

## Chapter 3. Random Variables and Distributions

**Basic idea** of introducing random variables: represent outcomes and/or random events by numbers.

### 3.1 Random variables and Distributions.

**Definition.** A random variable is a function defined on the sample space  $\Omega$ , which assigns a real number  $X(\omega)$  to each outcome  $\omega \in \Omega$ .

**Example 1.** Flip a coin 10 times. We may define random variables (r.v.s) as follows:

$X_1$  = no. of heads,

$X_2$  = no. of flips required to have the first head,

$X_3$  = no. of 'HT'-pairs,  
 $X_4$  = no. of tails.

For  $\omega = HHTHHTHHTT$ ,  $X_1(\omega) = 6$ ,  $X_2(\omega) = 1$ ,  $X_3(\omega) = 3$ ,  $X_4(\omega) = 4$ .  
Note  $X_1 \equiv 10 - X_4$ .

**Remark.** The values of a r.v. varies and cannot be pre-determined before an outcome occurs.

**Definition.** For any r.v.  $X$ , its (cumulative) distribution function (CDF) is defined as  $F_X(x) = P(X \leq x)$ .

**Example 2.** Toss a fair coin twice and let  $X$  be the number of heads. Then

$$P(X = 0) = P(X = 2) = 1/4, \quad P(X = 1) = 1/2.$$

Hence its CDF is  $F_X(x) = \begin{cases} 0 & x < 0, \\ 1/4 & 0 \leq x < 1, \\ 3/4 & 1 \leq x < 2, \\ 1 & x \geq 2. \end{cases}$

**Note.** (i)  $F_X(x)$  is right continuous, non-decreasing, and defined for all  $x \in (-\infty, \infty)$ . For example,  $F_X(1.1) = 0.75$ .

(ii) The CDF is a non-random function.

(iii) If  $F(\cdot)$  is the CDF of r.v.  $X$ , we simply write  $X \sim F$ .

**Properties of CDF.** A function  $F(\cdot)$  is a CDF iff

(i)  $F$  is non-decreasing:  $x_1 < x_2$  implies  $F(x_1) \leq F(x_2)$ ,

(ii)  $F$  is normalized:  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow \infty} F(x) = 1$ ,

(iii)  $F$  is right continuous:  $\lim_{y \downarrow x} F(y) = F(x)$ .

## Probabilities from CDF

$$(a) P(X > x) = 1 - F(x)$$

$$(b) P(x < X \leq y) = F(y) - F(x)$$

$$(c) P(X < x) = \lim_{h \downarrow 0} F(x - h) \equiv F(x-)$$

$$(d) P(X = x) = F(x) - F(x-).$$

**Note.** It is helpful for understanding (c) & (b) to revisit Example 2.

### 3.2 Discrete random variables

If r.v.  $X$  only takes some isolated values,  $X$  is called a discrete r.v. Its CDF is called a discrete distribution.

**Definition.** For a discrete r.v.  $X$  taking values  $\{x_1, x_2, \dots\}$ , we define the probability function (or probability mass function) as

$$f_X(x_i) = P(X = x_i), \quad i = 1, 2, \dots$$

Obviously,  $f_X(x_i) \geq 0$  and  $\sum_i f_X(x_i) = 1$ .

It is often more convenient to list a probability function in a table:

$X$	$x_1$	$x_2$	$\dots\dots$
Probability	$f_X(x_1)$	$f_X(x_2)$	$\dots\dots$

**Example 2** (continue). The probability function is be tabulated:

$X$	0	1	2
Probability	$1/4$	$1/2$	$1/4$

**Expectation or Mean**  $EX$  or  $E(X)$ : a measure for the ‘center’, ‘average value’ of a r.v.  $X$ , and is often denoted by  $\mu$ .

For a discrete r.v.  $X$  with probability function  $f_X(x)$ ,

$$\mu = EX = \sum_i x_i f_X(x_i).$$

**Variance**  $\text{Var}(X)$ : a measure for variation, uncertainty or ‘risk’ of a r.v.  $X$ , is often denoted by  $\sigma^2$ , while  $\sigma$  is called **standard deviation** of  $X$ .

