LECTURE 11
NORMAL DISTRIBUTION

The normal distribution is the most important distribution used in statistics. As we will discuss later in the course, many distributions that arise in practice can be approximated by the normal distribution.

1 Definition

Definition 1. (Normal distribution) We say that a continuously distributed random variable X has a normal distribution with parameters $\mu \in (-\infty, \infty)$ and $\sigma^2 \in (0, \infty)$, denoted

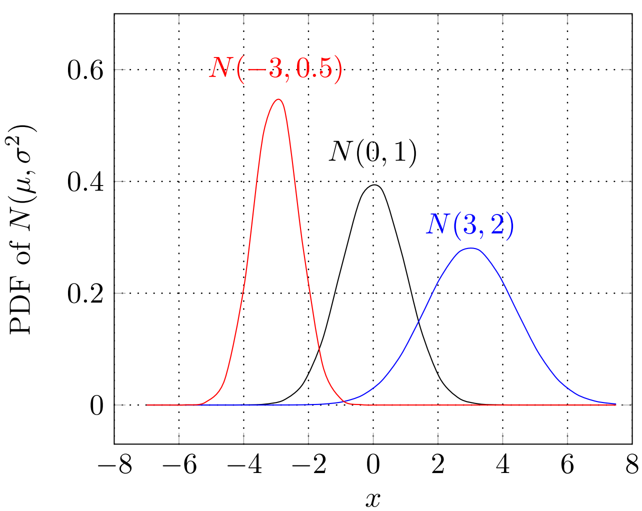
$$X \sim N(\mu, \sigma^2),$$

if the PDF of X is given by

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

Remark. 1. Definition 1 introduces the Normal *family of distributions*: different values of the parameters μ and σ^2 correspond to different normal distributions. Examples of several normal PDFs are plotted in Figure 1.

Figure 1: Normal PDFs for different values of μ and σ^2



2. When $\mu = 0$ and $\sigma^2 = 1$, the distribution is called *standard normal*, and the random

variable

$$Z \sim N(0,1)$$

is called a standard normal random variable.

3. The support of normal distributions is the entire real line $(-\infty, \infty)$. However, it is easy to see from equation (1) that the PDF of a normal distribution goes to zero exponentially fast as $x \to \pm\infty$. Consequently, the probabilities $P(X < x)$ or $P(X > x)$ also approach zero very fast as $x \to \pm\infty$. For example, in the case of standard normal random variable: about 95% of the probability mass is concentrated between -1.96 and 1.96, and approximately 99.9% of the probability mass is between -3.3 and 3.3.

2 CDF and quantiles

The CDF of a normal distribution does not have a closed-form expression. The CDF of $Z \sim N(0,1)$ is commonly denoted by Φ, and its PDF is commonly denoted by ϕ:

$$\begin{aligned} \Phi(z) &= \int_{-\infty}^{z} \phi(x)dx \\ &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx. \end{aligned}$$

This integral has been evaluated numerically, and the results are presented in normal tables (see Tables Va and Vb on pages 494-495 in Hogg, Tanis and Zimmerman).

Let z_τ denote the τ-th quantile of the standard normal distribution:

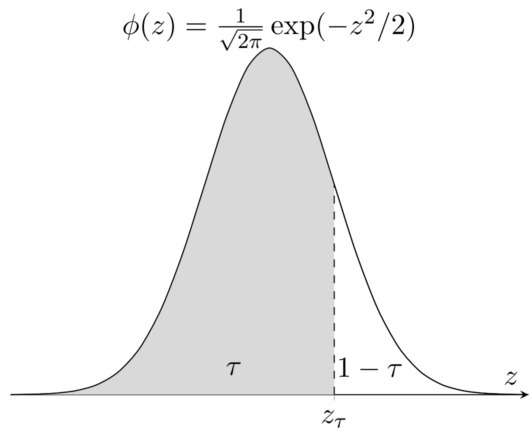
$$P(Z \leq z_\tau) = \tau$$

or using the notation for the standard normal CDF,

$$\Phi(z_\tau) = \tau.$$

Since the integral of a function represents the area under the graph of the function, the CDF at z can be shown graphically as the area under the PDF from $-\infty$ to z, see Figure 2.

Figure 2: CDF (Φ) as the area under the PDF (ϕ): $\Phi(z_\tau) = \tau$



In Lecture 10 we said that a distribution is symmetric around μ if its PDF satisfies $f_X(\mu - u) = f_X(\mu + u)$ for any $u \in \mathbb{R}$. In the case of $N(\mu, \sigma^2)$ distribution, the PDF is symmetric around μ:

$$\begin{aligned} f_X(\mu - u) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{((\mu-u)-\mu)^2}{\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(-u)^2}{\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{u^2}{\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{((\mu+u)-\mu)^2}{\sigma^2}\right) \\ &= f_X(\mu + u). \end{aligned}$$

From the results of Lecture 10 for symmetric distributions (Theorem 4), it follows that if $X \sim N(\mu, \sigma^2)$, then

