

Clase 14: La distribución normal

Matías Carrasco

28 de agosto de 2019

Resumen Veremos cómo la simetría rotacional da origen a la fórmula de la campana de Gauss. Luego estudiaremos sus propiedades.

Índice

La distribución normal o gaussiana	1
Variables normales independientes	7
Apéndice: La deducción original de Gauss	10

La distribución normal o gaussiana

La distribución normal es una de las distribuciones más importantes en probabilidad y estadística. Veamos con un ejemplo sencillo cómo aparece de forma natural bajo ciertos supuestos de simetría.

■ **Ejemplo 1 — Simetría rotacional.** Supongamos que lanzo un dardo a un tablero de dardos. Apunto al centro del tablero $(0,0)$ pero no soy tan bueno con los dardos, por lo que el dardo cae en una posición aleatoria (X, Y) que tiene una función de densidad conjunta $p : \mathbb{R}^2 \rightarrow \mathbb{R}$. Hagamos dos suposiciones sobre la forma en que juego a los dardos:

1. La densidad $p(x, y)$ tiene simetría rotacional, por lo que la distribución de dónde cae mi dardo solo depende de la distancia del dardo al centro. En otras palabras, no tengo ninguna tendencia a errarle en alguna dirección en particular.
2. Las variables aleatorias X e Y son independientes. Es decir, cuánto le erro a izquierda o derecha no hace ninguna diferencia a la distribución de cuánto le erro hacia arriba o hacia abajo.

Entonces, por el supuesto 1 y el teorema de Pitágoras debo poder expresar la densidad como

$$p(x, y) = g(x^2 + y^2)$$

ya que $r^2 = x^2 + y^2$ es la distancia al origen. Ahora, como las variables aleatorias X e Y son independientes e idénticamente distribuidas, debo poder expresar

$$p(x, y) = p(x)p(y).$$

Pero por lo anterior, $p(x) = g(x^2)$ y $p(y) = g(y^2)$, de donde combinando estas dos propiedades, obtenemos que para cada par (x, y) vale

la igualdad

$$g(x^2 + y^2) = g(x^2)g(y^2).$$

Esto significa que g debe ser una función exponencial¹

$$g(t) = Ae^{-Bt}$$

en donde A es una constante de normalización y B de alguna manera refleja las unidades en las que estoy midiendo las distancias.² La constante B debe ser positiva porque la densidad debería ser una función decreciente de la distancia (no soy tan malo en los dardos).

Es decir, la densidad de X es de la forma

$$p(x) = Ae^{-Bx^2}.$$

Para hallar A debemos integrar $p(x, y)$ en todo \mathbb{R}^2 e imponer la condición de que dicha integral sea igual a uno (pues es una densidad de probabilidad).

Recordar que en coordenadas polares de \mathbb{R}^2 , para integrar una función $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ hay que usar la fórmula (ver la Figura 1)

$$\iint f(x, y) dx dy = \int \int f(r, \theta) r dr d\theta.$$

Consideremos la función $f(x, y) = p(x, y)$. Entonces

$$\begin{aligned} \iint_{\mathbb{R}^2} p(x, y) dx dy &= \int_0^{2\pi} \int_0^\infty A^2 e^{-Br^2} r dr d\theta \\ &= 2\pi A^2 \int_0^\infty r e^{-Br^2} dr \\ &= (u = Br^2, du = 2Br dr) \\ &= \frac{\pi A^2}{B} \int_0^\infty e^{-u} du = \frac{\pi A^2}{B} = 1 \end{aligned}$$

De aquí resulta que $A = \sqrt{B/\pi}$.

La constante B está relacionada con la varianza de X e Y . Calculemos la varianza de X usando la simetría rotacional. Como $E(X) = 0$ también por simetría, la varianza de X es por definición

$$\sigma^2 = E(X^2) = \sqrt{\frac{B}{\pi}} \int_{-\infty}^{+\infty} x^2 e^{-Bx^2} dx.$$

Esta integral se puede calcular mediante integración por partes. Pero podemos usar la densidad conjunta de X e Y para calcularla más fácilmente.

Sea $R^2 = X^2 + Y^2$ la distancia del punto (X, Y) al origen, ver la Figura . Entonces

$$E(R^2) = E(X^2) + E(Y^2) = 2E(X^2) = 2\sigma^2,$$

¹ Si nunca vieron esta propiedad en los cursos de cálculo no entrar en pánico. Es muy simple probarla, pasando de los enteros a los racionales, y luego por continuidad a todos los reales.

