Probability

Normal Distribution

Prof. Dr. Mostafa Elhosseini

https://youtube.com/drmelhosseini

Normal Distribution

The normal distribution (also called the **Gaussian distribution**) is by far the **most commonly used distribution** in statistics.

- Heights, weights, and other physical characteristics
- Measurement errors in scientific experiments
- **Even** when the underlying distribution is **discrete**, the **normal curve** often gives an **excellent approximation**

Normal Distribution

- **De Moivre** presented this fundamental result, known as the **central limit theorem**, in 1733.
- Unfortunately, his work was **lost** for some time, and **Gauss** independently developed a normal distribution nearly 100 years later.
- Although De Moivre was later credited with the derivation, a **normal distribution** is also referred to as a **Gaussian distribution**

The German-born Carl Friedrich Gauss (1777–1855), shown here on a former German 10-Mark bill, was one of the first to use the normal distribution, with curve shown to his right.

Normal Distribution

A continuous random variable X is said to have a **normal distribution** with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, where $-\infty < \boldsymbol{\mu} < \infty$ and $\mathbf{0} < \boldsymbol{\sigma}$, if the *pdf* of X is

$$\boldsymbol{f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-{(x-\mu)^2}/{(2\sigma^2)}}} \quad -\infty < x < \infty$$

- The statement that $\boldsymbol{X}$ is normally distributed with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$ is often abbreviated $\boldsymbol{X \sim N(\mu, \sigma^2)}$

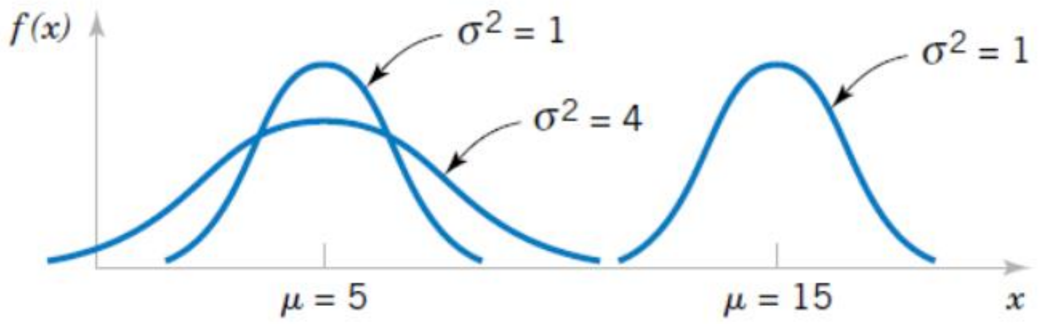
Normal Distribution

$$\boldsymbol{f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-{(x-\mu)^2}/{(2\sigma^2)}}} \qquad -\infty < x < \infty$$

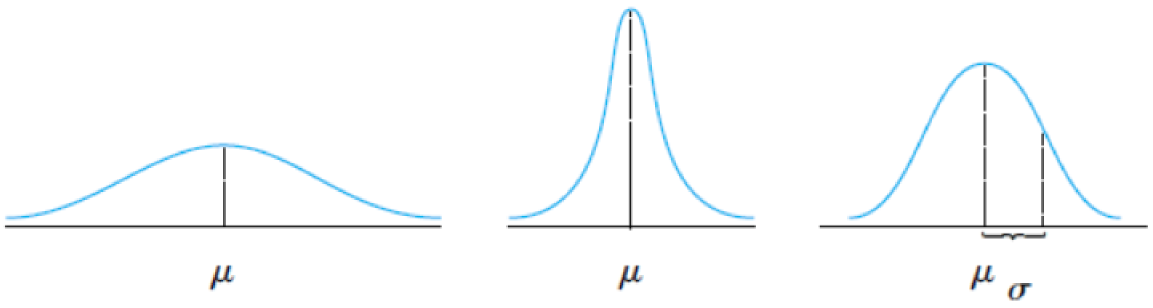
- $f(x;\mu,\sigma) \geq 0$
- $\int_{-\infty}^{\infty} f(x;\mu,\sigma)\,dx = 1$
- $E(X) = \mu$ and
- $V(X) = \sigma^2$,