For a discrete r.v.  $X$  with probability function  $f_X(x)$ ,

$$\sigma^2 = \text{Var}(X) = \sum_i (x_i - \mu)^2 f_X(x_i) = \sum_i x_i^2 f_X(x_i) - \mu^2.$$

**The  $k$ -th moment** of  $X$ :  $\mu_k \equiv E(X^k) = \sum_i x_i^k f_X(x_i)$ ,  $k = 1, 2, \dots$ .

Obviously,  $\mu = \mu_1$ , and  $\sigma^2 = \mu_2 - \mu_1^2$ .



## Some important discrete distributions

**Convention.** We often use upper case letters  $X, Y, Z, \dots$  to denote r.v.s, and lower case letters  $x, y, z, \dots$  to denote the values of r.v.s. In contrast letters  $a, b$  or  $A, B$  are often used to denote (non-random) constants.

*Degenerate distribution:*  $X \equiv a$ , i.e.  $F_X(x) = 1$  for any  $x \geq a$ , and 0 otherwise.

It is easy to see that  $\mu = a$  and  $\sigma^2 = 0$ .

*Bernoulli distribution:*  $X$  is binary,  $P(X = 1) = p$  and  $P(X = 0) = 1 - p$ , where  $p \in [0, 1]$  is a constant. It represents the outcome of flipping a coin.

$$\mu = 1 \cdot p + 0 \cdot (1 - p) = p, \quad \sigma^2 = p(1 - p).$$

**Note.** Bernoulli trial refers to an experiment of flipping a coin repeatedly.

*Binomial distribution*  $\text{Bin}(n, p)$ :  $X$  takes values  $0, 1, \dots, n$  only with the probability function

$$f_X(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n.$$

**Theorem.** If we toss a coin  $n$  times, let  $X$  be the number of heads. Then  $X \sim \text{Bin}(n, p)$ , where  $p$  is the probability that head occurs in tossing the coin once.

**Proof.** Let  $\omega = HTHTT \dots H$  denote an outcome of  $n$  tosses. Then  $X = k$  iff there are  $k$  'H' and  $(n - k)$  'T' in  $\omega$ . Therefore the probability of such a  $\omega$  is  $p^k (1 - p)^{n-k}$ . Since those  $k$  H's may occur in any  $n$  positions of the sequence, there are  $\binom{n}{k}$  such  $\omega$ 's. Hence

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Let us check if the probability function above is well defined. Obviously  $P(X = k) \geq 0$ , furthermore

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = \{p + (1-p)\}^n = 1^n = 1.$$

Let us work out the mean and the variance for  $X \sim \text{Bin}(n, p)$ .

$$\begin{aligned} \mu &= \sum_{k=0}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{j=0}^{n-1} np \frac{(n-1)!}{j!(n-1-j)!} p^j (1-p)^{n-1-j} \\ &= np \sum_{j=0}^m \frac{m!}{j!(m-j)!} p^j (1-p)^{m-j} = np. \end{aligned}$$

Note that  $\sigma^2 = E(X^2) - \mu^2 = E\{X(X-1)\} + \mu - \mu^2$ . We need to work out

$$\begin{aligned} E\{X(X-1)\} &= \sum_{k=0}^n k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=2}^n n(n-1) p^2 \frac{(n-2)!}{(k-2)! \{(n-2)-(k-2)\}} p^{k-2} (1-p)^{\{(n-2)-(k-2)\}} \\ &= n(n-1) p^2 \sum_{j=0}^{n-2} \frac{(n-2)!}{j! \{(n-2)-j\}} p^j (1-p)^{\{(n-2)-j\}} = n(n-1) p^2. \end{aligned}$$

This gives  $\sigma^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$ .

By the above theorem, we can see immediately

- (i) If  $X \sim \text{Bin}(n, p)$ ,  $n - X \sim \text{Bin}(n, 1 - p)$ .
- (ii) If  $X \sim \text{Bin}(n, p)$ ,  $Y \sim \text{Bin}(m, p)$ , and  $X$  and  $Y$  are independent, then  $X + Y \sim \text{Bin}(n + m, p)$ .