² Por ejemplo, si mido la distancia en cm, B será 10 veces más grande que si midiera en mm.

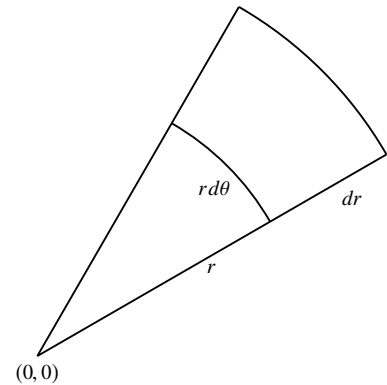


Figura 1: Elemento de área en polares es $dx dy = r dr d\theta$.

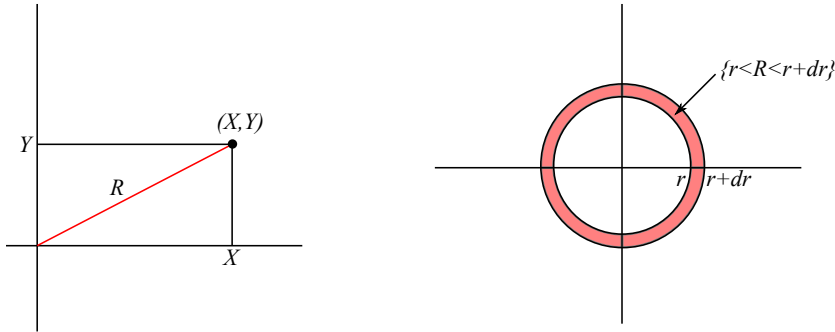


Figura 2: Distancia de (X, Y) al origen.

ya que X e Y tienen la misma distribución.

La densidad de R está dada por

$$p_R(r)dr = (2\pi r dr) \frac{B}{\pi} e^{-Br^2} = 2Bre^{-Br^2} dr, \quad (r > 0).$$

La densidad de $S = R^2$ se obtiene por cambio de variable:

$$p_S(s) = \frac{1}{ds/dr} p_R(r) = 2Bre^{-Br^2} \frac{1}{2r} = Be^{-Bs}, \quad (s > 0).$$

Es decir, $S = R^2$ tiene distribución exponencial de parámetro B . Luego $E(S) = 1/B$, de donde $\sigma^2 = 1/(2B)$. Dicho de forma equivalente $B = 1/(2\sigma^2)$.

Hemos llegado así a la famosa fórmula de la campana de Gauss:

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

Cuando la varianza $\sigma^2 = 1$ decimos que X tiene distribución normal estándar. ■

Densidad normal estándar

La densidad normal estándar es la función

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

definida para todo x real. Una variable aleatoria X tiene distribución normal estándar si es absolutamente continua y su densidad es φ . Escribimos $X \sim N(0, 1)$.

El gráfico de la función φ es conocido como *la campana de Gauss*. Es similar al gráfico de la densidad de Cauchy, salvo que tiende a cero exponencialmente cuando $|x|$ tiende a infinito.

Para hallar la fda de φ debemos integrarla. La forma natural de integrar una función es calcular una primitiva, pero un teorema de *Risch*

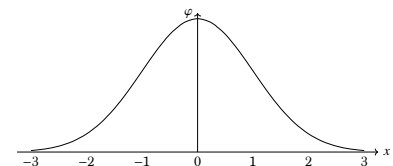


Figura 3: Gráfico de la densidad normal estándar.

afirma que φ no tiene una primitiva elemental. Esto quiere decir que la primitiva de φ no tiene una fórmula que se pueda escribir, mediante el uso de composiciones, sumas, restas, multiplicaciones y divisiones, a partir de un número finito de exponenciales, logaritmos, funciones trigonométricas y raíces n -ésimas. Es por eso que para calcular probabilidades con la normal estándar usamos tablas o computadoras.

