo XVIII, persistieron dudas sobre el método de obtención de un único valor de un conjunto de experimentos, el matemático británico Thomas Simpson, en 1756, intentó obtener curvas que especificaran funciones de densidades de probabilidad, asumiendo un valor máximo sobre el posible error cometido, además de probabilidades proporcionales a una secuencia discreta de errores o a progresiones de tipo aritmético. Propuso también intentos por obtener funciones continuas que modelaran la distribución de dichos errores.

Simpson calculó la probabilidad que el error en la medida de varias observaciones no exceda una cota dada cuando los errores individuales toman valores

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

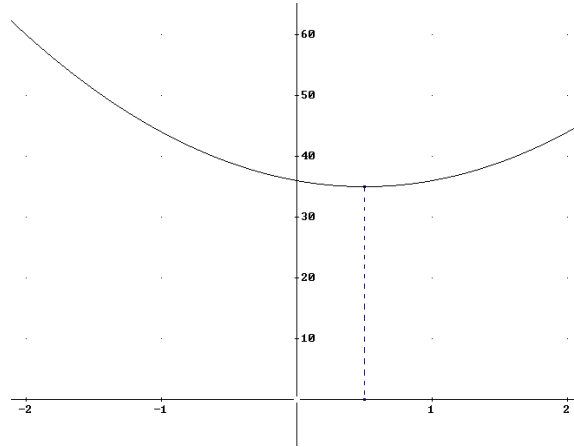


Figura 5: Gráfica de la función de la suma de desviaciones al cuadrado para $x_1 = 5$, $x_2 = -3$, $x_3 = 1$ y $x_4 = -1$.

con probabilidades que son proporcionales a cualesquiera de dos progresiones

$$r^{-v}, \dots, r^{-3}, r^{-2}, r^{-1}, r^0, r^1, r^2, r^3, \dots, r^v \quad (8)$$

o

$$r^{-v}, 2r^{1-v}, 3r^{2-v}, \dots, (v+1)r^0, \dots, 3r^{v-2}, 2r^{v-1}, r^v. \quad (9)$$

En la Figura 6, se pueden observar gráficas para tales probabilidades, asumiendo $r = 1$. La secuencia de puntos arriba de la gráfica ilustra la progresión en (8), mientras que la secuencia abajo ilustra la progresión en (9).

En 1744 Pierre Simon Laplace, astrónomo, físico y matemático francés, presentó una primera curva, obtenida bajo el supuesto que ésta debe ser simétrica y no decreciente para valores positivos de x . Además, no se tenía una relación conocida entre la función y su derivada, supuso sobre éstas la relación más sencilla, una relación de proporcionalidad. Al ser dicha función no decreciente, el cociente entre ella y su derivada debía ser una constante negativa.

Al suponer Laplace una función simétrica y decreciente (para $x > 0$), y al no conocer nada acerca de $\phi(x)$ y $\phi'(x)$, presumiblemente, asume que ellas son proporcionales, esto se puede escribir como

$$\frac{d\phi(x)}{dx} = -m\phi(x),$$

la resultante ecuación diferencial se resuelve facilmente, obteniendo

$$\phi(x) = \frac{m}{2} e^{-m|x|}. \quad (10)$$

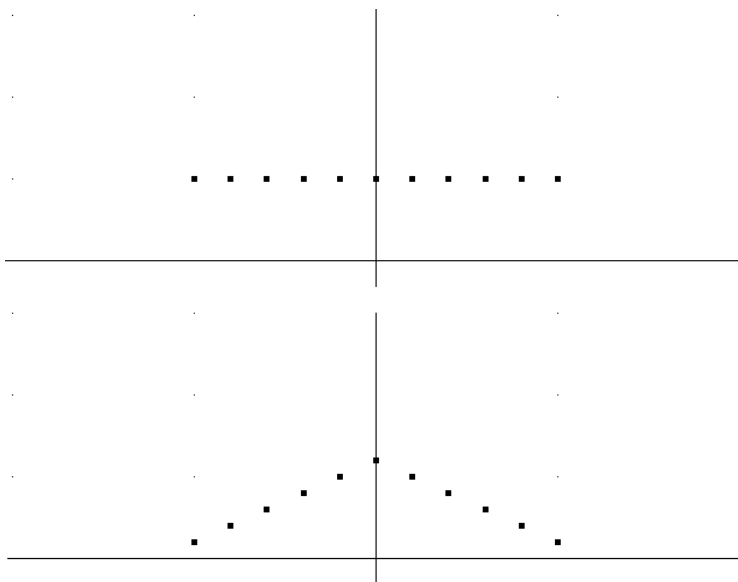


Figura 6: Gráficas para la distribución de los errores propuestas por Simpson.

