



4.2.3 Normal (Gaussian) Distribution

The normal distribution is by far the most important probability distribution. One of the main reasons for that is the *Central Limit Theorem* (CLT) that we will discuss later in the book. To give you an idea, the CLT states that if you add a large number of random variables, the distribution of the sum will be approximately normal under certain conditions. The importance of this result comes from the fact that many random variables in real life can be expressed as the sum of a large number of random variables and, by the CLT, we can argue that distribution of the sum should be normal. The CLT is one of the most important results in probability and we will discuss it later on. Here, we will introduce normal random variables.

We first define the **standard normal random variable**. We will then see that we can obtain other normal random variables by *scaling* and *shifting* a standard normal random variable.

A continuous random variable Z is said to be a *standard normal (standard Gaussian)* random variable, shown as $Z \sim N(0, 1)$, if its PDF is given by

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\}, \quad \text{for all } z \in \mathbb{R}.$$

The $\frac{1}{\sqrt{2\pi}}$ is there to make sure that the area under the PDF is equal to one. We will verify that this holds in the solved problems section. Figure 4.6 shows the PDF of the standard normal random variable.

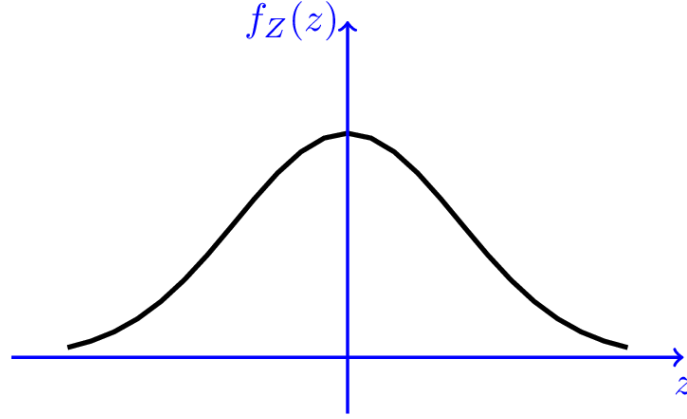


Fig.4.6 - PDF of the standard normal random variable.

Let us find the mean and variance of the standard normal distribution. To do that, we will use a simple useful fact. Consider a function $g(u) : \mathbb{R} \rightarrow \mathbb{R}$. If $g(u)$ is an odd function, i.e., $g(-u) = -g(u)$, and $|\int_0^\infty g(u)du| < \infty$, then

$$\int_{-\infty}^{\infty} g(u)du = 0.$$

For our purpose, let

$$g(u) = u^{2k+1} \exp\left\{-\frac{u^2}{2}\right\},$$

where $k = 0, 1, 2, \dots$. Then $g(u)$ is an odd function. Also $|\int_0^\infty g(u)du| < \infty$. One way to see this is to note that $g(u)$ decays faster than the function $\exp\{-u\}$ and since $|\int_0^\infty \exp\{-u\}du| < \infty$, we conclude that $|\int_0^\infty g(u)du| < \infty$. Now, let Z be a standard normal random variable. Then, we have

$$EZ^{2k+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^{2k+1} \exp\left\{-\frac{u^2}{2}\right\} du = 0,$$

for all $k \in \{0, 1, 2, \dots\}$. Thus, we have shown that for a standard normal random variable Z , we have

$$EZ = EZ^3 = EZ^5 = \dots = 0.$$

In particular, the standard normal distribution has zero mean. This is not surprising as we can see from Figure 4.6 that the PDF is symmetric around the origin, so we expect that $EZ = 0$. Next, let's find EZ^2 .

$$EZ^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 \exp\left\{-\frac{u^2}{2}\right\} du$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \left[-u \exp\left\{-\frac{u^2}{2}\right\} \right]_{-\infty}^{\infty} + \\
&+ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{u^2}{2}\right\} du \quad (\text{integration by parts}) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du \\
&= 1.
\end{aligned}$$