Pese a esto, hemos podido calcular la esperanza y la varianza de una variable X con distribución normal estándar sin problemas:

$$E(X) = 0 \quad \text{y} \quad \text{Var}(X) = 1.$$

Supongamos que queremos calcular la probabilidad de que X pertenezca a un cierto intervalo I . En teoría debemos calcular la integral

$$P(X \in I) = \int_I \varphi(x) dx$$

pero no conocemos la primitiva de φ . Notar que si $I = [a, b]$, la integral anterior se puede escribir como la diferencia

$$\int_I \varphi(x) dx = \int_{-\infty}^b \varphi(x) dx - \int_{-\infty}^a \varphi(x) dx = \Phi(b) - \Phi(a),$$

en donde hemos puesto

$$\Phi(y) = \int_{-\infty}^y \varphi(x) dx.$$

La función Φ es la función de distribución acumulada de la densidad normal estándar. Es decir, $\Phi(x)$ representa el área debajo el gráfico de φ desde $-\infty$ hasta x . En términos probabilísticos $\Phi(x)$ representa la probabilidad de que X sea menor o igual a x : $\Phi(x) = P(X \leq x)$.

■ **Ejemplo 2 — Lectura directa de la tabla.** La lectura directa de la tabla consiste en, dado x , hallar $\Phi(x)$. Así, por ejemplo $\Phi(2.31) = 0.9896$. Notar que la tabla no provee los valores $\Phi(x)$ para valores negativos de x . Sin embargo, podemos usar la simetría de la densidad normal (igualdad de áreas rojas en la figura) para deducir que

$$\Phi(-x) = 1 - \Phi(x).$$

Así, por ejemplo $\Phi(-0.67) = 1 - \Phi(0.67) = 1 - 0.7486 = 0.2514$. ■

La densidad normal de parámetros μ y σ^2

La densidad normal puede estar centrada en otro lugar y tener un “ancho” de campana distinto al de la normal estándar. La definición general de la distribución normal es la siguiente.

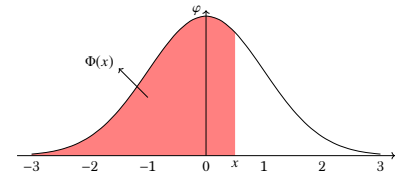


Figura 4: Interpretación gráfica de $\Phi(x)$.

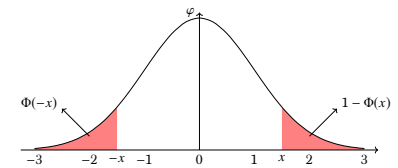


Figura 5: Simetría de la densidad normal estándar

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Figura 6: Tabla de la distribución normal estándar.

Densidad normal general

X tiene distribución normal de parámetros μ y σ^2 si su densidad es igual a

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

Esto lo escribimos $X \sim N(\mu, \sigma^2)$.

La esperanza y la varianza de una variable X con distribución normal $N(\mu, \sigma^2)$ son:

$$E(X) = \mu, \quad \text{y} \quad \text{Var}(X) = \sigma^2.$$

Ambas se pueden probar de forma sencilla observando que, de la fórmula de cambio de variable lineal, vemos que $X = \mu + \sigma Z$ con $Z \sim N(0,1)$.

Recordamos que la definición anterior quiere decir que para todo $a < b$ vale

$$P(a \leq X \leq b) = \int_a^b \varphi_{\mu,\sigma^2}(x) dx.$$

En las figuras que siguen vemos como cambia el gráfico de φ_{μ,σ^2} al variar los parámetros μ y σ .

El parámetro μ es el valor “más probable”, es el centro de simetría de la gráfica de φ_{μ,σ^2} , y los valores de X se concentran entorno a μ . El parámetro σ representa el ancho de la campana, y por lo tanto, cuán

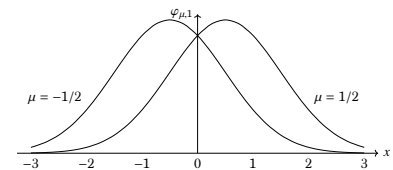


Figura 7: Variando μ

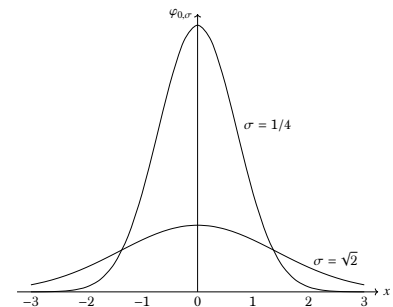


Figura 8: Variando σ

lejos de μ están los valores X . Por ejemplo, la probabilidad

$$P(|X - \mu| \leq k\sigma)$$

está representada en la gráfica de la Figura 9, para $k = 1, 2, 3$.

■ **Ejemplo 3 — Estandarización.** ¿Cómo se calcula una probabilidad para una normal? El método consiste en dos pasos:

- Estandarización: reducción al caso $N(0, 1)$.
- Uso de una tabla de la normal estándar.

