

# Statistics 1

D. González Sánchez  
Licenciatura en Economía  
CIDE

Fall 2022

*One can describe Statistics as the mathematical discipline whose purpose is to use empirical data generated by a random phenomenon, in order to make inferences about some deterministic characteristics of the phenomenon while simultaneously quantifying the uncertainty inherent in these inferences.*

Panaretos [5, p. xiii]

## Contents

<b>1</b>	<b>Basics of probability</b>	<b>3</b>
<b>2</b>	<b>Discrete random variables</b>	<b>3</b>
2.1	Mean, variance, and standard deviation . . . . .	3
2.2	Common discrete distributions . . . . .	5
2.2.1	Bernoulli distribution . . . . .	5
2.2.2	Binomial distribution . . . . .	5
2.2.3	Geometric distribution . . . . .	6
2.2.4	Poisson distribution . . . . .	7
2.3	Moment-generating functions . . . . .	7
2.4	Exercises . . . . .	8
<b>3</b>	<b>Continuous random variables</b>	<b>9</b>
3.1	Mean and variance . . . . .	9
3.2	Common continuous distributions . . . . .	11
3.2.1	Uniform distribution . . . . .	11
3.2.2	Exponential distribution . . . . .	11
3.2.3	Normal distribution . . . . .	11
3.2.4	Gamma distribution and its particular cases . . . . .	13
3.2.5	Beta distribution . . . . .	14
3.3	Exercises . . . . .	15
<b>4</b>	<b>Multivariate distributions</b>	<b>16</b>
4.1	Joint distribution functions . . . . .	16
4.1.1	Marginal and conditional distributions . . . . .	16
4.1.2	Independent random variables . . . . .	17
4.2	Functions of random variables . . . . .	18

4.2.1	Expectation . . . . .	18
4.2.2	Covariance . . . . .	19
4.3	The bivariate normal distribution . . . . .	20
4.4	Conditional expectation and conditional variance* . . . . .	22
4.5	Distribution of a function of random variables . . . . .	23
4.5.1	Direct method . . . . .	23
4.5.2	Moment-generating functions . . . . .	25
4.5.3	Jacobians . . . . .	25
4.6	Exercises . . . . .	28
<b>5</b>	<b>Limit theorems</b>	<b>29</b>
5.1	Convergence of sequences of random variables . . . . .	29
5.2	Laws of large numbers . . . . .	29
5.3	The central limit theorem . . . . .	31
5.4	Exercises . . . . .	33
<b>A</b>	<b>Gamma and beta functions</b>	<b>34</b>
A.1	The gamma function . . . . .	34
A.2	The beta function . . . . .	36
A.3	The Gaussian integral . . . . .	37
<b>B</b>	<b>Probability spaces</b>	<b>38</b>
B.1	The family of events . . . . .	38
B.2	Probability measures . . . . .	39
B.3	General random variables . . . . .	40
B.4	A note on simulation of random variables . . . . .	42
	<b>References</b>	<b>44</b>

# 1 Basics of probability

See Mendenhall et al. [11, Chapter 2] or Lefebvre [4, Chapter 2].

## 2 Discrete random variables

In this section, we assumed that each sample space  $\Omega$  is endowed with a probability  $P$  defined on some  $\sigma$ -field  $\mathcal{F}$ . See Appendix B.

A random variable  $X : \Omega \rightarrow \mathbb{R}$  is **discrete** if  $R_X$  is at most countable.

Given a discrete r.v.  $X$ , define the **probability mass function (pmf)**

$$p_X(x) := P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\}), \quad x \in R_X.$$

Notice that

(a)  $p_X(x) \geq 0$  for each  $x$  and

(b)  $\sum_{x \in R_X} p_X(x) = 1$ .

**Remark 2.1.** Given a function  $\pi : S \rightarrow \mathbb{R}$ , where  $S$  is at most countable, that satisfies properties (a) and (b), we can find a discrete r.v.  $X : \Omega \rightarrow \mathbb{R}$  and a probability  $P$  such that  $S = R_X$  and

$$\pi(x) = P(X = x) \quad \forall x \in S.$$

◇

### 2.1 Mean, variance, and standard deviation

**Definition 2.2.** Let  $X : \Omega \rightarrow \mathbb{R}$  be a discrete r.v., define the **expectation (mean or expected value)** of  $X$  as

$$E(X) = \sum_{x \in R_X} x p_X(x),$$

whenever the series is absolutely convergent.