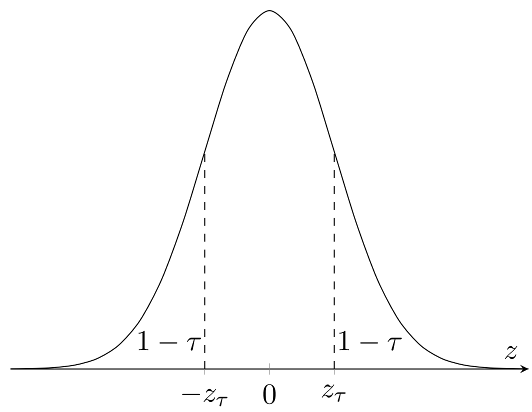
$$\begin{aligned} EX &= \mu, \\ \text{median: } q_{0.5} &= \mu, \\ q_{1-\tau} - \mu &= \mu - q_\tau, \end{aligned}$$

where for any $\tau \in (0,1)$, q_τ denotes the τ-th quantiles of X. In the case of standard normal distribution, since $\mu = 0$,

$$z_{1-\tau} = -z_\tau,$$

where for any $\tau \in (0,1)$, z_τ denotes the τ-th quantiles of $Z \sim N(0,1)$. This relationship is shown in Figure 3.

Figure 3: Quantiles of $Z \sim N(0,1)$: $z_{1-\tau} = -z_\tau$



3 MGF and moments

From the symmetry results, we have been able to deduce that for $X \sim N(\mu, \sigma^2)$,

$$EX = \mu.$$

To compute other moments, we can use the MGF. The MGF will be also very useful for some additional important results.

Theorem 2. *Let $X \sim N(\mu, \sigma^2)$. Then the MGF of X is given by*

$$M_X(t) = \exp\left(\mu t + \sigma^2 t^2/2\right), \quad t \in \mathbb{R}. \tag{2}$$

Remark. Due to uniqueness of the MGF, any distribution with the MGF described in (2) must be $N(\mu, \sigma^2)$.

Proof. By the definition of the MGF,

$$\begin{aligned} M_X(t) &= Ee^{tX} \\ &= \int e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{tx - \frac{(x-\mu)^2}{2\sigma^2}} dx. \end{aligned}$$

The integral is difficult to solve directly. Therefore, we will transform the expression inside the integral and bring it to the form of a normal PDF. Then, to eliminate the integral, we will use the fact that a PDF must integrate to one. Consider just the exponent for the expression inside the integral:

$$\begin{aligned} tx - \frac{(x-\mu)^2}{2\sigma^2} &= \frac{tx \cdot 2\sigma^2 - (x-\mu)^2}{2\sigma^2} \\ &= \frac{2x(\sigma^2 t + \mu) - x^2 - \mu^2}{2\sigma^2} \\ &= \frac{2x(\sigma^2 t + \mu) - x^2 - (\sigma^2 t + \mu)^2 + (\sigma^2 t + \mu)^2 - \mu^2}{2\sigma^2} \\ &= \frac{-(x - (\sigma^2 t + \mu))^2 + \sigma^4 t^2 + 2\mu t \sigma^2}{2\sigma^2} \\ &= -\frac{(x - (\sigma^2 t + \mu))^2}{2\sigma^2} + \left(\mu t + \frac{\sigma^2 t^2}{2}\right), \end{aligned}$$

where in the third line we subtracted and added back $(\sigma^2 t + \mu)^2$. Hence,

$$\begin{aligned} e^{tx - \frac{(x-\mu)^2}{2\sigma^2}} &= e^{-\frac{(x-(\sigma^2 t+\mu))^2}{2\sigma^2} + \left(\mu t + \frac{\sigma^2 t^2}{2}\right)} \\ &= e^{-\frac{(x-(\sigma^2 t+\mu))^2}{2\sigma^2}} e^{\mu t + \frac{\sigma^2 t^2}{2}}. \end{aligned}$$

However, note that the second exponent in the expression on the right-hand side does not involve x, and therefore the corresponding term can be moved outside the integral:

$$\begin{aligned} \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{tx - \frac{(x-\mu)^2}{2\sigma^2}} dx &= e^{\mu t + \frac{\sigma^2 t^2}{2}} \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{-\frac{(x-(\sigma^2 t+\mu))^2}{2\sigma^2}} dx \\ &= e^{\mu t + \frac{\sigma^2 t^2}{2}}, \end{aligned}$$

where the last equality follows from the fact that

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-(\sigma^2 t+\mu))^2}{2\sigma^2}}$$

is the PDF of $N(\mu + \sigma^2 t, \sigma^2)$ distribution and therefore

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int e^{-\frac{(x-(\sigma^2 t+\mu))^2}{2\sigma^2}} dx = 1.$$

□

We can now use the MGF to compute the moments of $N(\mu, \sigma^2)$ distribution.