Si X tiene distribución normal de parámetros μ y σ y $a < b$, entonces

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right).$$

Como $\frac{X - \mu}{\sigma}$ tiene distribución normal estándar, vemos que

$$P(a \leq X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

Veamos un ejemplo de cálculo. Supongamos que $X \sim N(1.5, 4)$, notar que $\sigma = 2$. Queremos calcular $P(0.16 < X < 6.12)$. Como vimos

$$\begin{aligned} P(0.16 < X < 6.12) &= \Phi\left(\frac{6.12 - 1.5}{2}\right) - \Phi\left(\frac{0.16 - 1.5}{2}\right) \\ &= \Phi(2.31) - \Phi(-0.67) = 0.9896 - 0.2514 = 0.7382. \end{aligned}$$

Por lo tanto $P(0.16 < X < 6.12) = 0.7382$. ■

■ **Ejemplo 4 — Leyendo la tabla al revés.** Por último, observamos que la tabla de la normal estándar se puede leer al revés. La lectura inversa de la tabla consiste en, dado un valor $p \in (0, 1)$ para una probabilidad, queremos hallar el valor x_p tal que $\Phi(x_p) = p$. Para valores que no aparecen en la tabla se usa interpolación lineal.

Por ejemplo, si $p = 0.95$, como $\Phi(1.64) = 0.945$, $\Phi(1.65) = 0.955$, y

$$0.95 = (0.945 + 0.955)/2$$

entonces $x_{0.95} \approx (1.64 + 1.65)/2 = 1.645$. ■

■ **Ejemplo 5 — Aproximando un histograma con la normal.** La Figura 10 el histograma de la longitud de la mano de varios estudiantes hombres del curso de PyE 2017. Sobre el histograma se muestra también la densidad de la distribución normal de parámetros $\mu = 22.6$ y $\sigma = 1.27$, que son el promedio y el desvío estándar respectivamente de la muestra de datos. Notar que la densidad normal aproxima muy bien el histograma.

La tabla de abajo muestra la comparación numérica. Si bien el total de mediciones no es extremadamente grande, son 120 en total, la aproximación es muy buena.

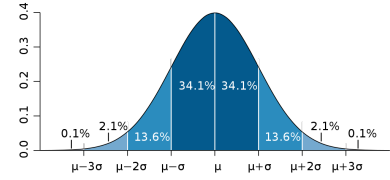


Figura 9: Regla del desvío

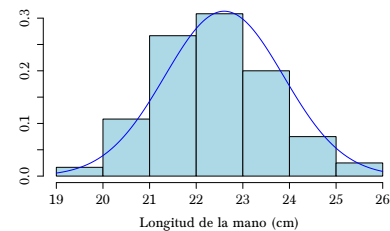


Figura 10: Histograma de la longitud de la mano (en cm) de los estudiantes hombres del curso de PyE 2017.

Intervalo	19-20	20-21	21-22	22-23	23-24	24-25	25-26
Frec. Abs.	2	13	32	37	24	9	3
Frec. Rel.	.017	.108	.267	.308	.200	.075	.025
Normal	.018	.084	0.214	.304	.241	.106	.026

Muchas mediciones de tipo biológico se ajustan bien a la distribución normal, como son la altura, el peso, la presión sanguínea, la temperatura corporal, y muchas otras. ■

Variables normales independientes

Las propiedades más importantes de la distribución normal involucran dos o más variables normales independientes.

Sean X e Y independientes con distribución normal estándar

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad (z \in \mathbb{R}).$$

La densidad conjunta de X e Y está dada por

$$p(x, y) = \varphi(x)\varphi(y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} \quad ((x, y) \in \mathbb{R}^2).$$

La propiedad clave de esta densidad es que es una función de $r^2 = x^2 + y^2$, en donde r es la distancia del punto (x, y) al origen. Esto hace que la gráfica de la función $p(x, y)$ se obtenga haciendo girar la campana de Gauss sobre el eje z (las secciones son de hecho proporcionales a la campana de Gauss, ver la Figura 11).