The r.v.  $X$  with values in  $\mathbb{N}$  and pmf

$$p_X(n) = \frac{1}{n(n+1)}, \quad n \in \mathbb{N},$$

does not have finite expectation.

**Theorem 2.3.** Let  $X : \Omega \rightarrow \mathbb{R}$  be a discrete r.v. and  $g : \mathbb{R} \rightarrow \mathbb{R}$ . Then the expectation of  $Y = g(X)$  is

$$E[Y] = \sum_{x \in R_X} g(x) p_X(x),$$

that is

$$\sum_{y \in R_Y} y p_Y(y) = \sum_{x \in R_X} g(x) p_X(x).$$

*Proof.* Notice that  $R_X = \cup_{y \in R_Y} \{x \in R_X \mid g(x) = y\}$  and

$$\{\omega \in \Omega \mid Y(\omega) = y\} = \bigcup_{x \in R_X, g(x)=y} \{\omega \in \Omega \mid X(\omega) = x\}, \quad y \in R_Y.$$

Then

$$\begin{aligned} \sum_{x \in R_X} g(x)p_X(x) &= \sum_{y \in R_Y} \left[ \sum_{x \in R_X, g(x)=y} g(x)P(X=x) \right] \\ &= \sum_{y \in R_Y} y \left[ \sum_{x \in R_X, g(x)=y} P(X=x) \right] \\ &= \sum_{y \in R_Y} yP(Y=y). \end{aligned}$$

□

**Proposition 2.4.** Let  $X : \Omega \rightarrow \mathbb{R}$  be a discrete r.v. and  $g_1, g_2 : \mathbb{R} \rightarrow \mathbb{R}$ .

(a) If  $X$  is constant, say  $X = c$ , then  $E(X) = c$ .

(b) If  $a \in \mathbb{R}$ , then  $E[aX] = aE(X)$ .

(c)  $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$

(d) If  $a, c \in \mathbb{R}$ , then  $E[aX + c] = aE(X) + c$ .

*Proof.* Parts (a) and (b) follow from the definition of expectation. Part (c) is a direct consequence of Theorem 2.3. Finally, (d) follows from (a), (b), and (c). □

The **variance**  $\sigma_X^2$  of the r.v.  $X$  is defined as

$$\sigma_X^2 := E[(X - \mu)^2]$$

where  $\mu_X = E(X)$ . The variance of  $X$  is also denoted  $\text{var}(X)$ .

**Proposition 2.5.** Let  $X : \Omega \rightarrow \mathbb{R}$  be a discrete r.v.

(a)  $\text{var}(X) = E(X^2) - \mu_X^2$ .

(b) If  $a \in \mathbb{R}$ , then  $\text{var}(aX) = a^2\text{var}(X)$ .

(c) If  $c \in \mathbb{R}$ , then  $\text{var}(X + c) = \text{var}(X)$ .

The **standard deviation**  $\sigma_X$  of a discrete r.v.  $X$  is defined as

$$\sigma_X = \sqrt{\text{var}(X)}.$$

**Theorem 2.6.** Let  $X$  be a discrete r.v.

(a) **(Markov's inequality).** If  $X \geq 0$  and  $a > 0$ , then

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

(b) If  $c \in \mathbb{R}$ ,  $\varepsilon > 0$ , and  $m > 0$ , then

$$P(|X - c| \geq \varepsilon) \leq \frac{E(|X - c|^m)}{\varepsilon^m}.$$

(c) **(Chebyshev's inequality)**. If  $\mu_X$  and  $\sigma_X^2$  are finite, then

$$P(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}.$$

*Proof.* (a) Since  $X \geq 0$ ,

$$E(X) \geq \sum_{x \in R_X, x \geq a} xp_X(x).$$

Then  $E(X) \geq aP(X \geq a)$  and hence Markov's inequality follows.

