MH1820 Week 5

1. Normal distribution

2. Chi-square distribution

The Normal distribution

The Normal distribution

The **Normal distribution** is the most important continuous probability distribution in the entire field of statistics.

It is also known as the **Gaussian distribution**.

Its graph is called the **normal curve** and is a bell-shaped curve.

The probability density function (PDF) of a **normal random variable** X is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty.$$

Notation:

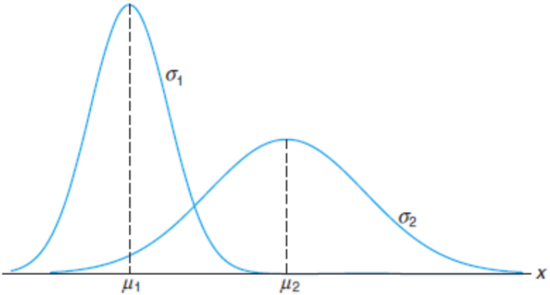
$$X \sim N(\mu, \sigma^2)$$

Theorem 1 (Normal distribution)

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

$$\mathbb{E}[X] = \mu, \quad \mathrm{Var}[X] = \sigma^2.$$

Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$.



If $\mu = 0$ and $\sigma = 1$, then $Z \sim N(0,1)$ is called a **standard normal random variable**, and its distribution is called a **standard normal distribution**. The CDF of Z is denoted by

$$\Phi(z) = \mathbb{P}(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

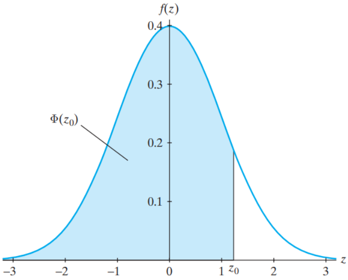


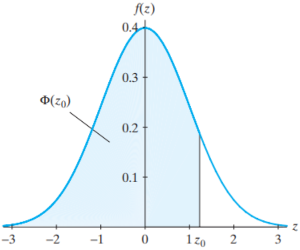
Figure 3.3-1 Standard normal pdf

It is useful to know that for any z, we have

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

The CDF $\Phi(z)$ cannot be calculated by hand, so in this course we will just look them up in the table.

Table Va The Standard Normal Distribution Function



$$P(Z \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw$$

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

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389

Theorem 2 (Transforming to standard normal)

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

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

The above allows us to compute probability of a normal random variable $X \sim N(\mu, \sigma^2)$ using $Z \sim N(0,1)$. E.g.

$$\begin{aligned}
\mathbb{P}(a \leq X \leq b) &= \mathbb{P}\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= \mathbb{P}\left({\color{red}Z} \leq \frac{b-\mu}{\sigma}\right) - \mathbb{P}\left({\color{red}Z} \leq \frac{a-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)
\end{aligned}$$

Example 3

If $X \sim N(3, 16)$, find $\mathbb{P}(X \geq 5)$ and $\mathbb{P}(4 \leq X \leq 8)$

Solution. Note that $\mu = 3$ and $\sigma = \sqrt{16} = 4$.

$$\begin{aligned}\mathbb{P}(X \geq 5) &= 1 - \mathbb{P}(X \leq 5) = 1 - \mathbb{P}\left(Z \leq \frac{5-3}{4}\right) \\ &= 1 - \Phi(0.5) = 1 - 0.6915 = 0.3085.\end{aligned}$$

$$\begin{aligned}\mathbb{P}(4 \leq X \leq 8) &= \Phi\left(\frac{8-3}{4}\right) - \Phi\left(\frac{4-3}{4}\right) \\ &= \Phi(1.25) - \Phi(0.25) = 0.8944 - 0.5987 = 0.2957.\end{aligned}$$

$\square$

Example 4

If $X \sim N(25, 36)$, find c such that

$$\mathbb{P}(|X - 25| \leq c) = 0.9544.$$

Solution. Note that $\mu = 25$, $\sigma = 6$, and so

$$\begin{aligned}
\mathbb{P}(|X - 25| \leq c) &= 0.9544 \\
\mathbb{P}\left(-\frac{c}{6} \leq \frac{X-25}{6} \leq \frac{c}{6}\right) &= 0.9544 \\
\Phi\left(\frac{c}{6}\right) - \Phi\left(-\frac{c}{6}\right) &= 0.9544 \\
\Phi\left(\frac{c}{6}\right) - 1 + \Phi\left(\frac{c}{6}\right) &= 0.9544 \\
\Phi\left(\frac{c}{6}\right) &= 0.9772.
\end{aligned}$$

From the table for $\Phi(z)$, we deduce that $\frac{c}{6} = 2$, and so $c = 12$.