Furthermore, let  $Y_i = 1$  if the  $i$ -th toss yields H, and 0 otherwise. Then  $Y_1, \dots, Y_n$  are *independent* Bernoulli r.v.s with mean  $p$  and variance  $p(1-p)$ . Since  $X = Y_1 + \dots + Y_n$ , we notice

$$EX = \sum_{i=1}^n EY_i = np, \quad \text{Var}(X) = \sum_{i=1}^n \text{Var}(Y_i) = np(1-p).$$

This is a much easier way to derived the means and variances for binomial distributions, which is based on the following general properties.

---

(i) For any r.v.s  $\xi_1, \dots, \xi_n$ , and any constants  $a_1, \dots, a_n$ ,

$$E\left(\sum_{i=1}^n a_i \xi_i\right) = \sum_{i=1}^n a_i E(\xi_i).$$

(ii) If, in addition,  $\xi_1, \dots, \xi_n$  are *independent*,

$$\text{Var}\left(\sum_{i=1}^n a_i \xi_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(\xi_i).$$

---

---

**Independence of random variables.** The r.v.s  $\xi_1, \dots, \xi_n$  are independent if

$$P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n) = P(\xi_1 \leq x_1) \times \dots \times P(\xi_n \leq x_n)$$

for any  $x_1, \dots, x_n$ .

---

---

**Moment generate function (MGF)** of r.v.  $X$ :

$$\psi_X(t) = E(e^{tX}), \quad t \in (-\infty, \infty).$$

(i) It is easy to see that  $\psi'_X(0) = E(X) = \mu$ . In general  $\psi_X^{(k)}(0) = E(X^k) = \mu_k$ .

(ii) If  $Y = a + bX$ ,  $\psi_Y(t) = E(e^{(a+bX)t}) = e^{at}\psi_X(bt)$ .

(iii) If  $X_1, \dots, X_n$  are independent,  $\psi_{\sum_i X_i}(t) = \prod_{i=1}^n \psi_{X_i}(t)$ , and vice versa

---

If  $X$  is discrete,  $\psi_X(t) = \sum_i e^{x_i t} f_X(x_i)$ .

To generate a r.v. from  $\text{Bin}(n, p)$ , we can flip a coin (with  $p$ -probability for H)  $n$  times, and count the number of heads. However R can do the flipping for us much more efficiently:

```
> rbinom(10, 100, 0.1) # generate 10 random numbers from \Bin(100, 0.1)
[1] 8 11 9 9 18 7 5 5 3 7
> rbinom(10, 100, 0.1) # do it again, obtain different numbers
[1] 11 13 6 7 11 9 9 9 12 10
> x <- rbinom(10, 100, 0.7); x; mean(x)
[1] 66 77 67 66 64 68 70 68 72 72
[1] 69 # mean close to np=70
> x <- rbinom(10, 100, 0.7); x; mean(x)
[1] 70 73 72 70 68 69 70 66 79 71
[1] 70.8
```



Note that `rbinom(10000, 1, 0.5)` is equivalent to toss a fair coin 10000 times:

```
> y <- rbinom(10000, 1, 0.5); length(y); table(y)
[1] 10000
y
 0      1
4990 5010 # about a half times with head
```

You may try with smaller sample size, such as

```
> y <- rbinom(10, 1, 0.5); length(y); table(y)
[1] 10
y
0 1
3 7 # 7 heads and 3 tails
```

Also try out `pbinom` (CDF), `dbinom` (probability function), `qbinom` (quantile) for Binomial distributions.

*Geometric Distribution*  $\text{Geom}(p)$ :  $X$  takes all positive integer values with probability function

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

Obviously,  $X$  is the number of tosses required in a Bernoulli trial to obtain the first head.

$$\mu = \sum_{k=1}^{\infty} k(1 - p)^{k-1} p = -p \frac{d}{dp} \sum_{k=1}^{\infty} (1 - p)^k = -p \frac{d}{dp} (1/p) = 1/p,$$

and it can be shown that  $\sigma^2 = (1 - p)/p^2$ .