Normal Distribution

- Each density curve is **symmetric** about $\boldsymbol{\mu}$ and **bell-shaped**, so the **center** of the bell (point of symmetry) is both the **mean** of the distribution and the **median**.
- The value of $\boldsymbol{\sigma}$ is the **distance from $\boldsymbol{\mu}$ to the inflection points** of the curve (the points at which the curve changes from turning downward to turning upward)

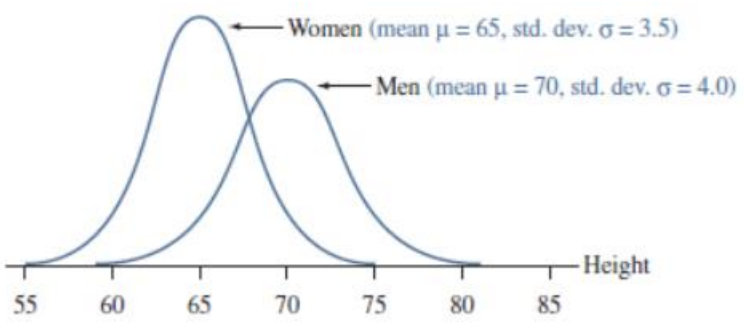


Normal Distribution

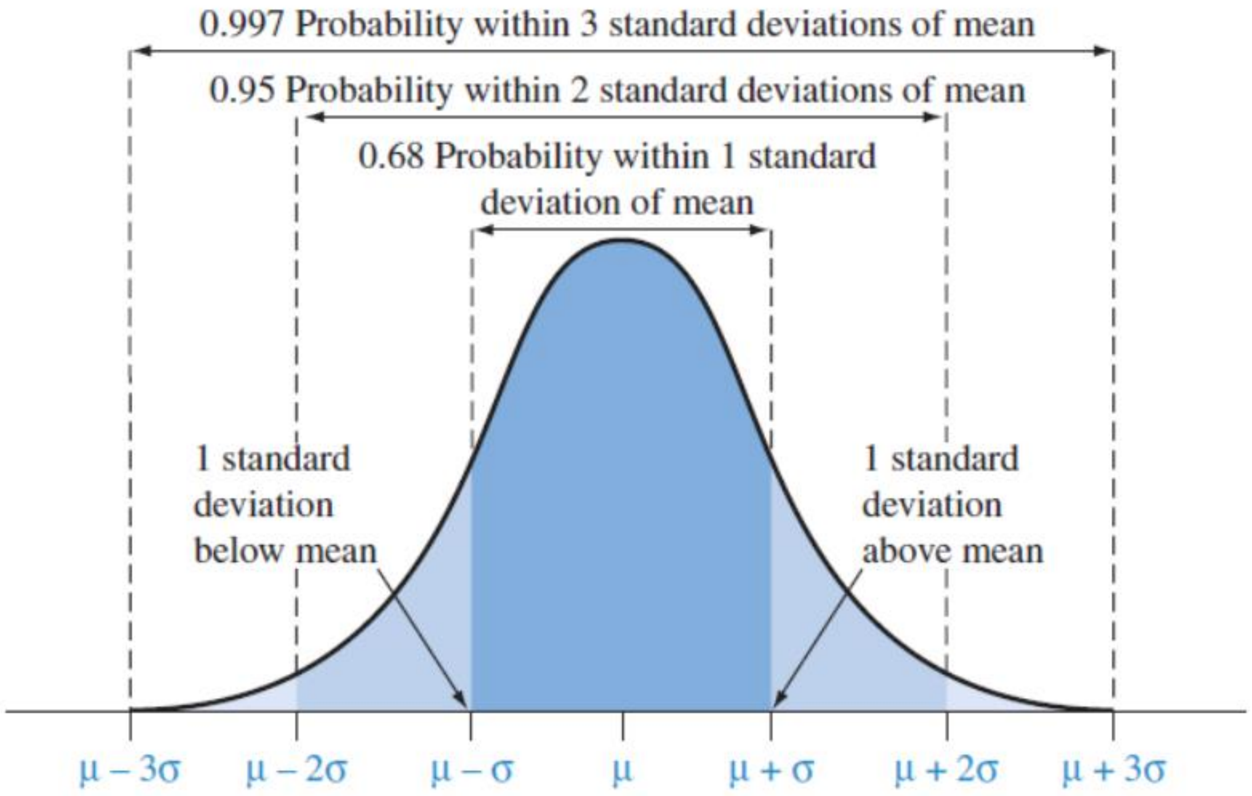


- Large values of $\boldsymbol{\sigma}$ yield graphs that are quite spread out about $\boldsymbol{\mu}$, whereas small