Example 5

Suppose X, the grade on a midterm exam, is normally distributed with mean 70 and standard deviation 10. The instructor wants to give 15% of the class an A (the best grade). What cutoff should the instructor use to determine who gets an A?

Solution. We have $X \sim N(70, 10^2)$. We want to find the cutoff C such that

$$\begin{aligned} \mathbb{P}(X \geq C) &= 0.15 \\ \mathbb{P}\left(Z \geq \frac{C-70}{10}\right) &= 0.15 \end{aligned}$$

From the table (for right-tail probabilities),

$$\frac{C-70}{10} \approx 1.04$$

$$C = 80.4$$

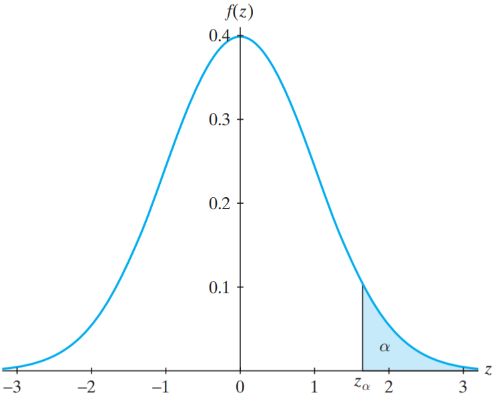
$\square$

In statistical applications, we are often interested in the numbers called percentiles. Let $Z \sim N(0, 1)$.

- The $100(1 - \alpha)$**th percentile** (or **the upper** 100α**th percentage point**) for the standard normal is the number z_α such that

$$\mathbb{P}(Z > z_\alpha) = \mathbb{P}(Z \geq z_\alpha) = \alpha.$$

- The 100α**th percentile** is the number $z_{1-\alpha}$.



Note that by symmetry we have $\mathbb{P}(Z \le -z_\alpha) = \mathbb{P}(Z \ge z_\alpha) = \alpha$.
For a table of $\mathbb{P}(Z \ge z_\alpha)$, see see NTULearn $->$ Content $->$ TABLES.pdf.

Example 6

Find $z_{0.0125}$.

Solution. Note that

$$\mathbb{P}(Z > z_{0.0125}) = 0.0125.$$

From the table, we have $z_{0.0125} = 2.24$. $\square$

Example 7

Let $X = Z^2$, where $Z \sim N(0,1)$. Compute the CDF of X. Hence, deduce the PDF of X.

Solution. The CDF of X is given by

$$\begin{aligned} F(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(Z^2 \leq x) \\ &= \mathbb{P}(-\sqrt{x} \leq Z \leq \sqrt{x}) \\ &= \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) \\ &= \Phi(\sqrt{x}) - 1 + \Phi(\sqrt{x}) \\ &= 2\Phi(\sqrt{x}) - 1. \end{aligned}$$

The PDF $f(x)$ is given by

$$\begin{aligned}
f(x) &= F'(x) \\
&= \frac{d}{dx}\left(2\Phi(\sqrt{x}) - 1\right) \\
&= 2\frac{d}{du}\Phi(u) \cdot \frac{du}{dx} \quad (\text{where } u = \sqrt{x}) \\
&= 2\frac{1}{\sqrt{2\pi}}e^{-u^2/2} \cdot \frac{1}{2\sqrt{x}} \\
&= \frac{1}{\sqrt{2\pi}}x^{-1/2}e^{-x/2}.
\end{aligned}$$

$\square$

Chi-square distribution

Chi-square distribution

The preceding example is a **chi-square distribution** with 1 degree of freedom, denoted by $\chi^2(1)$. In general, we have the following:

Theorem 8 (Chi-square distribution)