$$
\begin{aligned}
M_X'(t) &= \frac{d}{dt}\left(e^{\mu t+\frac{\sigma^2 t^2}{2}}\right) \\
&= e^{\mu t+\frac{\sigma^2 t^2}{2}}(\mu+\sigma^2 t). \\
M_X''(t) &= \frac{d^2}{dt^2}\left(e^{\mu t+\frac{\sigma^2 t^2}{2}}\right) \\
&= \frac{d}{dt}\left(e^{\mu t+\frac{\sigma^2 t^2}{2}}(\mu+\sigma^2 t)\right) \\
&= e^{\mu t+\frac{\sigma^2 t^2}{2}}(\mu+\sigma^2 t)^2 + e^{\mu t+\frac{\sigma^2 t^2}{2}}\sigma^2.
\end{aligned}
$$

As expected,

$$M_X'(0) = \mu.$$

For the second moment,

$$
\begin{aligned}
EX^2 &= M_X''(0) \\
&= \mu^2 + \sigma^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
Var(X) &= EX^2 - (EX)^2 \\
&= (\mu^2+\sigma^2) - \mu^2 \\
&= \sigma^2.
\end{aligned}
$$

Thus, the parameters μ and σ^2 in $N(\mu, \sigma^2)$ correspond to the mean and the variance respectively. Furthermore, a normal distribution is completely determined by the mean and the variance. In other words, if X is normally distributed, one needs to know only the mean and the variance of the distribution to describe the behavior of X.

Since the distribution is symmetric around μ, all odd moments are equal to μ:

$$EX^r = \mu, \quad r = 1, 3, 5, \ldots. \tag{3}$$

This can be shown using the exactly the same arguments as in the proof of $EX = \mu$ (see Lecture 10, Theorem 4(a)). Alternatively, (3) can be stated as all odd *central* moments being equal to zero:

$$E(X-\mu)^r = 0, \quad r = 1, 3, 5, \ldots.$$

The result is true not only for normal, but any symmetric distribution, and the third central moment is often used to check deviations from symmetry.

Definition 3. (Coefficient of skewness) The coefficient of skewness is defined as the third standardized central moment:

$$skewness = \frac{E(X-EX)^3}{(Var(X))^{3/2}} = \frac{E(X-EX)^3}{(E(X-EX)^2)^{3/2}}.$$

Remark. A negative value for the skewness coefficient indicates that the distribution is skewed left (has a longer left tail), and a positive value indicates that the distribution is skewed right (has a longer right tail).

Higher moments contain additional information on the shape of a distribution. The fourth moment moment is sometimes used to measure the peakedness or flatness of a distribution.

Definition 4. (Kurtosis)

$$kurtosis = \frac{E(X-EX)^4}{(Var(X))^2} = \frac{E(X-EX)^4}{(E(X-EX)^2)^2}.$$

Remark. It can be shown that the kurtosis of a standard normal distribution is 3. Hence, the values of kurtosis different from 3 can be used for detecting deviations from normality. For that reason, we often use the *excess kurtosis*, which defined as the kurtosis minus 3:

$$excess\ kurtosis = \frac{E(X-EX)^4}{(Var(X))^2} - 3.$$

In the case of symmetric distributions, a positive value for the excess kurtosis indicates that the ratio of the fourth and second central moments is smaller than that of a normal distribution. As a result, the PDF is more peaked than the PDF of a normal distribution. Similarly, a negative value of the excess kurtosis indicates that the PDF is flatter and the tails are fatter than those of a normal distribution.

4 Linear transformations of normal variables

In statistical applications, we often have to work with linear functions of normal variables. The following result establishes that a linear function of a normal random variable is also normally distributed.

Theorem 5. *Suppose that $X \sim N(\mu, \sigma^2)$. Let a and $b \neq 0$ be two constants. Then*

$$Y = a + bX \sim N(a + b\mu, b^2\sigma^2).$$

Proof. We will prove the result using the MGF (since every distribution has a unique MGF).

Note that by (2), the MGF of $N(a + b\mu, b^2\sigma^2)$ is

$$e^{(a+b\mu)t+(b\sigma)^2t^2/2}.$$

Hence, we need to show that the MGF of Y is equal to the above expression.

$$\begin{aligned} M_Y(t) &= Ee^{tY} \\ &= Ee^{t(a+bX)} \\ &= e^{ta}Ee^{tbX} \\ &= e^{ta}M_X(tb) \\ &= e^{ta}e^{\mu tb+\sigma^2(tb)^2/2} \\ &= e^{(a+\mu b)t+(\sigma b)^2t^2/2}, \end{aligned}$$

where the third equality holds since e^{ta} is a constant, the fourth equality holds by the definition of the MGF since b is a constant, and the equality in the line before the last holds by (2) and since $X \sim N(\mu, \sigma^2)$. □

The theorem implies that if Z is a standard normal random variable, then

$$X = \mu + \sigma Z \sim N(\mu, \sigma^2),$$

provided that $\sigma \neq 0$. Alternatively, suppose that $X \sim N(\mu, \sigma^2)$. Then,

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

Hence, any random variable with $N(\mu, \sigma^2)$ distribution can be written as $X = \mu + \sigma Z$, where $Z \sim N(0, 1)$. We often refer to $(X - \mu)/\sigma$ as a standardized variable.