Using the MGF provides an alternative way to find mean and variance: for

$t < -\log(1 - p)$  (i.e.  $e^t(1 - p) < 1$ ),

$$\begin{aligned}\psi_X(t) &= E(e^{tX}) = \sum_{i=1}^{\infty} e^{ti}(1 - p)^{i-1}p = \frac{p}{1 - p} \sum_{i=1}^{\infty} \{e^t(1 - p)\}^i \\ &= \frac{p}{1 - p} \frac{e^t(1 - p)}{1 - e^t(1 - p)} = \frac{pe^t}{1 - e^t(1 - p)} = \frac{p}{e^{-t} - 1 + p}.\end{aligned}$$

Now  $\mu = \psi'_X(0) = \left[ \frac{pe^{-t}}{(e^{-t} - 1 + p)^2} \right]_{t=0} = 1/p$ , and

$$\mu_2 = \psi''_X(0) = \left[ \frac{2pe^{-2t}}{(e^{-t} - 1 + p)^3} - \frac{pe^{-t}}{(e^{-t} - 1 + p)^2} \right]_{t=0} = 2/p^2 - 1/p.$$

Hence  $\sigma^2 = \mu_2 - \mu^2 = (1 - p)/p^2$ .

The R functions for  $\text{Geom}(p)$ : `rgeom`, `dgeom`, `pgeom` and `qgeom`.

*Poisson Distribution*  $\text{Poisson}(\lambda)$ :  $X$  takes all non-negative integers with probability function

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots,$$

where  $\lambda > 0$  is a constant, called parameter.

The MGF  $X \sim \text{Poisson}(\lambda)$ :

$$\psi_X(t) = \sum_{k=0}^{\infty} e^{kt} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \lambda)^k}{k!} = e^{-\lambda} e^{e^t \lambda} = \exp\{\lambda(e^t - 1)\}.$$

Hence

$$\mu = \psi'_X(0) = [\exp\{\lambda(e^t - 1)\} \lambda e^t]_{t=0} = \lambda,$$

$$\mu_2 = \psi''_X(0) = [\exp\{\lambda(e^t - 1)\} \lambda e^t + \exp\{\lambda(e^t - 1)\} (\lambda e^t)^2]_{t=0} = \lambda + \lambda^2.$$

Therefore  $\sigma^2 = \mu_2 - \mu^2 = \lambda$ .

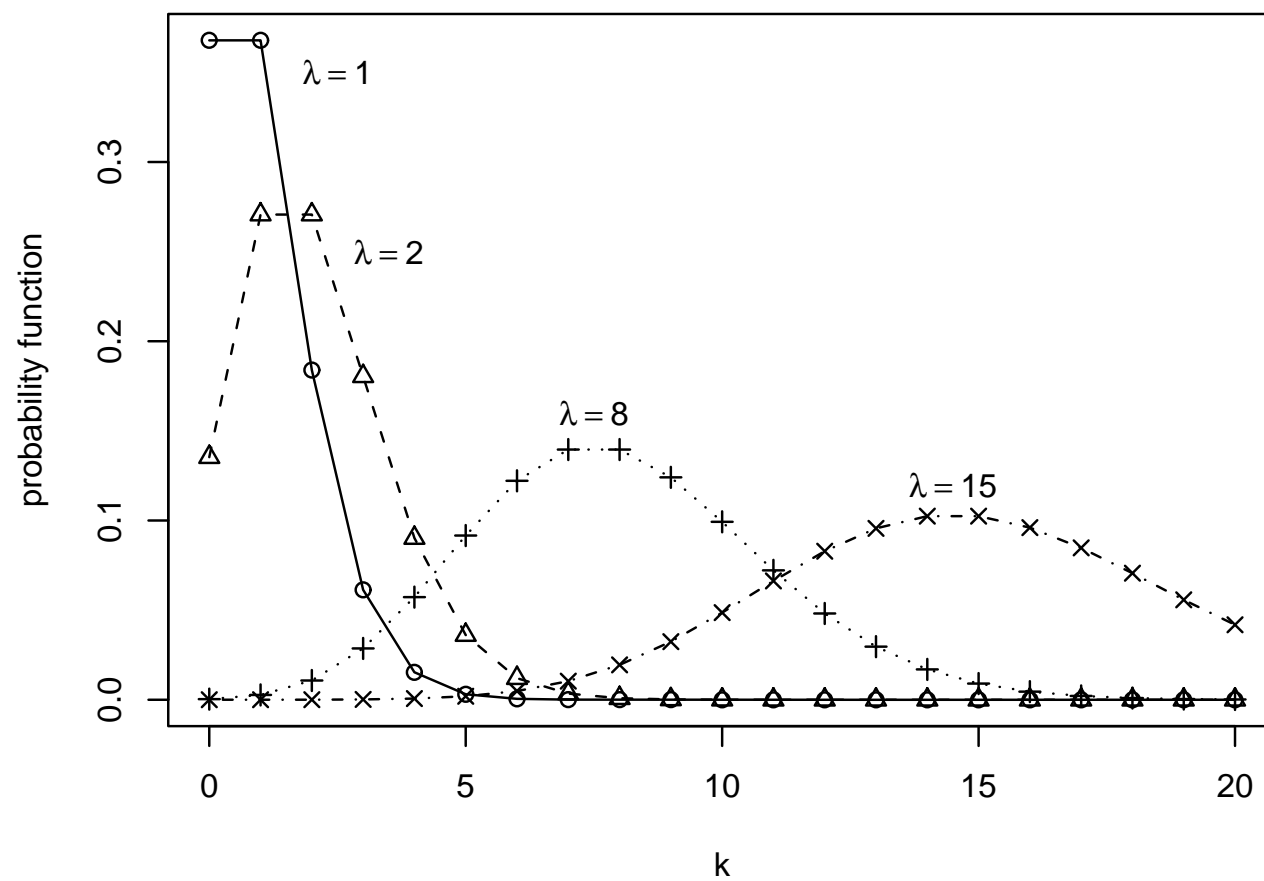
**Remark.** For Poisson distributions,  $\mu = \sigma^2$ .