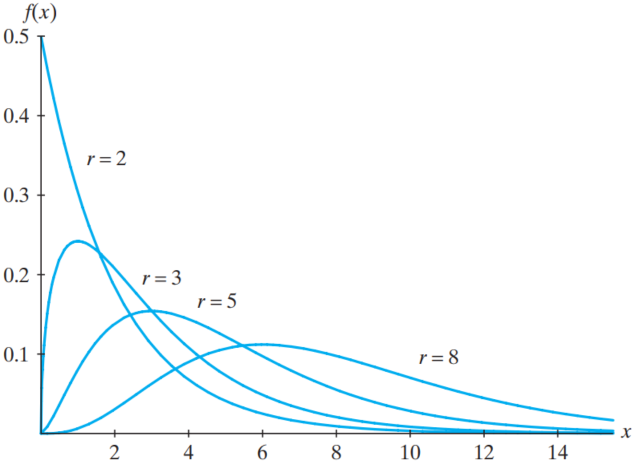
Suppose $X_i = Z^2$ *where* $Z \sim N(0,1)$*, for* $i = 1, \ldots, r$*, are independently and identically distributed (i.i.d).*

Then $X = \sum_{i=1}^{r} X_i$ *has a* **chi-square distribution with** r **degree of freedom**, *denoted by* $X \sim \chi^2(r)$.

In fact,

$$\chi^2(r) = Gamma\left(\alpha = \frac{r}{2}, \theta = 2\right).$$

Chi-square distribution $\chi^2(r)$ for different degrees of freedom $r = 2, 3, 5, 8$.



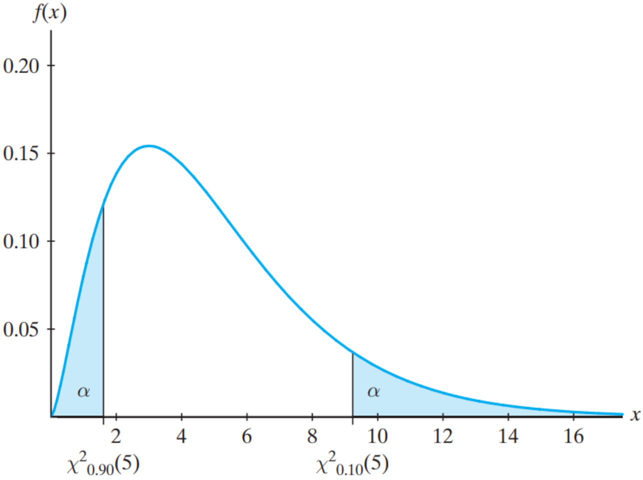
Let $X \sim \chi^2(r)$. Similar to standard normal, we define

- the $100(1-\alpha)$**th percentile** (or **upper** 100α**th percentage point**) $\chi^2_\alpha(r)$ to be the number such that

$$\mathbb{P}(X \geq \chi^2_\alpha(r)) = \alpha.$$

- the 100α**th percentile** to be the number $\chi^2_{1-\alpha}(r)$.

Chi-square tails, $r = 5$, $\alpha = 0.10$:



For a table of $\chi^2_\alpha(r)$, see NTULearn $->$ Content $->$ TABLES.pdf

Example 9

Let X have a chi-square distribution with $r = 5$ degree of freedom. Find the probabilty that X is between 1.145 and 12.83.

Solution. From the table for $\chi^2_\alpha(5)$, we have

$$
\begin{aligned}
\mathbb{P}(1.145 \le X \le 12.83) &= \mathbb{P}(\chi^2_{0.95}(5) \le X \le \chi^2_{0.025}(5)) \\
&= 0.975 - 0.05 = 0.925.
\end{aligned}
$$

$\square$

Example 10

If customers arrive at a shop on the average of 30 per hour in accordance with a Poisson process, what is the probability that the shopkeeper will have to wait longer than 9.390 minutes for the first nine customers to arrive?

Solution. The average number of customers arriving at the shop is $\lambda = \frac{30}{60} = \frac{1}{2}$ per minutes. Let X be the waiting time in minutes until the ninth customer. Then $X \sim \textit{Gamma}(\alpha = 9, \theta = 2)$. Let $r = 18$. Then

$$X \sim \chi^2(18) = \textit{Gamma}(\alpha = \frac{r}{2} = 9, \theta = 2).$$

We now use the tabel for $\chi^2_\alpha(18)$ to calculate the probabilty $\mathbb{P}(X > 9.390)$.

From the table, we have

$$\mathbb{P}(X > 9.390) = \mathbb{P}(X > \chi^2_{0.95}(18)) = 0.95.$$

$\square$