values of $\boldsymbol{\sigma}$ yield graphs with a high peak above $\boldsymbol{\mu}$ and most of the area under the graph quite close to $\boldsymbol{\mu}$.
- Thus a large $\boldsymbol{\sigma}$ implies that a value of X far from $\boldsymbol{\mu}$ may well be observed, whereas such a value is quite unlikely when $\boldsymbol{\sigma}$ is small

Normal Distribution

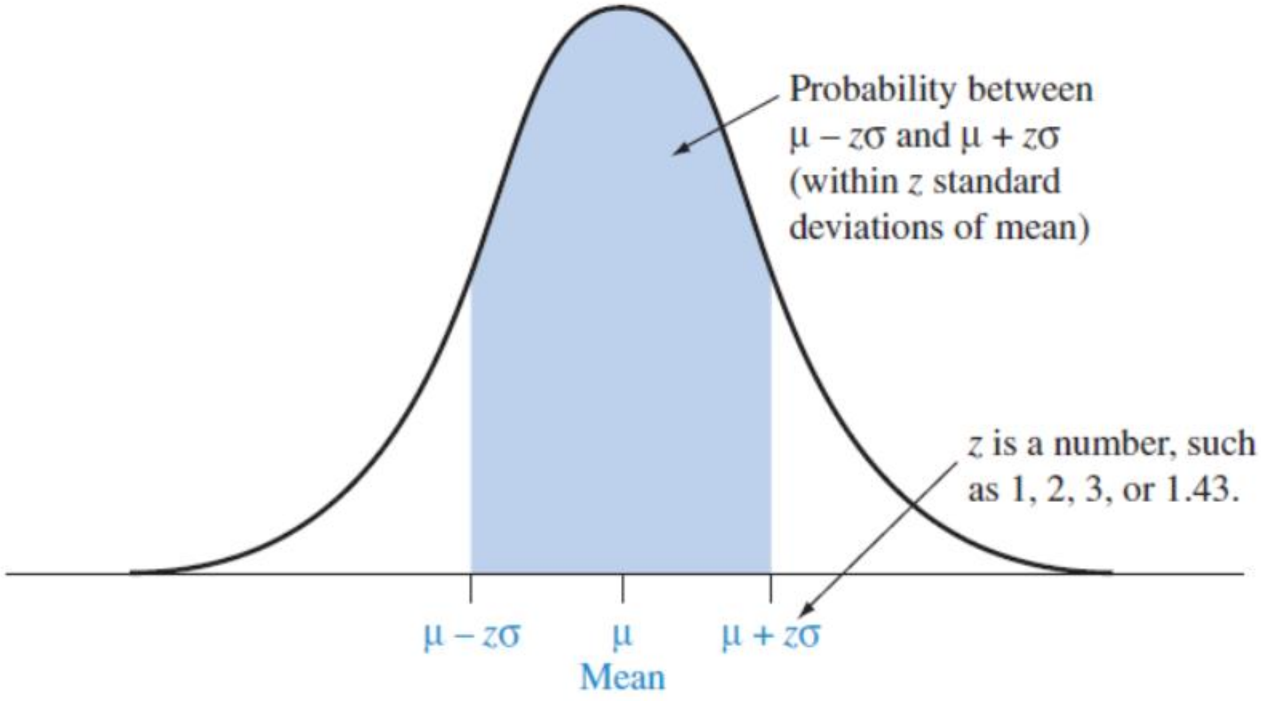


Normal Distribution



- About 68% of the population is in the interval $\mu \pm \sigma$.
- About 95% of the population is in the interval $\mu \pm 2\sigma$.
- About 99.7% of the population is in the interval $\mu \pm 3\sigma$