The R functions for Poisson( $\lambda$ ): `rpois`, `dpois`, `ppois` and `qpois`.

To understand the role of the parameter  $\lambda$ , we plot the probability function of Poisson( $\lambda$ ) for different values of  $\lambda$ .

```
> x <- c(0:20)
> plot(x,dpois(x,1),type='o',xlab='k', ylab='probability function')
> text(2.5,0.35, expression(lambda==1))
> lines(x,dpois(x,2),typ='o',lty=2, pch=2)
> text(3.5,0.25, expression(lambda==2))
> lines(x,dpois(x,8),typ='o',lty=3, pch=3)
> text(7.5,0.16, expression(lambda==8))
> lines(x,dpois(x,15),typ='o',lty=4, pch=4)
> text(14.5, 0.12, expression(lambda==15))
```

## Plots of $\lambda^k e^{-\lambda} / k!$ against $k$



### Three ways of computing probability and distribution functions:

- calculators — for simple calculation
- statistical tables — for, e.g. the final exam
- R — for serious tasks such as real application

## 3.2 Continuous random variables

A r.v.  $X$  is *continuous* if there exists a function  $f_X(\cdot) \geq 0$  such that

$$P(a < X < b) = \int_a^b f_X(x) dx, \quad \forall a < b.$$

We call  $f_X(\cdot)$  the *probability density function* (PDF) or, simply, density function. Obviously

$$F_X(x) = \int_{-\infty}^x f_X(u) du.$$

### Properties of continuous random variables

(i)  $F_X(x) = P(X \leq x) = P(X < x)$ , i.e.  $P(X = x) = 0 \neq f_X(x)$ .



(ii) The PDF  $f_X(\cdot) \geq 0$ , and  $\int_{-\infty}^{\infty} f_X(x)dx = 1$ .

(iii)  $\mu = E(X) = \int_{-\infty}^{\infty} xf_X(x)dx$ ,

$$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x)dx = \int_{-\infty}^{\infty} x^2 f_X(x)dx - \mu^2.$$

Furthermore the MGF of  $X$  is equal to

$$\psi_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx.$$

## Some important continuous distributions

*Uniform distribution*  $U(a, b)$ :  $X$  takes any values between  $a$  and  $b$  equally likely. Its PDF is

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

Then the CDF is

$$F(x) = \int_{-\infty}^x f(u)du = \begin{cases} 0 & x < a, \\ \frac{1}{b-a} \int_a^x du = \frac{x-a}{b-a} & a \leq x \leq b, \\ 1 & x > b. \end{cases},$$

and

$$\mu = \int_a^b \frac{x dx}{b-a} = \frac{a+b}{2}, \quad \mu_2 = \int_a^b \frac{x^2 dx}{b-a} = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

Hence  $\sigma^2 = \mu_2 - \mu^2 = (b-a)^2/12$ .