La curva de error extractada y que corresponde a la ecuación (10) se muestra en la Figura 7.

Laplace no se perturbó por haber obtenido, con base en estos supuestos, una función no derivable en cero. Tres años más tarde, es el mismo Laplace quien propuso una segunda curva, procedió sobre un intervalo unitario, realizando sobre él particiones en subintervalos y trabajó sobre sus longitudes. Bajo este proceso y asumiendo un valor a , como cota superior para los errores, obtuvo una curva proporcional al $\ln(a/x)$ en el intervalo $(0, a]$. Más exactamente él procede de las siguiente manera, 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, dividiendo el intervalo en $n + 1$ espacios. Ordenando estos espacios de tal forma que

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

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 haga n infinitamente grande. El límite de las configuraciones es una curva que resulta 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, y su ecuación se presenta a continuación,

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

La curva obtenida, ver Figura 8, corresponde a una función con dominio finito, y una singularidad infinita en cero. Esta segunda curva fue considerada un retroceso en la búsqueda de la tan anhelada curva normal.

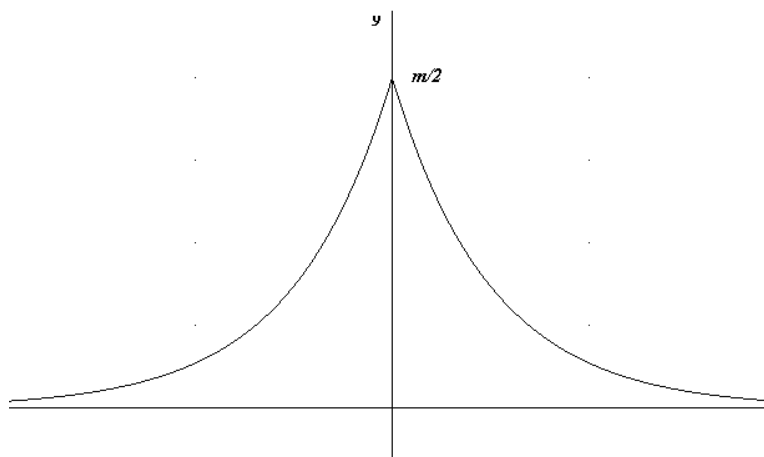


Figura 7: Primera curva de Laplace

Para 1777, Daniel Bernoulli, el más multifacético de los magníficos geómetras de la familia Bernoulli, bajo el mismo supuesto de errores acotados, supuso sobre su curva propuesta, que fuera derivable en el valor donde alcanza su máximo, es decir que la pendiente de la recta tangente fuera allí horizontal, y grande en valor absoluto en vecindades de los extremos del dominio, es decir una recta tangente próxima a ser vertical en cercanías a los límites del dominio. Una primera curva propuesta por Bernoulli fue una semielipse y luego de algunos argumentos lógicos adicionales propuso como curva un semicírculo.

El final de la búsqueda se dio hacia 1801. Para esta época se observa en el firmamento un cuerpo celeste, el cual se sospechó fuera un planeta. Tomaron mediciones respecto a su órbita y luego de seis meses el cuerpo desapareció próximo al sol. El fenómeno celeste llamó la atención de los mejores astrónomos en Europa. El objetivo fue, con la información que se tenía respecto a su órbita, pronosticar en qué región del cielo éste aparecería. Es en este momento en donde hizo su aparición el joven matemático alemán, Karl Friedrich Gauss, conocido como “El príncipe de las matemáticas”. Gauss señaló una región del cielo donde haría su aparición el supuesto nuevo planeta, lejos de aquella región que propusieron los diferentes astrónomos de la época. De manera acertada, fue la región del cielo que propuso Gauss correcta, dicho planeta fue bautizado por Giuseppe Piazzi como Ceres de Ferdinanda, por la diosa griega siciliana, y el rey Fernando IV de Nápoles y Sicilia. Más adelante, el Ferdinanda se eliminó por razones políticas. Ceres resultó ser el primer asteroide que se observaba del cinturón de asteroides y, hasta ahora, el de mayor tamaño conocido. Gauss explicó que la metodología utilizada para la estimación mencionada, se basa en el método denominado “mínimos cuadrados” y que permite obtener la órbita que mejor se ajusta a los datos que se tenían. Este criterio es justificado por las tres afirmaciones siguientes:

1. Pequeños errores son más probables que grandes errores.
2. Para un valor real ϵ , la probabilidad que se cometa en la medición un error de magnitud ϵ y $-\epsilon$ resultan iguales.