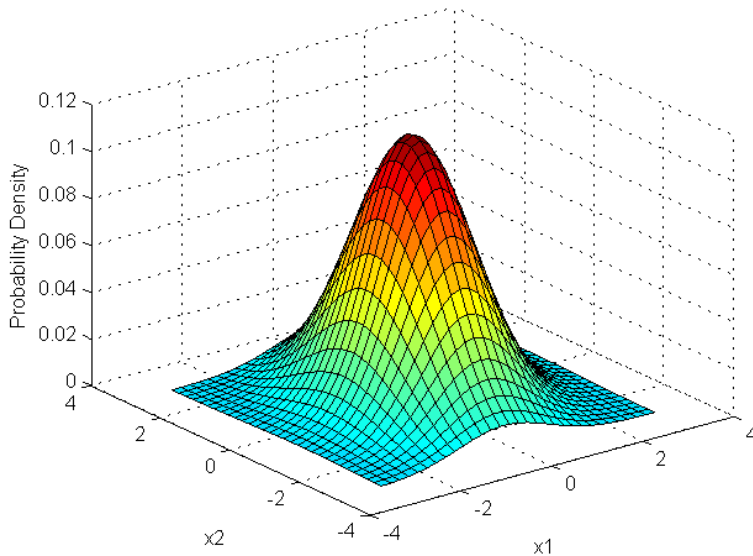
La simetría rotacional de esta densidad bi-variada, obtenida a partir de dos variables normales independientes, es una propiedad muy especial. Como vimos en el Ejemplo 1, esta propiedad distingue a la normal entre todas aquellas distribuciones bi-variadas obtenidas a partir de dos variables independientes.

Esta simetría rotacional es clave para entender muchas de las propiedades importantes de la distribución normal que estudiaremos a partir de ahora.

■ **Ejemplo 6 — Otra vez tirando al blanco.** Un experto tirador tira al blanco, y el tiro tiene coordenadas (X, Y) que se distribuye como dos normales estándar independientes. La probabilidad de que el tiro caiga en el interior del círculo de radio r centrado en el origen es $1/2$. ¿Cuál es la probabilidad de que el tiro caiga en el interior del círculo de radio $2r$ centrado en el origen?

Debemos calcular el radio r . Como en el ejemplo anterior, consideremos la variable $R^2 = X^2 + Y^2$. Un cálculo directo muestra que

$$P(R \leq r) = 1 - e^{-r^2/2}.$$

Figura 11: Gráfico de la densidad conjunta $p(x, y)$.

Entonces r debe ser solución de $1 - e^{-r^2/2} = 1/2$, o lo que es lo mismo $e^{-r^2/2} = 1/2$. Luego $r = \sqrt{2 \ln(2)} \approx 1.77$.

Cambiando r por $2r$ tenemos que

$$P(R \leq 2r) = 1 - e^{-(2r)^2/2} = 1 - e^{-2r^2} = 15/16 \approx 0.9375.$$

¿Cuál es la distancia promedio de los tiros al origen?

$$\begin{aligned} E(R) &= \int_0^{\infty} r p_R(r) dr = \int_0^{\infty} r^2 e^{-r^2/2} dr = \frac{1}{2} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2} dx \\ &= \frac{\sqrt{2\pi}}{2} \int_{-\infty}^{+\infty} x^2 \varphi(x) dx = \sqrt{\frac{\pi}{2}} \approx 1.253. \end{aligned}$$

■

Combinaciones lineales de normales independientes

Combinaciones lineales de variables normales independientes son siempre normales. Este hecho importante es otra consecuencia de la simetría rotacional de la distribución conjunta de variables independientes X e Y con distribución normal estándar.

Para ver esto, sea X_θ la primer coordenada de (X, Y) relativa a un nuevo sistema de ejes que forma un ángulo θ con los ejes originales.

De la Figura 12 vemos que $X_\theta = X \cos \theta + Y \sin \theta$. Pero en vista de la simetría rotacional de la densidad conjunta de (X, Y) , la distribución de X_θ debe ser la misma que la distribución de X , que es normal estándar, cualquiera sea el ángulo θ . Entonces $X_\theta \sim N(0, 1)$.