Normal Distribution



- The multiples 1, 2, and 3 of the number of standard deviations from the mean are denoted by the symbol z in general. For instance, $z = 2$ for 2 standard deviations. For each fixed number z, the probability within z standard deviations of the mean is the area under the normal curve between $\mu - z\sigma$ and $\mu + z\sigma$

Normal Distribution

The **normal distribution** is symmetric, bell-shaped, and characterized by its mean μ (expressing the center) and standard deviation σ (expressing the variability). The probability of observing values within any given number of standard deviations from the mean μ is the same for all normal distributions (i.e., for any μ and any $\sigma > 0$). This probability equals 0.68 for the interval within 1 standard deviation of the mean, 0.95 for the interval within 2 standard deviations of the mean, and 0.997 for the interval within 3 standard deviations of the mean. See Figure 6.5.

Normal Distribution

- The **proportion** of a normal population that is **within a given number of standard deviations of the mean** is the same for any normal population.
- For this reason, when dealing with normal populations, we often **convert from** the units in which the population items were originally measured **to standard units**.
- **Standard units** tell how many standard deviations an observation is from the population mean

Normal Distribution

- To compute $\boldsymbol{P(a \leq X \leq b)}$ when X is a normal random variable with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, we must determine

$$\int_a^b \frac{\mathbf{1}}{\sqrt{\mathbf{2}\boldsymbol{\pi}\boldsymbol{\sigma}}} \boldsymbol{e}^{-\boldsymbol{(x-\mu)^2}/\boldsymbol{(2\sigma^2)}}\, dx$$

- None of the standard integration techniques can be used to evaluate the above Expression.
- Instead, for $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\sigma} = \mathbf{1}$, the above Expression has been calculated using numerical techniques and tabulated for certain values of $\boldsymbol{a}$ and $\boldsymbol{b}$

Standard Normal Distribution

- The normal distribution with parameter values $\mu = 0$and $\sigma = 1$ is called the **standard normal distribution**.
- A random variable having a standard normal distribution is called a standard normal random variable and will be denoted by **Z**.
- The pdf of Z is

$$\boldsymbol{f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}} \quad -\infty < z < \infty$$

- The **standard normal distribution** does **not** frequently **serve** as a model for a naturally arising population. Instead, it is a **reference distribution** from which information about other normal distributions can be **obtained**

Standard Normal Distribution

- The normal distribution with parameter values $\mu = 0$and $\sigma = 1$is called the **standard normal distribution**.
- A random variable having a standard normal distribution is called a standard normal random variable and will be denoted by **Z**.
- The pdf of Z is

$$\boldsymbol{f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}} \quad -\infty < z < \infty$$

- The **standard normal distribution** does **not** frequently **serve** as a model for a naturally arising population. Instead, it is a **reference distribution** from which information about other normal distributions can be **obtained**

The Standard Normal Distribution

To compute $P(a \leq X \leq b)$ when X is a normal rv with parameters μ and σ, we must determine

$$\int_a^b \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/(2\sigma^2)}\,dx \tag{4.4}$$

None of the standard integration techniques can be used to evaluate Expression (4.4). Instead, for $\mu = 0$ and $\sigma = 1$, Expression (4.4) has been calculated using numerical techniques and tabulated for certain values of a and b. This table can also be used to compute probabilities for any other values of μ and σ under consideration.

A normal random variable with

$$\mu = 0 \quad \text{and} \quad \sigma^2 = 1$$

is called a **standard normal random variable** and is denoted as Z.

The cumulative distribution function of a standard normal random variable is denoted as

$$\Phi(z) = P(Z \leq z)$$