The last equality holds because we are integrating the standard normal PDF from $-\infty$ to ∞ . Thus, we conclude that for a standard normal random variable Z , we have

$$\text{Var}(Z) = 1.$$

So far we have shown the following:

$$\text{If } Z \sim N(0, 1), \text{ then } EZ = 0 \text{ and } \text{Var}(Z) = 1.$$

CDF of the standard normal

To find the CDF of the standard normal distribution, we need to integrate the PDF function. In particular, we have

$$F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left\{-\frac{u^2}{2}\right\} du.$$

This integral does not have a closed form solution. Nevertheless, because of the importance of the normal distribution, the values of $F_Z(z)$ have been tabulated and many calculators and software packages have this function. We usually denote the standard normal CDF by Φ .

The CDF of the standard normal distribution is denoted by the Φ function:

$$\Phi(x) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left\{-\frac{u^2}{2}\right\} du.$$

As we will see in a moment, the CDF of any normal random variable can be written in terms of the Φ function, so the Φ function is widely used in probability. Figure 4.7 shows the Φ function.

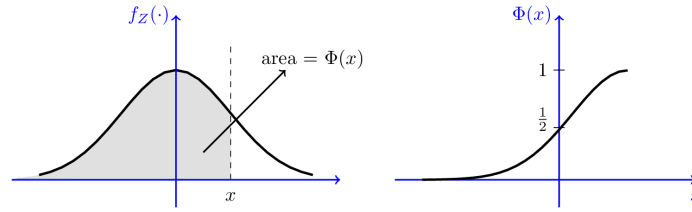


Fig.4.7 - The Φ function (CDF of standard normal).

Here are some properties of the Φ function that can be shown from its definition.

1. $\lim_{x \rightarrow \infty} \Phi(x) = 1, \quad \lim_{x \rightarrow -\infty} \Phi(x) = 0;$
2. $\Phi(0) = \frac{1}{2};$
3. $\Phi(-x) = 1 - \Phi(x),$ for all $x \in \mathbb{R}.$

Also, since the Φ function does not have a closed form, it is sometimes useful to use upper or lower bounds. In particular we can state the following bounds (see Problem 7 in the Solved Problems section). For all $x \geq 0$,

$$\frac{1}{\sqrt{2\pi}} \frac{x}{x^2 + 1} \exp\left\{-\frac{x^2}{2}\right\} \leq 1 - \Phi(x) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} \exp\left\{-\frac{x^2}{2}\right\} \quad (4.7)$$

As we mentioned earlier, because of the importance of the normal distribution, the values of the Φ function have been tabulated and many calculators and software packages have this function. For example, you can use the `normcdf` command in MATLAB to compute $\Phi(x)$ for a given number x . More specifically, `normcdf(x)` returns $\Phi(x)$. Also, the function `norminv` returns $\Phi^{-1}(x)$. That is, if you run $x = \text{norminv}(y)$, then x will be the real number for which $\Phi(x) = y$.

Normal random variables

Now that we have seen the standard normal random variable, we can obtain any normal random variable by shifting and scaling a standard normal random variable. In particular, define

$$X = \sigma Z + \mu, \quad \text{where } \sigma > 0.$$

Then

$$EX = \sigma EZ + \mu = \mu,$$

$$\text{Var}(X) = \sigma^2 \text{Var}(Z) = \sigma^2.$$

We say that X is a normal random variable with mean μ and variance σ^2 . We write $X \sim N(\mu, \sigma^2)$.

If Z is a standard normal random variable and $X = \sigma Z + \mu$, then X is a normal random variable with mean μ and variance σ^2 , i.e.,

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

Conversely, if $X \sim N(\mu, \sigma^2)$, the random variable defined by $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable, i.e., $Z \sim N(0, 1)$. To find the CDF of $X \sim N(\mu, \sigma^2)$, we can write

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(\sigma Z + \mu \leq x) \quad (\text{where } Z \sim N(0, 1)) \\ &= P\left(Z \leq \frac{x-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right). \end{aligned}$$

To find the PDF, we can take the derivative of F_X ,

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= \frac{d}{dx} \Phi\left(\frac{x-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma} \Phi'\left(\frac{x-\mu}{\sigma}\right) \quad (\text{chain rule for derivative}) \\ &= \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}. \end{aligned}$$

If X is a normal random variable with mean μ and variance σ^2 , i.e, $X \sim N(\mu, \sigma^2)$, then

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

$$F_X(x) = P(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

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

Figure 4.8 shows the PDF of the normal distribution for several values of μ and σ .