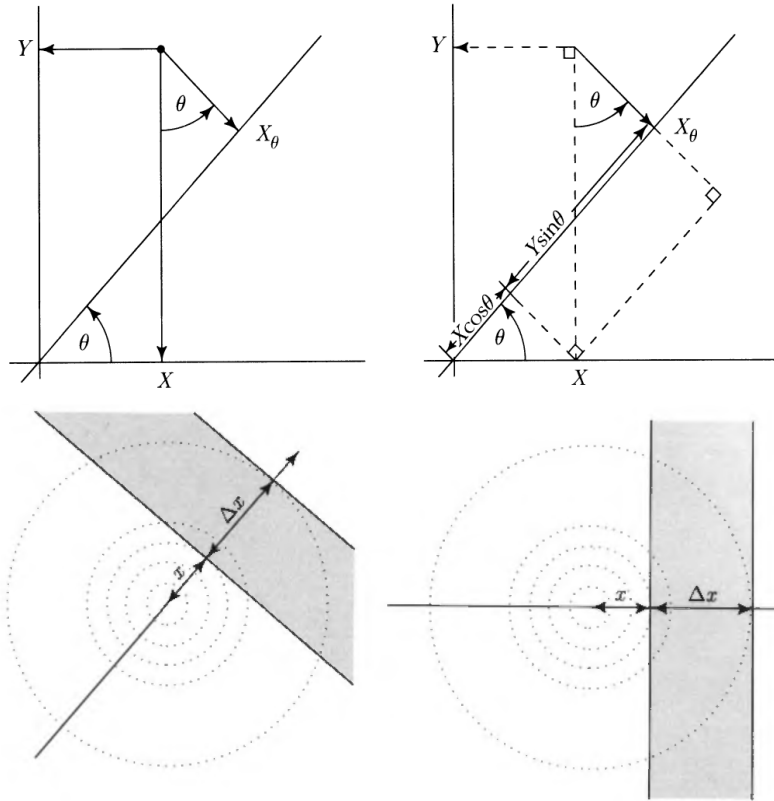


Figura 12: Proyección X_θ sobre el eje de ángulo θ con X .

Por ejemplo, el evento $x \leq X_\theta \leq x + \Delta x$ corresponde a que el punto (X, Y) caiga en la banda que se muestra en la Figura 12. Es decir que

$$P(x \leq X_\theta \leq x + \Delta x) = P(x \leq X \leq x + \Delta x).$$

Como $\cos \theta$ y $\sin \theta$ pueden ser dos números cualesquiera α y β con $\alpha^2 + \beta^2 = 1$, la simetría rotacional de (X, Y) implica: si X e Y son normales estándar independientes, entonces $\alpha X + \beta Y$ tiene distribución normal estándar para todo α y β con $\alpha^2 + \beta^2 = 1$.

Tomando $\alpha = \beta = 1/\sqrt{2}$, vemos que si X e Y son normales estándar independientes, entonces $(X + Y)/\sqrt{2}$ es normal estándar. Recordar que si Z es normal estándar, entonces $\sigma Z \sim N(0, \sigma^2)$, por lo que si X e Y son normales estándar, entonces $X + Y \sim N(0, 2)$.

Este argumento se extiende a sumas más generales. Si $X \sim N(\mu, \sigma^2)$ e $Y \sim N(\lambda, \tau^2)$ son independientes, entonces

$$X + Y \sim N(\mu + \lambda, \sigma^2 + \tau^2).$$

Para probarlo, observar que

$$U = \frac{X - \mu}{\sigma} \sim N(0, 1), \quad V = \frac{Y - \lambda}{\tau} \sim N(0, 1),$$

y son independientes. Tomando

$$\alpha^2 = \frac{\sigma^2}{\sigma^2 + \tau^2}, \quad \beta^2 = \frac{\tau^2}{\sigma^2 + \tau^2},$$

tenemos que $W = \alpha U + \beta V \sim N(0, 1)$ ya que $\alpha^2 + \beta^2 = 1$. Pero podemos ver que

$$W = \frac{X + Y - (\mu + \lambda)}{\sqrt{\sigma^2 + \tau^2}}$$

de donde $X + Y \sim N(\mu + \lambda, \sigma^2 + \tau^2)$.

Por inducción se demuestra que:

Combinación lineal de normales independientes

Si X_1, \dots, X_n son independientes con $X_i \sim N(\mu_i, \sigma_i^2)$, entonces

$$X_1 + \dots + X_n \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Un error frecuente es decir que si X e Y son normales estándar independientes, entonces $X - Y \sim N(1 - 1, 1 - 1) = N(0, 0)$. Esto está rotundamente mal. Notar que $X - Y = X + (-Y)$ y $-Y$ también tiene distribución normal estándar. La distribución de $X - Y$ es $N(0, 2)$.

Apéndice: La deducción original de Gauss

Si bien la historia de su descubrimiento es enormemente rica en anécdotas, nos centraremos en la forma original con la cual Gauss derivó una fórmula para la densidad de probabilidad de esta distribución.

