Normal distributions.

- Normal distributions are commonly used because distributions of averages of IID random variables can often be approximated well by normal distributions.
- Normal distribution $N(\mu, \sigma^2)$. Suppose that μ and $\sigma > 0$ are constants. Suppose that X is a random variable with PDF f_X, where

  $$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty.$$

  Then the distribution of X is called the normal distribution with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$. Here $e = \lim_{n\to\infty}(1 + 1/n)^n \approx 2.718$.
- Fact 1 If $X \sim N(\mu, \sigma^2)$, then

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

  $N(0, 1)$ is called the standard normal distribution (標準常態分布). The result follows from Fact 1 in the handout "The probability density function (PDF) of a continuous random variable".
- Fact 1 implies the probabilities, mean and variance for $N(\mu, \sigma^2)$ can be obtained using the probabilities, mean and variance for $N(0, 1)$.
- The mean and variance for $N(0, 1)$.

  Example 1. Suppose that $X \sim N(0, 1)$. Find $E(X)$ and $Var(X)$ using R.

  Sol. Let f be the PDF for $N(0, 1)$. Then

  $$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

  and

  $$Var(X) = E(X^2) - (E(X))^2 = \int_{-\infty}^{\infty} x^2 f(x) dx - (E(X))^2.$$

  Running the following R codes gives $\int_{-\infty}^{\infty} x f(x) dx$:

```
f <- function(x){ exp(-x^2/2)/sqrt(2*pi) }
g <- function(x){ x*f(x) }
integrate(g,-Inf,Inf)$value
```

The output is 0, so

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = 0.$$

Running the following R codes gives $\int_{-\infty}^{\infty} x^2 f(x) dx$:

```
f <- function(x){ exp(-x^2/2)/sqrt(2*pi) }
g <- function(x){ x^2*f(x) }
integrate(g,-Inf,Inf)$value
```

The output is 1, so

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = 1$$

and $Var(X) = E(X^2) - (E(X))^2 = 1 - 0^2 = 1$.

- Mean and variance for $N(\mu, \sigma^2)$. Suppose that $X \sim N(\mu, \sigma^2)$. Then

$$E(X) = \mu \text{ and } Var(X) = \sigma^2. \tag{1}$$

  One may derive (1) using Fact 1 and the following results.

  1. The mean and variance for $N(0, 1)$ are 0 and 1 respectively.
  2. For any constants a and b,

$$E(a + bX) = a + bE(X) \text{ and } Var(a + bX) = b^2 Var(X). \tag{2}$$

- From Fact 1,

$$P(a < N(\mu, \sigma^2) < b) = P\left(\frac{a - \mu}{\sigma} < N(0, 1) < \frac{b - \mu}{\sigma}\right).$$

- $P(a < N(0, 1) < b)$ can be found by using R (or other software packages) to compute

$$\int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

  Or, one can use normal tables to find $P(a < N(0, 1) < b)$.

- Example 2. Find $P(0 < N(0,1) < 1.5)$ using R.
  Sol. Running the following R codes gives $P(0 < N(0,1) < 1.5)$:

```
f <- function(x){ exp(-x^2/2)/sqrt(2*pi) }
integrate(f, 0, 1.5)$value
#Output: 0.4331928
```

- Example 3. Use the table "Normal probabilities" at

  `https://stat.walkup.tw/teaching/statistics/tables/normal.pdf`

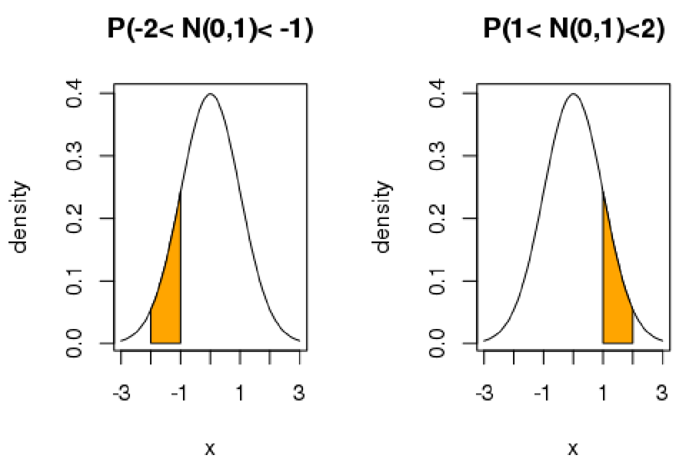
  to find $P(0 < N(0,1) < 1.5)$ and $P(0 < N(0,1) < 1.51)$.

z	0.00	0.01	0.02	$\cdots$	0.09
0.0	0.0000	0.0040	0.0080	$\cdots$	0.0359
0.1	0.0398	0.0438	0.0478	$\cdots$	0.0753
$\vdots$			$\vdots$		
1.5	0.4332	0.4345	0.4357	$\cdots$	0.4441
$\vdots$			$\vdots$		
3.0	0.4987	0.4987	0.4987	$\cdots$	0.4990

  - $P(0 < N(0,1) < 1.5) \doteq 0.4332.$
  - $P(0 < N(0,1) < 1.51) = P(0 < N(0,1) < (1.5 + 0.01)) \doteq 0.4345.$

- The $N(0,1)$ PDF is symmetric about 0. Therefore, for $0 \le a < b \le \infty$,

$$P(-b < N(0,1) < -a) = P(a < N(0,1) < b).$$



- Example 4. Find $P(-1.5 < N(0,1) < 1.3)$.