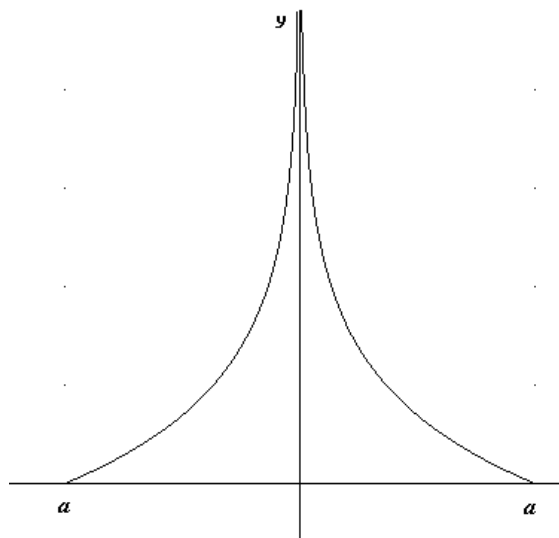


Figura 8: Segunda curva del error de Laplace

3. En la presencia de varias observaciones sobre la misma cantidad, el más probable valor de esa cantidad es el promedio de las observaciones.

Sobre la base de estas afirmaciones, el joven científico concluye que la densidad de probabilidad para el error (esto es, la curva del error) esta dada por

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

donde h es una constante positiva que es pensada por Gauss como la "precisión del proceso de medida". Él reconoce ésta como la curva de campana determinada por $\mu = 0$ y $\sigma = 1/\sqrt{2}h$. El proceso de deducción de esta función hizo uso de algunos conceptos básicos de la probabilidad y el cálculo.

La obtención de esta función está basada en el método de mínimos cuadrados, Gauss se refirió a dicho proceso como "...el más importante problema de la aplicación de las matemáticas en la filosofía natural..." y que trata de buscar los valores más probables para k parámetros basado en n observaciones, teniendo en cuenta la incertidumbre introducida por los errores en las observaciones. La imagen de Gauss, junto con su principal resultado, la distribución en forma de campana, aparece estampada en los billetes, ya extintos, de 10 marcos alemanes, ver Figura 9.

Durante la primera mitad del siglo XIX, los científicos utilizaron la teoría de errores en distintas ramas del saber, ampliando sus resultados. Una de las primeras aplicaciones de la curva normal fue debida al alemán Friedrich Bessel en 1818, que comprobó que los errores cometidos en 300 medidas astronómicas coincidían con bastante aproximación a los previstos por Gauss mediante la curva normal.



Figura 9: Billeto de 10 marcos alemanes, con la figura de Gauss y en el fondo la distribución acampanada.

Finalmente y para concluir, se presenta a continuación, el desarrollo matemático realizado por Gauss, para la deducción de la ecuación de la famosa curva, conocida en la literatura como la campana de Gauss.

3. Proceso seguido por Gauss para la deducción matemática de la función de la distribución normal

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 asumir 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

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

Note 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

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

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

$$f(M_1 - \bar{M}) + f(M_2 - \bar{M}) + \dots + f(M_n - \bar{M}) = 0. \quad (11)$$

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

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

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

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

con lo que la prueba se completa. \square

Referencias

- [1] Castro, I. Cómo hacer matemáticas con DERIVE. Reverté de Colombia S.A. Bogotá. 1992.
- [2] De Moivre, A. Approximatio ad Summam Terminorum Binomii $(a + b)^n$ in Seriem Expansi. Printed for private circulation, 1733.

- [3] Feller, W. An Introduction to Probability Theory and its Applications vol. 2. John Wiley & Sons. New York.1968.
- [4] Hacking, I. El surgimiento de la probabilidad. Barcelona, España. Editorial Gedisa. 1995.
- [5] Hald, A. A History of Probability and Statistics and Their Applications before 1750. New York, John Wiley & Sons, 1990.
- [6] Pearson, K. "Historical Note on the Origin of the Normal Curve of Errors". Biometrika 16. (1924) 402-404.
- [7] Stahl, S. (2006), "The evolution of the normal distribution", Mathematics Magazine. 79(2), 96-113.