Gauss estaba interesado en el problema de estimar el valor real de una cierta cantidad medible a partir de observaciones empíricas. Específicamente, imaginemos que el valor real, y desconocido, de una cierta cantidad medible es μ , y que disponemos de las observaciones

$$M_1, M_2, \dots, M_n.$$

El objetivo es estimar μ a partir de estas observaciones. La idea es la siguiente: supongamos que nuestras mediciones empíricas son de la forma

$$M_i = \mu + E_i,$$

en donde E_i es el error, que supondremos aleatorio, que cometemos en la i -ésima medición. Gauss se propuso derivar una fórmula para la densidad de la distribución de los errores, un problema conocido en la época como el problema de *la curva de error*. Sus criterios para realizar esto se basaron en supuestos sobre las propiedades generales de los errores:

1. Los errores pequeños son más comunes que los errores grandes.
2. Es igual de probable cometer un error positivo que uno negativo de igual magnitud.
3. Los errores cometidos en mediciones distintas son independientes.

Si llamamos de $\varphi(\epsilon)$ a la densidad de probabilidad del error, estos supuestos se traducen en:

1. Si $0 < \epsilon_1 < \epsilon_2$, entonces $\varphi(\epsilon_1) \geq \varphi(\epsilon_2)$.
2. Para todo ϵ , se tiene $\varphi(\epsilon) = \varphi(-\epsilon)$.
3. Las variables E_1, \dots, E_n son independientes.

Existen muchas curvas de error φ posibles que cumplan estos supuestos. Algunas de ellas se muestran en las Figuras 13, 14, y 15. Sin embargo, Gauss supuso un cuarto ingrediente fundamental:

En presencia de varias mediciones de la misma magnitud, el valor más probable de la cantidad que se mide es su promedio.

Con este último ingrediente Gauss fue capaz de derivar una fórmula para φ .

Gauss asumió que la densidad φ era derivable, cosa que por supuesto también haremos nosotros. Consideremos la cantidad

$$f(\epsilon) = \frac{\varphi'(\epsilon)}{\varphi(\epsilon)},$$

en donde φ' es la derivada de φ . Si derivamos la igualdad $\varphi(\epsilon) = \varphi(-\epsilon)$ dada en el supuesto 2, obtenemos $\varphi'(\epsilon) = -\varphi'(-\epsilon)$, de donde

$$f(\epsilon) = \frac{\varphi'(\epsilon)}{\varphi(\epsilon)} = \frac{-\varphi'(-\epsilon)}{\varphi(-\epsilon)} = -f(-\epsilon).$$

En otras palabras, $f(-\epsilon) = -f(\epsilon)$.

El supuesto 3 sobre la independencia de los errores en las distintas mediciones, implica que la probabilidad de observar

- el error E_1 en un intervalo I_1 de tamaño $\Delta\epsilon_1$ centrado en ϵ_1 ,
- el error E_2 en un intervalo I_2 de tamaño $\Delta\epsilon_2$ centrado en ϵ_2 ,
- ⋮
- el error E_n en un intervalo I_n de tamaño $\Delta\epsilon_n$ centrado en ϵ_n ,

esta dada por (ver la Figura 16)

$$P(E_1 \in I_1, \dots, E_n \in I_n) = \prod_{i=1}^n P(E_i \in I_i) \approx \prod_{i=1}^n \varphi(\epsilon_i) \Delta\epsilon_i.$$

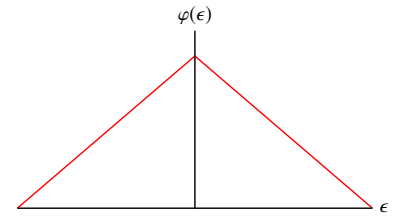


Figura 13: Curva de error triangular de Simpson. En este caso $\varphi(\epsilon) = -k^2\epsilon + k$, con k una constante positiva.

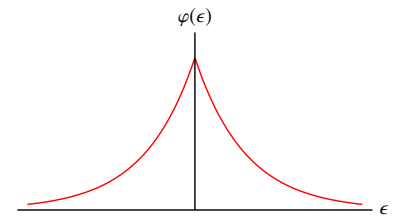


Figura 14: Curva de error exponencial de Laplace. En este caso $\varphi(\epsilon) = \frac{k}{2}e^{-k|\epsilon|}$, con k una constante positiva.

