4.4 Normal

Unit 4: Continuous Random Variables

Adapted from Blitzstein-Hwang Chapter 5.

The Normal distribution is a famous continuous distribution with a bell-shaped PDF. It is extremely widely used in statistics because of a theorem, the *central limit theorem*, which says that under very weak assumptions, the sum of a large number of i.i.d. random variables has an approximately Normal distribution, *regardless* of the distribution of the individual r.v.s. This means we can start with independent r.v.s from almost any distribution, discrete or continuous, but once we add up a bunch of them, the distribution of the resulting r.v. looks like a Normal distribution.

Definition 4.4.1 (Standard Normal distribution).

A continuous r.v. Z is said to have the *standard Normal distribution* if its PDF φ is given by

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty.$$

We write this as $Z \sim \mathcal{N}(0,1)$.

The constant $\frac{1}{\sqrt{2\pi}}$ in front of the PDF may look surprising (why is something with π needed in front of something with e, when there are no circles in sight?), but it turns out to be what is needed to make the PDF integrate to 1. Such constants are called *normalizing constants* because they normalize the total area under the PDF to 1. The standard Normal CDF Φ is the accumulated area under the PDF:

$$\Phi(z) = \int_{-\infty}^{z} \varphi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

We need to leave this as an integral: it turns out to be mathematically impossible to find a closed-form expression for the antiderivative of φ, meaning that we cannot express Φ as a finite sum of more familiar functions like polynomials or exponentials. But closed-form or no, it's still a well-defined function: if we give Φ an input z, it returns the accumulated area under the PDF from $-\infty$ up to z.

Notation 4.4.2.

By convention, we use φ for the standard Normal PDF and Φ for the standard Normal CDF. We will often use Z to denote a standard Normal random variable. The standard Normal PDF and CDF are plotted in Figure 4.4.3. The PDF is bell-shaped and symmetric about 0, and the CDF is S-shaped. These have the same general shape as the Logistic PDF and CDF that we saw in a couple of previous examples, but the Normal PDF decays to 0 much more quickly: notice that nearly all of the area under φ is between -3 and 3, whereas we had to go out to -5 and 5 for the Logistic PDF.

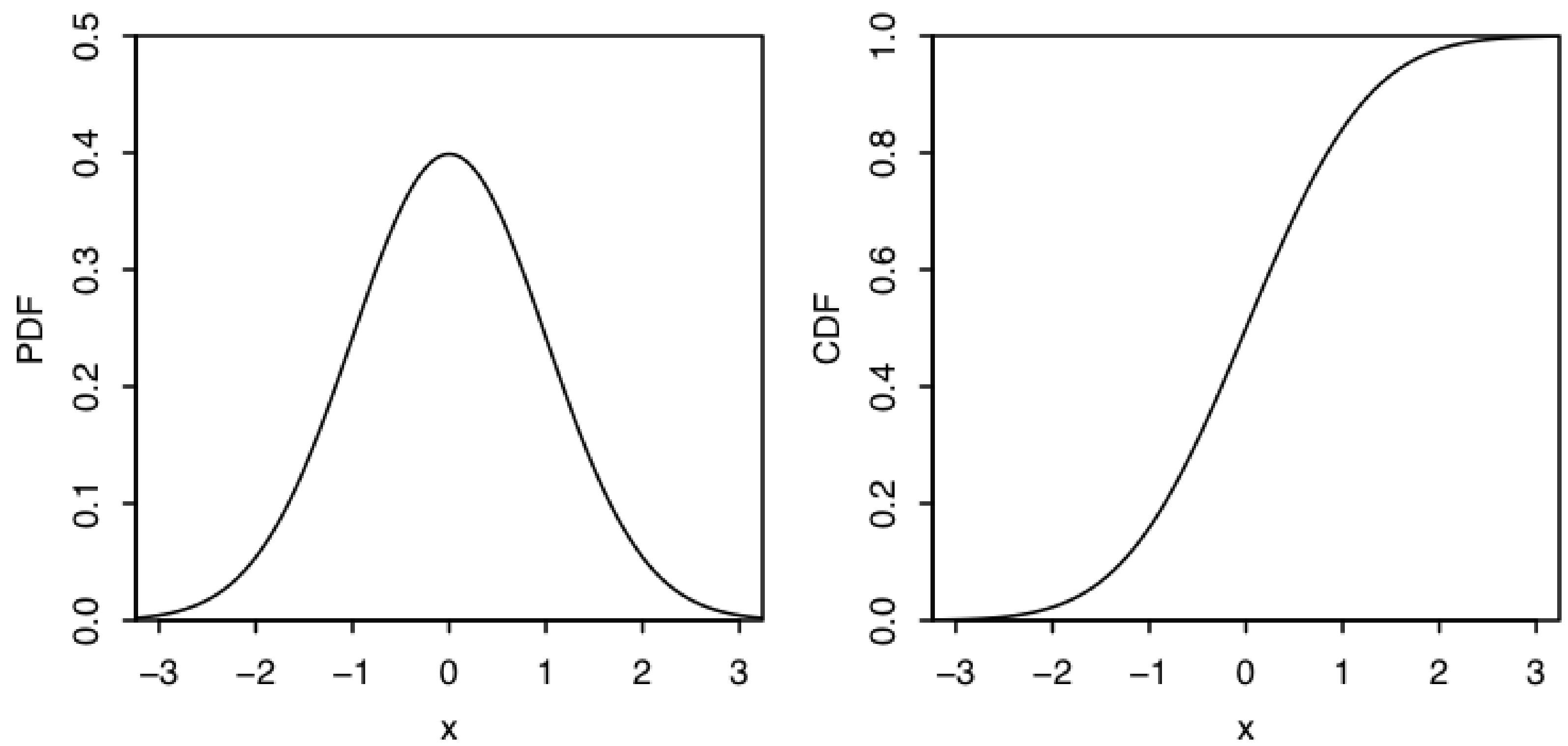

Figure 4.4.3: Standard Normal PDF φ (left) and CDF Φ (right).