  Sol.

$$\begin{aligned} P(-1.5 < N(0,1) < 1.3) &= P(-1.5 < N(0,1) < 0) + P(0 \le N(0,1) < 1.3) \\ &= P(0 < N(0,1) < 1.5) + P(0 < N(0,1) < 1.3) \\ &\doteq 0.4332 + 0.4032 = 0.8364. \end{aligned}$$

- Example 5. Find $P(N(0,1) > 1.3)$ and $P(N(0,1) < -1.5)$.

  Sol.

$$\begin{aligned} P(N(0,1) > 1.3) &= P(N(0,1) > 0) - P(0 < N(0,1) \le 1.3) \\ &= P(N(0,1) > 0) - P(0 < N(0,1) < 1.3) \\ &\doteq 0.5 - 0.4032 = 0.0968, \end{aligned}$$

  and

$$\begin{aligned} P(N(0,1) < -1.5) &= P(N(0,1) > 1.5) \\ &= P(N(0,1) > 0) - P(0 < N(0,1) \le 1.5) \\ &= P(N(0,1) > 0) - P(0 < N(0,1) < 1.5) \\ &\doteq 0.5 - 0.4332 = 0.0668. \end{aligned}$$

- Example 6. Suppose that $X \sim N(25, 100)$.

  (a) Find $P(25 < X < 35)$.

  (b) Find $P(X < 4)$.

  Sol.

  (a)

$$\begin{aligned} P(25 < X < 35) &= P\left(\frac{25-25}{10} < \frac{X-25}{10} < \frac{35-25}{10}\right) \\ &= P(0 < N(0,1) < 1) \doteq 0.3413. \end{aligned}$$

  (b)

$$\begin{aligned} P(X < 4) &= P\left(\frac{X-25}{10} < \frac{4-25}{10}\right) \\ &= P(N(0,1) < -2.1) \\ &= P(N(0.1) > 2.1) \\ &= P(N(0,1) > 0) - P(0 < N(0,1) < 2.1) \\ &\doteq 0.5 - 0.4821 = 0.0179. \end{aligned}$$

- Example 7. Suppose that the daily average temperatures at a weather station in July follow (approximately) the normal distribution with mean 25 degrees Celsius (攝氏25度) and standard deviation 10 degrees Celsius.

  (a) What is the probability that in a future day in July, the average temperature at the weather station is between 25 degrees Celsius and 35 degrees Celsius?

  (b) What is the probability that in a future day in July, the average temperature at the weather station is below 4 degrees Celsius?

  Sol. See the solution to Example 6.

- The empirical rule. Suppose that $X \sim N(\mu, \sigma^2)$ and $\sigma > 0$. Then

$$P(\mu - k\sigma < X < \mu + k\sigma) \doteq \begin{cases} 0.68 & \text{if } k = 1; \\ 0.95 & \text{if } k = 2; \\ 1 & \text{if } k = 3. \end{cases}$$

- Recall that for a random variable X such that $E(X) = \mu$ and $\sqrt{Var(X)} = \sigma$, Chebyshev's theorem states that

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2} = \begin{cases} 0 & \text{if } k = 1; \\ 0.75 & \text{if } k = 2; \\ \doteq 0.89 & \text{if } k = 3. \end{cases} \quad (3)$$

  If $X \sim N(\mu, \sigma^2)$, then from (1), (3) holds. However, the guaranteed coverage probabilities from (3) are not very close to the exact coverage probability.

- Consider the daily average temperature distribution in Example 7.

  (a) Apply the empirical rule to find two temperatures such that for about 95 percent of future days in July, the daily average temperatures are between the two values.

  (b) Apply the empirical rule to find two temperatures such that for almost all of future days in July, the daily average temperatures are between the two values.

  Ans. For about 95 percent of future days in July, the daily average temperatures are between 5 degrees Celsius and 45 degrees Celsius ($25 \pm 2(10)$). For almost all of future days in July, the daily average temperatures are between -5 degrees Celsius and 55 degrees Celsius ($25 \pm 3(10)$).

- Fact 2 If $X \sim N(\mu, \sigma^2)$, then $a + bX \sim N(E(a + bX), Var(a + bX))$ for $b \neq 0$.

Example 8. Suppose that $X \sim N(1,4)$ and $Y = 5 - 2X$. Find $P(3 < Y < 9)$.

Sol. $E(Y) = 5 - 2E(X) = 3$ and $Var(Y) = 4Var(X) = 16$, so

$$P(3 < Y < 9) = P\left(0 < \frac{Y-3}{\sqrt{16}} < 1.5\right) = P(0 < N(0,1) < 1.5) = 0.4332.$$