R-functions related to uniform distributions: runif, dunif, punif, qunif.

---

**Quantile.** For a given CDF  $F(\cdot)$ , its quantile function is defined as

$$F^{-1}(p) = \inf\{x : F(x) \geq p\}, \quad p \in [0, 1]$$

---

```
> x <- c(1, 2.5, 4)
> punif(x, 2, 3)      # CDF of U(2, 3) at 1, 2.5 and 4
[1] 0 0.5 1
> dunif(x, 2, 3)      # PDF of U(2, 3) at 1, 2.5 and 4
[1] 0 1 0
> qunif(0.5, 2, 3)    # quantile of U(2, 3) at p=0.5
[1] 2.5
```

*Normal Distribution*  $N(\mu, \sigma^2)$ : the PDF is

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

where  $\mu \in (-\infty, \infty)$  is the ‘centre’ (or mean) of the distribution, and  $\sigma > 0$  is the ‘spread’ (standard deviation).

**Remarks.** (i) The most important distribution in statistics: Many phenomena in nature have approximately normal distributions. Furthermore, it provides asymptotic approximations for the distributions of sample means (Central Limit Theorem).

(ii) If  $X \sim N(\mu, \sigma^2)$ ,  $EX = \mu$ ,  $\text{Var}(X) = \sigma^2$ , and  $\psi_X(t) = e^{\mu t + \sigma^2 t^2/2}$ .

We compute  $\psi_X(t)$  below, the idea is applicable in general.

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

(iii) Standard normal distribution:  $N(0, 1)$ .

If  $X \sim N(\mu, \sigma^2)$ ,  $Z \equiv (X - \mu)/\sigma \sim N(0, 1)$ . Hence

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

where

$$\Phi(x) = P(Z < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du$$

is the CDF of  $N(0, 1)$ . Its values are tabulated in all statistical tables.

**Example 3.** Let  $X \sim N(3, 5)$ .

$$P(X > 1) = 1 - P(X < 1) = 1 - P\left(Z < \frac{1-3}{\sqrt{5}}\right) = 1 - \Phi(-0.8944) = 0.81.$$