(b) Part (a) and the equality  $\{\omega \in \Omega \mid |X(\omega) - c|^m \geq \varepsilon^m\} = \{\omega \in \Omega \mid |X(\omega) - c| \geq \varepsilon\}$  yield (b).

(c) Chebyshev's inequality follows from (b) with  $m = 2$ . □

## 2.2 Common discrete distributions

### 2.2.1 Bernoulli distribution

Let  $S \subseteq \Omega$  be an event such that  $P(S) = p$ . A *Bernoulli trial* consists of two possible outcomes: *success*  $S$  or *failure*  $S^c$ . Define the **Bernoulli random variable**

$$B(\omega) = \begin{cases} 1 & \text{if } \omega \in S, \\ 0 & \text{if } \omega \in S^c. \end{cases}$$

Thus  $p_B(1) = p$  and  $p_B(0) = 1 - p$ . The distribution of  $B$  is called **Bernoulli distribution**; it is also said that  $B$  has Bernoulli distribution with parameter  $p$ . We use the notation

$$B \sim \text{Ber}(p).$$

Notice that

$$E(B) = p$$

and

$$\text{var}(B) = p(1 - p).$$

### 2.2.2 Binomial distribution

A *Binomial experiment* with parameters  $(n, p)$  consists of  $n$  independent Bernoulli trials with parameter  $p$ . A **Binomial random variable**  $Y$  gives the number of successes of a Binomial experiment. We use the notation

$$Y \sim \text{Bin}(n, p).$$

Then  $R_Y = \{0, 1, \dots, n\}$  and the probability mass function is

$$P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad 0 \leq k \leq n.$$

**Lemma 2.7.** Let  $m \in \mathbb{N}$  and  $x, y \in \mathbb{R}$ .

(a) **(Newton's Binomial)** For  $(x + y)^m = \sum_{k=0}^m \binom{m}{k} x^k y^{m-k}$ .

(b) For  $1 \leq k \leq m$ ,

$$k \binom{m}{k} = m \binom{m-1}{k-1}.$$

(c) For  $2 \leq k \leq m$ ,

$$k^2 \binom{m}{k} = m(m-1) \binom{m-2}{k-2} + k \binom{m}{k}.$$

*Proof.* Newton's Binomial is well known. Equalities (b) and (c) follow from direct calculations.  $\square$

Notice that  $\sum_{k=0}^n P(Y = k) = 1$  due to Newton's Binomial and the equality  $1 = [p + (1 - p)]^n$ .

**Proposition 2.8.** If  $Y \sim \text{Bin}(n, p)$ , then

$$E(Y) = np \quad \text{and} \quad \text{var}(Y) = np(1 - p).$$

*Proof.* It follows from Lemma 2.7. Proposition 3.6(a) is useful to compute the variance.  $\square$

### 2.2.3 Geometric distribution

Consider the number  $G$  of independent Bernoulli trials, with parameter  $p$ , until we obtain the *first success*. Thus

$$R_G = \{1, 2, 3, \dots\}$$

and  $G$  is called **Geometric random variable**. We use the notation  $G \sim \text{Geo}(p)$ . The probability mass function is given by

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

**Lemma 2.9.** Let  $0 < r < 1$ . Then

$$(a) \sum_{k=0}^{\infty} r^k = \frac{1}{1-r},$$

$$(b) \sum_{k=1}^{\infty} k r^{k-1} = \frac{1}{(1-r)^2}, \text{ and}$$

$$(c) \sum_{k=1}^{\infty} k^2 r^{k-1} = \frac{1+r}{(1-r)^3}.$$

**Proposition 2.10.** If  $G \sim \text{Geo}(p)$ , then

$$E(G) = \frac{1}{p} \quad \text{and} \quad \text{var}(G) = \frac{1-p}{p^2}.$$

*Proof.* It follows from Lemma 2.9.  $\square$

### 2.2.4 Poisson distribution

Let  $Y$  be a discrete r.v. such that

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

where  $\lambda > 0$ . This r.v. is said to have **Poisson distribution**, written as  $Y \sim \text{Poi}(\lambda)$ .

