

Teoría – Tema 10

Distribución normal o gaussiana

Índice de contenido

¿Por qué se llama “normal”?.....	2
Características de la función de densidad normal.....	3
Distribución normal estándar y tipificación de la variable aleatoria.....	4
Cálculo de probabilidades.....	5

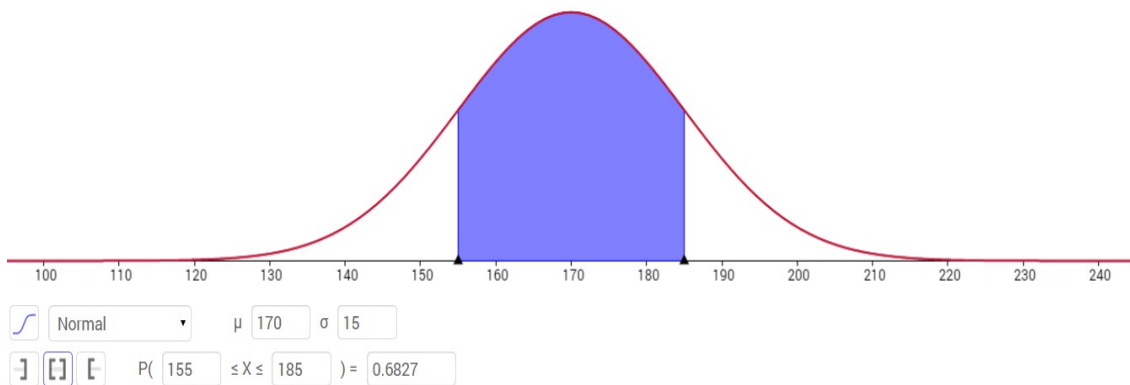
¿Por qué se llama "normal"?

Existen muchas variables aleatorias continuas que siguen con cierta fidelidad la curva de la distribución gaussiana. De ahí su nombre de "normal", ya que se llegó a pensar que casi todas las variables aleatorias seguirían este caso particular de distribución.

Fue propuesta por Moivre en el s. XVIII (el mismo de la fórmula de complejos), pero a la historia ha pasado con el sobrenombre de distribución gaussiana por el uso sistemático en los errores de medida de datos astronómicos realizados por Gauss.

----- Distribución normal con media 170 y desviación 15

$\mu = 170$ $\sigma = 15$



Características de la función de densidad normal

Una variable aleatoria que sigue una distribución normal de media μ y desviación típica σ se designa por $N(\mu, \sigma)$ si se cumplen las siguientes condiciones.

- La variable x puede tomar cualquier valor real $(-\infty, +\infty)$.

- La función de densidad es $f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}$.

El dominio de esta función es toda la recta real.

La función es simétrica respecto a la recta vertical $x = \mu$. Este eje de simetría divide a la gráfica en dos partes iguales.

La gráfica no corta al eje horizontal OX, ya que la recta $y = 0$ es una asíntota horizontal de $f(x)$.

El punto de corte con el eje vertical OY es $(0, f(0))$, con $f(0) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} e^{-\frac{1}{2} \left(\frac{\mu}{\sigma}\right)^2}$.

La función es creciente en el intervalo $(-\infty, \mu)$ y decreciente en $(\mu, +\infty)$. El máximo se alcanza en $x = \mu$. El valor de la función en el punto máximo es $f(\mu) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}}$.

Los puntos de inflexión, donde cambia la curvatura de la gráfica, acontecen en los valores $x = \mu - \sigma$ y $x = \mu + \sigma$. El valor de la función en los puntos de inflexión es

$$f(\sigma) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \frac{1}{\sqrt{e}}$$

El área encerrada entre la curva gaussiana y el eje horizontal OX es la unidad cuadrada (como toda función de densidad). Al ser simétrica respecto a la recta vertical $x = \mu$, el área se divide en dos mitades iguales de $0,5 u^2$ a la izquierda y a la derecha del valor medio.

Distribución normal estándar y tipificación de la variable aleatoria

La pareja de valores (μ, σ) puede tomar infinitos valores distintos. Por lo que tendremos infinitas curvas de distribución normal distintas. Para facilitar los cálculos y poder comparar de manera cómoda procesos aleatorios entre sí, es común transformar cualquier distribución normal $N(\mu, \sigma)$ a otra equivalente con media 0 y desviación típica 1. Es decir, $N(0, 1)$:

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2} \rightarrow \text{transformación} \rightarrow f(z) = \frac{1}{\sqrt{2 \cdot \pi}} e^{-\frac{1}{2} \cdot z^2} \rightarrow \text{nueva variable } z$$

Los valores que toma $N(0, 1)$ en toda la recta real están **tabulados**, por lo que agiliza el cálculo y el manejo de los datos.