Now find  $x = \Phi^{-1}(0.2)$ , i.e.  $x$  satisfies the equation

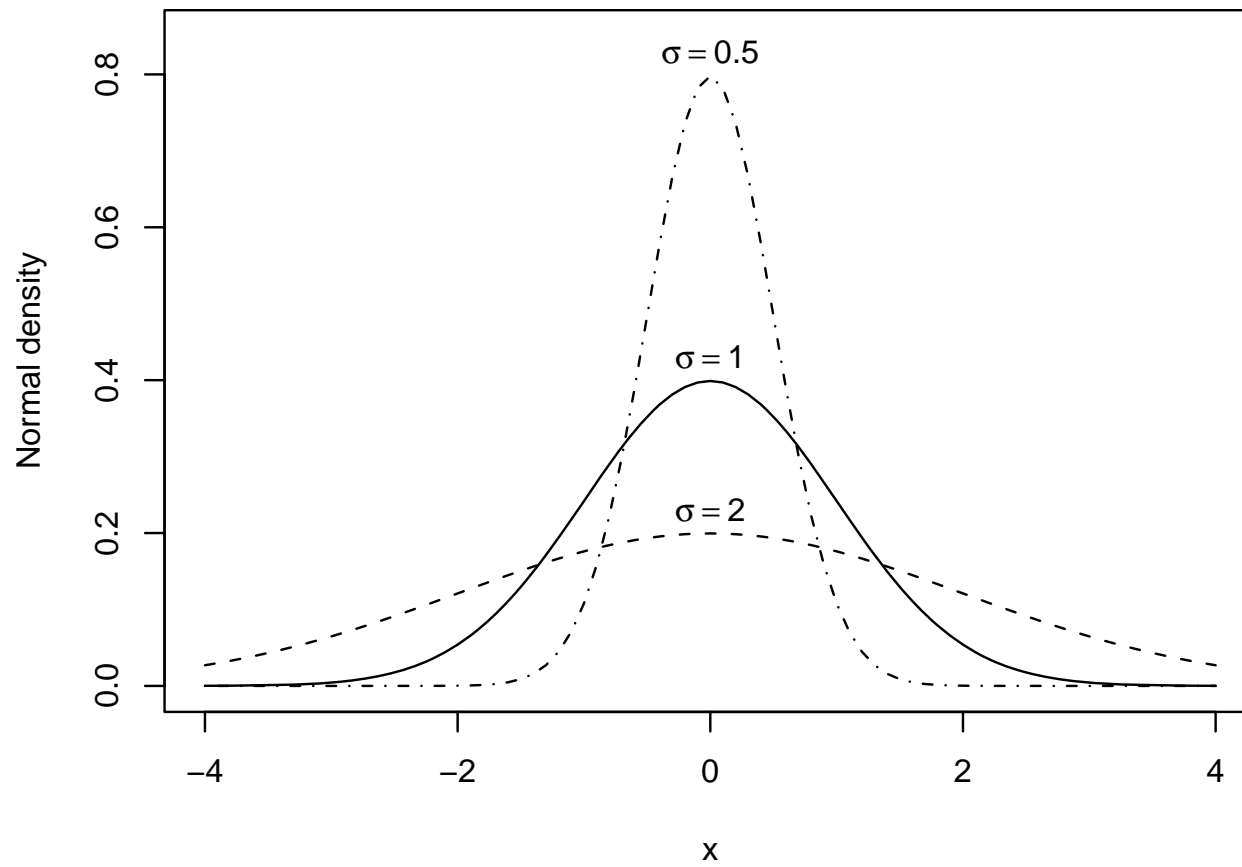
$$0.2 = P(X < x) = P\left(Z < \frac{x-3}{\sqrt{5}}\right).$$

From the normal table,  $\Phi(-0.8416) = 0.2$ . Therefore  $(x-3)/\sqrt{5} = -0.8416$ , leading to the solution  $x = 1.1181$ .

**Note.** You may check the answers using R:

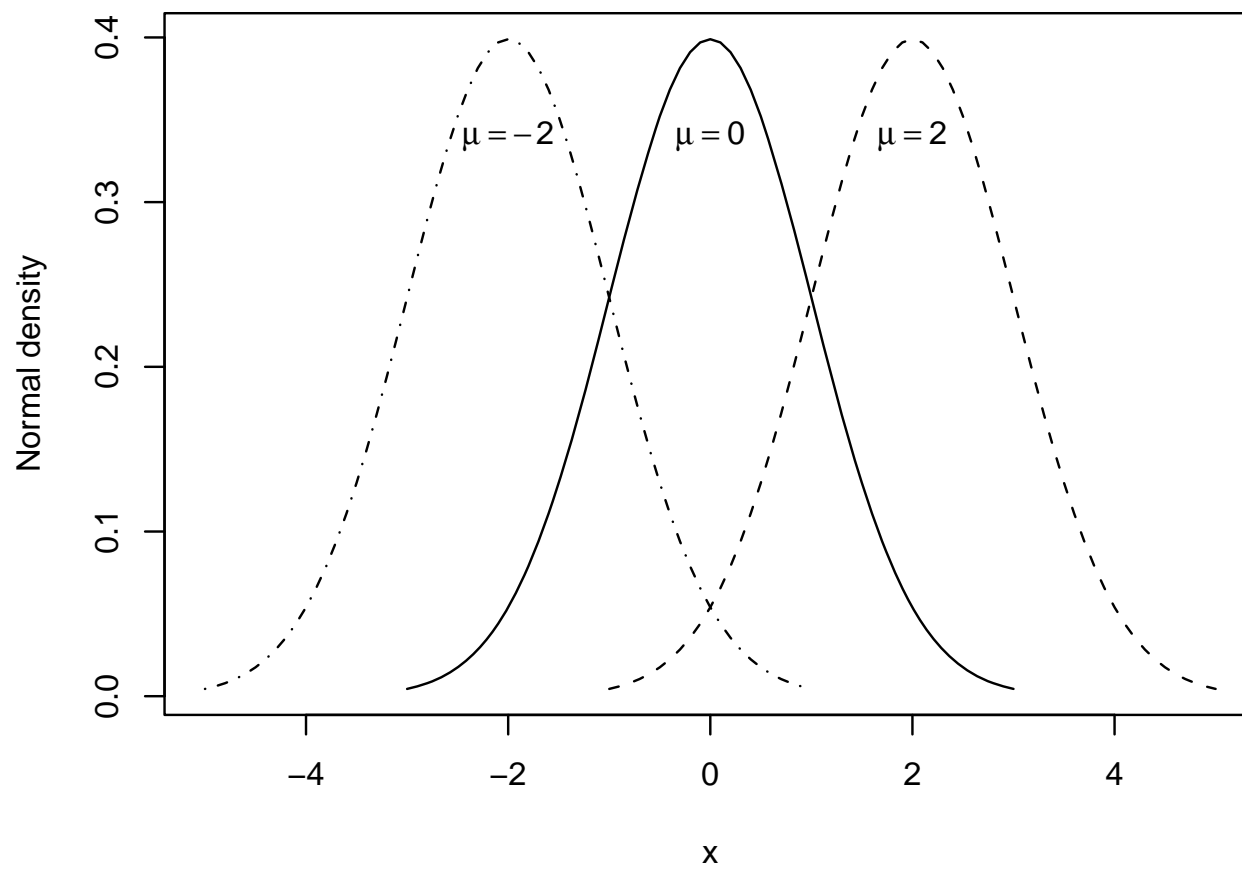
```
1 - pnorm(1, 3, sqrt(5)),  
  qnorm(0.2, 3, sqrt(5))
```