Poisson random variables are used to count the (random) number of events that occur in a given interval of time.

Recall that, for each  $\lambda \in \mathbb{R}$ ,

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}. \quad (1)$$

**Proposition 2.11.** *If  $Y \sim \text{Poi}(\lambda)$ , then*

$$E(Y) = \lambda \quad \text{and} \quad \text{var}(Y) = \lambda.$$

*Proof.* The expectation is easily obtained from (1). For the variance, we use (1) again to get

$$E[Y(Y-1)] = \lambda^2.$$

Thus  $E(Y^2) - \lambda = \lambda^2$  and hence  $\text{var}(Y) = \lambda$  because of Proposition 3.6(a).  $\square$

## 2.3 Moment-generating functions

The  $k$ -th moment about the origin of the r.v.  $X$  is given by

$$\mu_k := E(X^k)$$

and the  $k$ -th moment about the mean  $\mu$  of  $X$  is

$$E[(X - \mu)^k].$$

**Definition 2.12.** *Given a discrete r.v.  $X$ , the **moment-generating function**  $m_X$  of  $X$  is given by*

$$m_X(t) = E(e^{tX})$$

for each  $t$  such that  $E(e^{tX}) < \infty$ .

Observe that

$$\begin{aligned} E(e^{tX}) &= E\left(1 + tX + \frac{t^2 X^2}{2} + \frac{t^3 X^3}{3!} + \frac{t^4 X^4}{4!} + \dots\right) \\ &= 1 + tE(X) + \frac{t^2 E(X^2)}{2} + \frac{t^3 E(X^3)}{3!} + \frac{t^4 E(X^4)}{4!} + \dots \\ &= 1 + t\mu + \frac{t^2}{2}\mu_2 + \frac{t^3}{3!}\mu_3 + \frac{t^4}{4!}\mu_4 + \dots, \end{aligned} \quad (2)$$

whenever the function  $m_X$  is well defined on a neighborhood about the origin. Further, if  $m_X$  has derivatives  $m_X^{(k)}$  at  $t = 0$ , then we can compare the Taylor expansion of  $m_X$  and (2) to conclude that

$$\mu_k = m_X^{(k)}(0), \quad k = 1, 2, \dots$$

**Proposition 2.13.** Let  $B \sim \text{Bin}(n, p)$ ,  $G \sim \text{Geo}(p)$ , and  $X \sim \text{Poi}(\lambda)$ . Then

$$m_B(t) = (pe^t + 1 - p)^n, \quad t \in \mathbb{R},$$

$$m_G(t) = \frac{pe^t}{1 - (1 - p)e^t}, \quad (1 - p)e^t < 1,$$

and

$$m_X(t) = e^{\lambda(e^t - 1)}, \quad t \in \mathbb{R}.$$

*Proof.* The moment-generating functions are obtained by direct calculations.  $\square$

## 2.4 Exercises

Solve Exercises 3.15, 3.29, 3.37, 3.40, 3.41, 3.70, 3.71, 3.88, 3.130, 3.155 in Wackerly et al. [11].



### 3 Continuous random variables

Let  $X$  be a random variable, on the probability space  $(\Sigma, \mathcal{F}, P)$  (see Appendix B), and distribution  $F$ . We say that  $X$  is a **continuous random variable** if  $F$  is a continuous function.

In this section we deal with a subclass of continuous random variables for which there is an integrable **probability density function** (pdf) (or simply **density**)  $f : \mathbb{R} \rightarrow [0, \infty)$ , that is,

$$P(X \leq x) = \int_{-\infty}^x f(t)dt, \quad x \in \mathbb{R}.$$

In particular, by the Fundamental Theorem of Calculus,

$$F'(x) = f(x), \quad x \in \mathbb{R},$$

whenever  $f$  is continuous at  $x$ .

**Remark 3.1.** If the random variable  $X$  has a probability density function, then we say that  $X$  is **absolutely continuous**.

Given a r.v.  $X$  with a continuous density  $f$ , we have

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) \\ &= P(a < X < b) \\ &= \int_a^b f(x)dx \end{aligned}$$

for  $a < b$ .