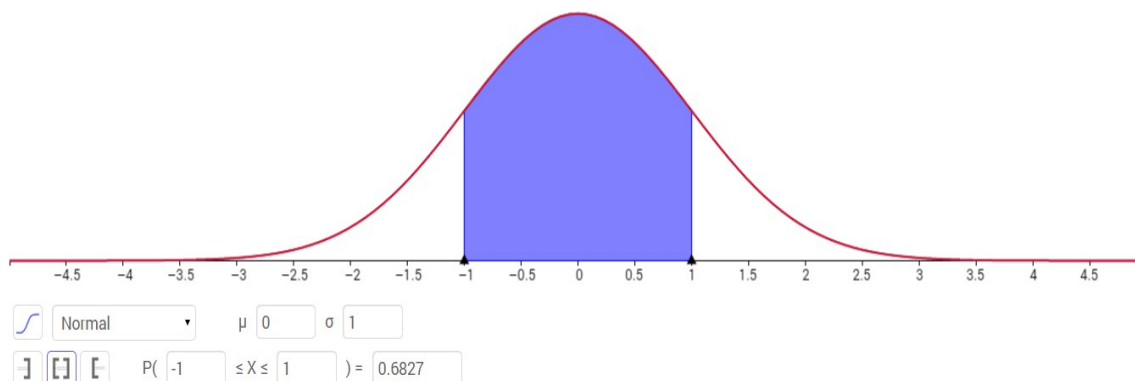
Esta transformación desplaza el valor medio desde μ hasta 0. Es decir, centra la curva gaussiana en el eje vertical vertical ($z=0$). Además convierte la desviación típica σ en 1, por lo que ensancha o contrae la gráfica de la distribución.

Esta doble transformación de parámetros se conoce como **tipificación de la variable aleatoria**, y se consigue con el siguiente cambio de variable:

$$z = \frac{x - \mu_x}{\sigma_x} \rightarrow \mu_z = \frac{\mu_x - \mu_x}{\sigma_x} = 0, \quad \sigma_z = \frac{\mu_x + \sigma_x - \mu_x}{\sigma_x} = 1 \rightarrow N(0, 1)$$

Distribución normal tipificada con media 0 y desviación 1

$\mu = 0 \quad \sigma = 1$



■ Cálculo de probabilidades

Tradicionalmente se han utilizado tablas para obtener las distintas probabilidades de la curva normal $N(0,1)$. Hoy día, con las nuevas tecnologías, tenemos varios recursos online que nos ofrecen estos valores de manera más cómoda. Por ejemplo, la **herramienta estadística de GeoGebra**:

<http://web.geogebra.org/app>

Tanto si usamos tablas como herramientas online, debemos entender qué significan los valores de probabilidad que nos ofrecen.

La probabilidad coincide con el área encerrada entre la curva y el eje horizontal. Por lo tanto, la probabilidad de obtener un valor igual o inferior al valor medio 0 es 50%.

$$P(z \leq 0) = 0,5 \rightarrow \text{área } 0,5 \text{ multiplicada por } 100$$

Y la probabilidad de obtener un valor igual o superior al valor medio 0 es 50%.

$$P(z \geq 0) = 0,5$$

Es decir, en la distribución normal **el valor medio coincide con la mediana**.

La **probabilidad de obtener un valor inferior o igual a uno dado** z_{max} será $P(z \leq z_{max})$. Y **las tablas suelen ofrecer este valor** $P(z \leq z_{max})$, con $z_{max} > 0$.

Es decir, no nos vamos a preguntar por la probabilidad de obtener un valor concreto sino por la probabilidad de obtener un **intervalo de valores**.

De hecho, por propia definición de la función de densidad, la probabilidad de obtener un valor concreto es 0 . Lo cual tiene lógica si pensamos que sería la probabilidad de obtener un valor real concreto elegido entre los infinitos valores reales comprendidos en $(-\infty, +\infty)$ (variable continua aleatoria).

Para valores negativos a la izquierda de la media, es decir $-z_{max} < 0$, el valor $P(z \leq -z_{max})$ se calcula recordando que la distribución normal es simétrica respecto la media y que la probabilidad total es igual a 1 (área encerrada por la curva a lo largo de toda la recta real). Es decir:

$$P(z \leq -z_{max}) = P(z \geq z_{max}) = 1 - P(z \leq z_{max})$$

La probabilidad de obtener un valor acotado entre un valor mínimo $z=z_{min}$ y un valor máximo $z=z_{max}$ se expresa:

$$P(z_{min} \leq z \leq z_{max}) = P(z \leq z_{max}) - P(z \leq z_{min})$$

¿Y si deseamos saber qué intervalo corresponde a una probabilidad fijada p ? Es decir, ahora el valor desconocido no es la probabilidad sino el valor z_{max} .

Si la probabilidad p es mayor o igual a 0,5 obtenemos el valor z_{max} positivo mirando directamente las tablas.

Y si la probabilidad p es menor a 0,5 significa que buscamos un valor $-z_{max}$ negativo. En este caso podremos usar las tablas si nos preguntamos por la probabilidad acumulada de su valor opuesto: $1-p = P(z \leq z_{max}) \rightarrow$ y el valor obtenido z_{max} será el opuesto al que buscábamos inicialmente.