## Density functions of $N(0, \sigma^2)$





## Density functions of $N(\mu, 1)$



Below are the R codes which produce the two normal density plots.

```
x <- seq(-4, 4, 0.1)      # x = (-4, -3.9, -3.8, ..., 3.9, 4)
plot(x, dnorm(x, 0, 1), type='l', xlab='x', ylab='Normal density',
      ylim=c(0, 0.85))
text(0,0.43, expression(sigma==1))
lines(x, dnorm(x, 0, 2), lty=2)
text(0,0.23, expression(sigma==sqrt(2)))
lines(x, dnorm(x, 0, 0.5), lty=4)
text(0,0.83, expression(sigma==sqrt(0.5)))
```

```
x <- seq(-3, 3, 0.1)
plot(x, dnorm(x, 0, 1), type='l', xlab='x', ylab='Normal density',
      xlim=c(-5, 5))
text(0,0.34, expression(mu==0))
lines(x+2, dnorm(x+2, 2, 1), lty=2)
text(2,0.34, expression(mu==2))
lines(x-2, dnorm(x-2, -2, 1), lty=4)
text(-2,0.34, expression(mu==-2))
```

*Exponential Distribution*  $\text{Exp}(\lambda)$ :  $X \sim \text{Exp}(\lambda)$  if  $X$  has the PDF

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda} & x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where  $\lambda > 0$  is a parameter.

$$E(X) = \lambda, \quad \text{Var}(X) = \lambda^2, \quad \psi_X(t) = 1/(1 - t\lambda).$$

**Background.**  $\text{Exp}(\lambda)$  is used to model the lifetime of electronic components and the waiting times between rare events.

*Gamma Distribution*  $\text{Gamma}(\alpha, \beta)$ :  $X \sim \text{Gamma}(\alpha, \beta)$  if  $X$  has the PDF

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where  $\alpha, \beta > 0$  are two parameters, and  $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$ .

$$E(X) = \alpha\beta, \quad \text{Var}(X) = \alpha\beta^2, \quad \psi_X(t) = (1 - t\beta)^{-\alpha}.$$

**Note.**  $\text{Gamma}(1, \beta) = \text{Exp}(\beta)$ .

*Cauchy Distribution:* the PDF of the Cauchy distribution is

$$f(x) = \frac{1}{\pi(1 + x^2)}, \quad x \in (-\infty, \infty).$$

As  $E(|X|) = \infty$ , the mean and variance of the Cauchy distribution do not exist. Cauchy Distribution is particularly useful to model the data with excessively large, or negatively large outliers.