#### 3.1 Mean and variance

**Definition 3.2.** Let  $X : \Omega \rightarrow \mathbb{R}$  be a r.v. with density  $f$ , define the **expectation (mean or expected value)** of  $X$  as

$$E(X) = \int_{\mathbb{R}} xf(x)dx,$$

whenever  $\int_{\mathbb{R}} |x|f(x)dx < \infty$ .

**Example 3.3.** The distribution of a r.v.  $X$  with density

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R},$$

is known as **Cauchy distribution**. The r.v.  $X$  has undefined expectation. ◇

A proof of Theorem 3.4 can be found in Rosenthal [8, Proposition 6.2.3]. The proof of Proposition 3.5 is analogous to that of Proposition 2.4.

**Theorem 3.4.** Let  $X : \Omega \rightarrow \mathbb{R}$  be a r.v. with density  $f$ . If  $h : \mathbb{R} \rightarrow \mathbb{R}$ , then

$$E[h(X)] = \int_{\mathbb{R}} h(x)f(x)dx,$$

whenever the integral is well defined.

**Proposition 3.5.** Let  $X : \Omega \rightarrow \mathbb{R}$  be a r.v. with density  $f$  and  $g_1, g_2 : \mathbb{R} \rightarrow \mathbb{R}$ . Assume the expectations below are well defined.

- (a) If  $X$  is constant, say  $X = c$ , then  $E(X) = c$ .
- (b) If  $a \in \mathbb{R}$ , then  $E[aX] = aE(X)$ .
- (c)  $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$
- (d) If  $a, c \in \mathbb{R}$ , then  $E[aX + c] = aE(X) + c$ .

The **variance**  $\sigma_X^2$  of the r.v.  $X$ , with density  $f$ , is defined as

$$\sigma_X^2 := E[(X - \mu_X)^2] = \int_{\mathbb{R}} (x - \mu_X)^2 f(x) dx$$

where  $\mu_X = E(X)$ . The variance of  $X$  is also denoted  $\text{var}(X)$ . The **standard deviation**  $\sigma_X$  of  $X$  is defined as

$$\sigma_X = \sqrt{\text{var}(X)}.$$

**Proposition 3.6.** Let  $X : \Omega \rightarrow \mathbb{R}$  be a continuous r.v., with density  $f$ . Assume the mean and variance of  $X$  are finite.

- (a)  $\text{var}(X) = E(X^2) - \mu_X^2$ .
- (b) If  $a \in \mathbb{R}$ , then  $\text{var}(aX) = a^2 \text{var}(X)$ .
- (c) If  $c \in \mathbb{R}$ , then  $\text{var}(X + c) = \text{var}(X)$ .

*Proof.* The assertions follow from properties of the integral. □

**Theorem 3.7.** Let  $X$  be a continuous r.v. with density  $f$ . Assume the expectations below are finite.

- (a) **(Markov's inequality).** If  $X \geq 0$  and  $a > 0$ , then

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

- (b) If  $c \in \mathbb{R}$ ,  $\varepsilon > 0$ , and  $m > 0$ , then

$$P(|X - c| \geq \varepsilon) \leq \frac{E(|X - c|^m)}{\varepsilon^m}.$$

- (c) **(Chebyshev's inequality).** If  $\mu_X$  and  $\sigma_X^2$  are finite, then

$$P(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}.$$

*Proof.* (a) Since  $X \geq 0$ ,

$$E(X) \geq \int_a^\infty xf(x)dx.$$

Then  $E(X) \geq aP(X \geq a)$  and hence Markov's inequality follows.

- (b) Part (a) and the equality  $\{\omega \in \Omega \mid |X(\omega) - c|^m \geq \varepsilon^m\} = \{\omega \in \Omega \mid |X(\omega) - c| \geq \varepsilon\}$  yield (b).