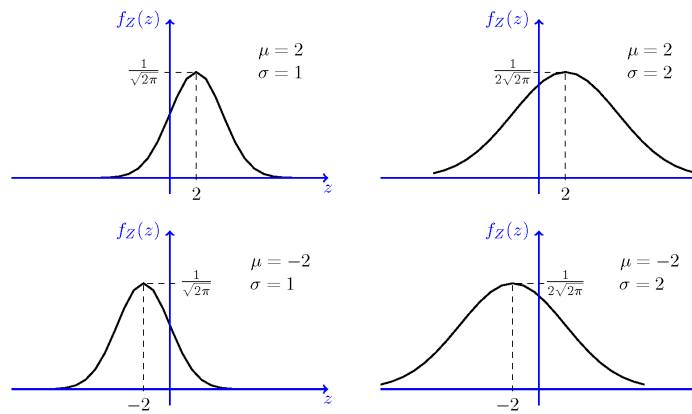


Fig.4.8 - PDF for normal distribution.

Example 4.11

Let $X \sim N(-5, 4)$.

- Find $P(X < 0)$.
- Find $P(-7 < X < -3)$.
- Find $P(X > -3 | X > -5)$.

Solution

X is a normal random variable with $\mu = -5$ and $\sigma = \sqrt{4} = 2$, thus we have

a. Find $P(X < 0)$:

$$\begin{aligned} P(X < 0) &= F_X(0) \\ &= \Phi\left(\frac{0 - (-5)}{2}\right) \\ &= \Phi(2.5) \approx 0.99 \end{aligned}$$

b. Find $P(-7 < X < -3)$:

$$\begin{aligned} P(-7 < X < -3) &= F_X(-3) - F_X(-7) \\ &= \Phi\left(\frac{(-3) - (-5)}{2}\right) - \Phi\left(\frac{(-7) - (-5)}{2}\right) \\ &= \Phi(1) - \Phi(-1) \\ &= 2\Phi(1) - 1 \quad (\text{since } \Phi(-x) = 1 - \Phi(x)) \\ &\approx 0.68 \end{aligned}$$

c. Find $P(X > -3 | X > -5)$:

$$\begin{aligned} P(X > -3 | X > -5) &= \frac{P(X > -3, X > -5)}{P(X > -5)} \\ &= \frac{P(X > -3)}{P(X > -5)} \\ &= \frac{1 - \Phi\left(\frac{(-3) - (-5)}{2}\right)}{1 - \Phi\left(\frac{(-5) - (-5)}{2}\right)} \\ &= \frac{1 - \Phi(1)}{1 - \Phi(0)} \\ &\approx \frac{0.1587}{0.5} \approx 0.32 \end{aligned}$$

An important and useful property of the normal distribution is that a linear transformation of a normal random variable is itself a normal random variable. In particular, we have the following theorem:

Theorem 4.3

If $X \sim N(\mu_X, \sigma_X^2)$, and $Y = aX + b$, where $a, b \in \mathbb{R}$, then $Y \sim N(\mu_Y, \sigma_Y^2)$ where

$$\mu_Y = a\mu_X + b, \quad \sigma_Y^2 = a^2\sigma_X^2.$$

Proof

We can write

$$X = \sigma_X Z + \mu_X \quad \text{where } Z \sim N(0, 1).$$

Thus,

$$\begin{aligned} Y &= aX + b \\ &= a(\sigma_X Z + \mu_X) + b \\ &= (a\sigma_X)Z + (a\mu_X + b). \end{aligned}$$

Therefore,

$$Y \sim N(a\mu_X + b, a^2 \sigma_X^2).$$