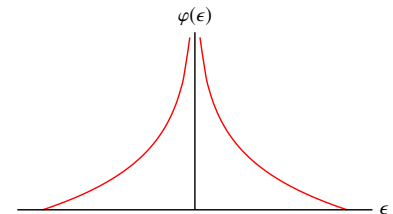


Figura 15: Curva de error logarítmica de Laplace. En este caso $\varphi(\epsilon) = \frac{1}{2k} \ln\left(\frac{k}{|\epsilon|}\right)$, con k una constante positiva.

Fijemos de ahora en más la longitud de los intervalos I_1, \dots, I_n como siendo todas iguales a $\Delta\epsilon$, suficientemente pequeña para que la aproximación en la expresión anterior sea válida. Lo importante aquí es simplemente fijar este valor para que no intervenga en los argumentos que haremos a continuación. De este modo, la probabilidad será máxima cuando el producto $\prod_{i=1}^n \varphi(\epsilon_i)$ sea máximo.

Observar que podemos escribir los errores en función de las mediciones y del valor desconocido μ , como $E_i = M_i - \mu$. El cuarto y último supuesto de Gauss, se traduce entonces en que el valor más probable para μ , esto es el que maximiza el producto

$$F(\mu) = \prod_{i=1}^n \varphi(M_i - \mu),$$

es el promedio

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

de las n mediciones. Esta condición implica que si derivamos F y la evaluamos en $\mu = \bar{M}$ el resultado debe ser cero.

Calcular la derivada de F no es difícil pero sí un poco largo, así que lo dejamos como ejercicio. Al derivar, obtenemos

$$\frac{dF}{d\mu}(\mu) = \left(f(M_1 - \mu) + \dots + f(M_n - \mu) \right) F(\mu).$$

Entonces, si sustituimos $\mu = \bar{M}$ e igualamos a cero, obtenemos la ecuación

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

Esta ecuación debe ser satisfecha cualesquiera sean las mediciones, y estas pueden ser cualquier valor real. En particular, si

$$M_1 = \alpha, \text{ y } M_2 = \dots = M_n = \beta,$$

entonces la condición anterior implica

$$f\left((n-1)\left(\frac{\alpha-\beta}{n}\right)\right) = (n-1)f\left(\frac{-\alpha+\beta}{n}\right) = -(n-1)f\left(\frac{\alpha-\beta}{n}\right).$$

Si llamamos $x = (\alpha - \beta)/n$, que puede ser cualquier real eligiendo adecuadamente α y β ; y llamamos $m = n - 1$ que puede ser cualquier entero, llegamos a la conclusión de que la función f debe cumplir la ecuación

$$f(mx) = mf(x),$$

para todo real x y todo entero positivo m . Se puede ver que las únicas funciones que cumplen con esta condición son las lineales: es decir $f(x) = -kx$.

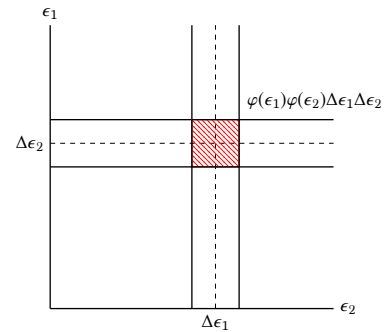


Figura 16: La probabilidad de que los errores E_1 y E_2 caigan en intervalos centrados en ϵ_1 y ϵ_2 y de longitud $\Delta\epsilon_1$ y $\Delta\epsilon_2$.

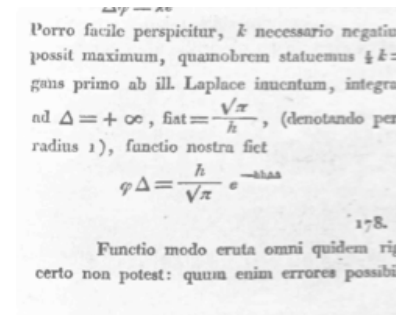


Figura 17: Extracto del artículo original de Gauss en donde aparece la fórmula de la campana.

Observemos a lo que hemos llegado: hemos probado que existe una constante $k > 0$ tal que

$$\frac{\varphi'(\epsilon)}{\varphi(\epsilon)} = -k\epsilon.$$

Si integramos en ambos lados de esta igualdad, luego de algunos cálculos, concluimos que

$$\varphi(\epsilon) = Ce^{-k\epsilon^2},$$

en donde C es una constante de integración. Esta es la fórmula mágica a la cual Gauss llegó y que conocemos hoy como densidad de la distribución normal. Ver la Figura 17.