View Larger Image
Image Description

There are several important symmetry properties that can be deduced from the standard Normal PDF and CDF.

1. *Symmetry of PDF*: φ satisfies $\varphi(z) = \varphi(-z)$, i.e., φ is an even function.
2. *Symmetry of tail areas*: The area under the PDF curve to the left of -2, which is $P(Z \leq -2) = \Phi(-2)$ by definition, equals the area to the right of 2, which is $P(Z \geq 2) = 1 - \Phi(2)$. In general, we have

$$\Phi(z) = 1 - \Phi(-z)$$

for all z. This can be seen visually by looking at the PDF curve, and mathematically by substituting $u = -t$ below and using the fact that PDFs integrate to 1:

$$\Phi(-z) = \int_{-\infty}^{-z} \varphi(t)dt = \int_{z}^{\infty} \varphi(u)du = 1 - \int_{-\infty}^{z} \varphi(u)du = 1 - \Phi(z).$$

3. *Symmetry of Z and $-Z$*: If $Z \sim \mathcal{N}(0,1)$, then $-Z \sim \mathcal{N}(0,1)$ as well. To see this, note that the CDF of $-Z$ is

$$P(-Z \le z) = P(Z \ge -z) = 1 - \Phi(-z),$$

but that is $\Phi(z)$, according to what we just argued. So $-Z$ has CDF Φ.

The general Normal distribution has two parameters, denoted μ and σ^2, which are the mean and variance (the mean and variance of a distribution are measures of the average and how spread out the distribution is, respectively; these are defined and explored in the next unit). Starting with a standard Normal r.v. $Z \sim \mathcal{N}(0,1)$, we can convert to a Normal r.v. with any desired parameters μ and σ^2 by a location-scale transformation.

Definition 4.4.4 (Normal distribution).

If $Z \sim \mathcal{N}(0,1)$, then

$$X = \mu + \sigma Z$$

is said to have the *Normal distribution* with mean parameter μ and variance parameter σ^2, for any real μ and σ^2 with $\sigma > 0$. We denote this by $X \sim \mathcal{N}(\mu, \sigma^2)$.

Of course, if we can get from Z to X, then we can get from X back to Z. The process of getting a standard Normal from a non-standard Normal is called, appropriately enough, *standardization*. For $X \sim \mathcal{N}(\mu, \sigma^2)$, the *standardized version* of X is

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1).$$

We can use standardization to find the CDF and PDF of X in terms of the standard Normal CDF and PDF.

Theorem 4.4.5 (Normal CDF and PDF).

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then the CDF of X is

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

and the PDF of X is

$$f(x) = \varphi\left(\frac{x - \mu}{\sigma}\right)\frac{1}{\sigma}.$$

Proof

For the CDF, we start from the definition $F(x) = P(X \leq x)$, standardize, and use the CDF of the standard Normal:

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

Then we differentiate to get the PDF, remembering to apply the chain rule:

$$\begin{aligned} f(x) &= \frac{d}{dx}\Phi\left(\frac{x-\mu}{\sigma}\right) \\ &= \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}. \end{aligned}$$

We can also write out the PDF as

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

Three important benchmarks for the Normal distribution are the probabilities of falling within one, two, and three standard deviations of the mean parameter μ. The 68-95-99.7 rule tells us that these probabilities are what the name suggests.

Theorem 4.4.6 (68-95-99.7 rule).

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\begin{aligned} P(|X-\mu| < \sigma) &\approx 0.68 \\ P(|X-\mu| < 2\sigma) &\approx 0.95 \\ P(|X-\mu| < 3\sigma) &\approx 0.997. \end{aligned}$$

Example 4.4.7.

Let $X \sim \mathcal{N}(-1, 4)$. What is $P(|X| < 3)$, exactly (in terms of Φ) and approximately?

Solution

The event $|X| < 3$ is the same as the event $-3 < X < 3$. We use standardization to express this event in terms of the standard Normal r.v. $Z = (X - (-1))/2$, then apply the 68-95-99.7 rule to get an approximation. The exact answer is

$$P(-3 < X < 3) = P\left(\frac{-3-(-1)}{2} < \frac{X-(-1)}{2} < \frac{3-(-1)}{2}\right) = P(-1 < Z < 2),$$

which is $\Phi(2) - \Phi(-1)$. The 68-95-99.7 rule tells us that $P(-1 < Z < 1) \approx 0.68$ and $P(-2 < Z < 2) \approx 0.95$. In other words, going from ± 1 standard deviation to ± 2 standard deviations adds approximately $0.95 - 0.68 = 0.27$ to the area under the curve. By symmetry, this is evenly divided between the areas $P(-2 < Z < -1)$ and $P(1 < Z < 2)$. Therefore,

$$P(-1 < Z < 2) = P(-1 < Z < 1) + P(1 < Z < 2) \approx 0.68 + \frac{0.27}{2} = 0.815.$$

This is close to the correct value, $\Phi(2) - \Phi(-1) \approx 0.8186$.

One more useful property of the Normal distribution is that the sum of independent Normals is Normal.

Theorem 4.4.8 (Sum of Independent Normals).

If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2),$$
$$X_1 - X_2 \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2).$$