The Normal Distribution - Fall 2024

Instructor: Dr. Juhee Lee
TA: Antonio Aguirre

University of California, Santa Cruz
Fall 2024

1. Normal Distribution

The normal distribution is one of the most fundamental probability distributions. It is defined by its **mean** (μ) and **variance** (σ^2). Its probability density function (PDF) is:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

Key features:

– It is symmetric around μ (the center).

– The **spread** is determined by σ^2 (variance).

– The standard normal distribution has $\mu = 0$ and $\sigma^2 = 1$.

Example:

If $X \sim \mathcal{N}(5, 4)$ ($\mu = 5$, $\sigma^2 = 4$), then the PDF is:

$$f_X(x) = \frac{1}{\sqrt{8\pi}} \exp\left(-\frac{(x-5)^2}{8}\right).$$

—

2. Properties and Operations for Normal Probabilities

2.1. Symmetry

The normal distribution is symmetric about its mean (μ):

$$\Pr(X \le \mu - a) = \Pr(X \ge \mu + a).$$

Example:

If $X \sim \mathcal{N}(10, 16)$, then:

$$\Pr(X \leq 6) = \Pr(X \geq 14).$$

2.2. Complement Rule

For any event:

$$\Pr(X \geq x) = 1 - \Pr(X \leq x).$$

Example:

If $\Pr(X \leq 12) = 0.8$, then:

$$\Pr(X \geq 12) = 1 - 0.8 = 0.2.$$

2.3. Probability Between Two Values

The probability that X falls between two values a and b is:

$$\Pr(a \leq X \leq b) = \Pr(X \leq b) - \Pr(X \leq a).$$

Example:

If $\Pr(X \leq 15) = 0.9$ and $\Pr(X \leq 10) = 0.6$, then:

$$\Pr(10 \leq X \leq 15) = 0.9 - 0.6 = 0.3.$$

2.4. Standardization

For $X \sim \mathcal{N}(\mu, \sigma^2)$, the standardization:

$$Z = \frac{X - \mu}{\sigma},$$

transforms X into $Z \sim \mathcal{N}(0, 1)$.

Example:

If $X \sim \mathcal{N}(20, 25)$ and $x = 30$, the standardized value is:

$$Z = \frac{30 - 20}{\sqrt{25}} = 2.$$

—

3. Linear Transformation of a Normal Distribution

Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ and consider a linear transformation:

$$Y = aX + b,$$

where $a \neq 0$ and b are constants. The transformed random variable Y is also normally distributed:

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2).$$

Example:

If $X \sim \mathcal{N}(10, 4)$, and $Y = 2X + 3$, then:

$$Y \sim \mathcal{N}(2 \cdot 10 + 3, (2^2) \cdot 4) = \mathcal{N}(23, 16).$$

—

4. Linear Combination of Independent Normals

If $X_1, X_2, \ldots, X_n$ are independent random variables, where $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, and we define:

$$Y = \sum_{i=1}^{n} a_i X_i,$$

then Y is also normally distributed:

$$Y \sim \mathcal{N}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right).$$

Example:

If $X_1 \sim \mathcal{N}(2, 1)$ and $X_2 \sim \mathcal{N}(3, 4)$, and $Y = X_1 + 2X_2$, then:

$$Y \sim \mathcal{N}(2 + 2 \cdot 3, 1 + 4 \cdot 4) = \mathcal{N}(8, 17).$$

—

5. Sample Mean Distribution

For $X_1, X_2, \ldots, X_n$ i.i.d. random variables:

$$X_i \sim \mathcal{N}(\mu, \sigma^2),$$

the **sample mean** is defined as:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Since the X_i's are normal, the sample mean $\bar{X}_n$ is also normally distributed:

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Example:

If $X_i \sim \mathcal{N}(50, 25)$ and $n = 10$, then the sample mean:

$$\bar{X}_n \sim \mathcal{N}(50, \frac{25}{10}) = \mathcal{N}(50, 2.5).$$

—

6. Determining Minimum Sample Size

Problem

We want the minimum sample size n such that:

$$P(|\bar{X}_n - \mu| \leq E) \geq p.$$

Solution

1. Rewrite the event:

$$P(|\bar{X}_n - \mu| \leq E) = 2\Phi(z) - 1,$$

where:

$$z = \frac{E\sqrt{n}}{\sigma}.$$

2. To satisfy $P(|\bar{X}_n - \mu| \leq E) \geq p$:

$$2\Phi(z) - 1 \geq p \implies \Phi(z) \geq \frac{p+1}{2}.$$

3. Let $z_p = \Phi^{-1}\left(\frac{p+1}{2}\right)$. Then:

$$z = z_p \implies \sqrt{n} = \frac{z_p\sigma}{E}.$$

4. Solve for n:

$$n = \left(\frac{z_p\sigma}{E}\right)^2.$$

Example

For $\sigma^2 = 9$ ($\sigma = 3$), $E = 1$, and $p = 0.95$:

- $z_p = \Phi^{-1}(0.975) \approx 1.96$,
- $n = \left(\frac{1.96 \cdot 3}{1}\right)^2 = (5.88)^2 = 34.5744$,
- Minimum $n = 35$ (rounded up).