- (c) Chebyshev's inequality follows from (b) with  $m = 2$ . □

## 3.2 Common continuous distributions

### 3.2.1 Uniform distribution

The r.v.  $U$  has **uniform distribution** on the interval  $[a, b]$ , with  $a < b$ , if it has density of the form

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

We use the notation  $U \sim \text{Unif}([a, b])$ . Thus

$$F(u) = \begin{cases} 0 & \text{if } u < a, \\ \frac{u-a}{b-a} & \text{if } a \leq u \leq b, \\ 1 & \text{if } u > b. \end{cases}$$

Further,  $E(U) = (a + b)/2$  and  $\text{var}(U) = (b - a)^2/12$ . Finally, the moment-generating function is

$$m_U(t) = \begin{cases} \frac{e^{tb} - e^{ta}}{(b-a)t} & \text{if } t \neq 0, \\ 1 & \text{if } t = 0. \end{cases}$$

### 3.2.2 Exponential distribution

Let  $\beta > 0$ . If the r.v.  $Y$  has density

$$f(y) = \begin{cases} \beta e^{-\beta y} & \text{if } y \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

then we say that  $Y$  has **exponential distribution**, denoted  $Y \sim \text{Exp}(\beta)$ . The distribution becomes

$$F(y) = \begin{cases} 0 & \text{if } y < 0, \\ 1 - e^{-\beta y} & \text{if } y \geq 0. \end{cases}$$

Integrating by parts, we see that

$$E(Y) = \int_0^\infty e^{-\beta y} = \frac{1}{\beta}.$$

Further, integrating by parts again, we have  $E(Y^2) = 2/\beta^2$ . Hence

$$\text{var}(Y) = \frac{1}{\beta^2}.$$

The moment-generating function  $m_Y$  is defined for  $t < \beta$ ,

$$m_Y(t) = \frac{\beta}{\beta - t}.$$

### 3.2.3 Normal distribution

Let  $\mu \in \mathbb{R}$  and  $\sigma > 0$ . We say that  $Z$  has **normal distribution** with parameters  $(\mu, \sigma)$  whenever the density is

$$f(z) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( \frac{z - \mu}{\sigma} \right)^2 \right), \quad z \in \mathbb{R}. \quad (3)$$

We use the notation  $Z \sim N(\mu, \sigma^2)$ . In particular, the **standard normal distribution** happens when  $\mu = 0$  and  $\sigma = 1$ .

**Remark 3.8.** From the so-called Gaussian integral (see Theorem A.8)

$$\int_{-\infty}^{\infty} e^{-s^2} ds = \sqrt{\pi},$$

we can see that  $\int_{-\infty}^{\infty} f(z) dz = 1$ , where  $f$  is given by (3). In particular,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = 1.$$

◇

**Theorem 3.9.** Let  $Z \sim N(\mu, \sigma^2)$ . Then  $E(Z) = \mu$  and  $\text{var}(Z) = \sigma^2$ .

*Proof.* In order to compute the expectation, set the change of variable  $x = (z - \mu)/\sigma$  to obtain

$$E(Z) = \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + \int_{-\infty}^{\infty} \frac{\sigma}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}} dx.$$

The first integral equals  $\mu$ . For the second integral, notice that  $\int_0^{\infty} x e^{-\frac{x^2}{2}} dx$  is finite because

$$\begin{aligned} \int_0^b x e^{-\frac{x^2}{2}} dx &= 1 - e^{-b^2/2} \\ &\rightarrow 1 \end{aligned}$$

as  $b \rightarrow \infty$ . Since the integrand is an odd function,

$$\int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} dx = 0.$$

Then  $E(Z) = \mu$ .

We now compute the variance

$$\begin{aligned} \text{var}(Z) &= E[(Z - \mu)^2] \\ &= \int_{-\infty}^{\infty} (z - \mu)^2 \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{z - \mu}{\sigma}\right)^2\right) dz \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx, \end{aligned} \tag{4}$$

where the latter equality follows from the change of variable  $x = \frac{z - \mu}{\sigma}$ . On the other hand, integrate both sides of the equality

$$\frac{d}{dx}(x e^{-x^2/2}) = -x^2 e^{-x^2/2} + e^{-x^2/2}$$

on the interval  $[-b, b]$ , then let  $b \rightarrow \infty$  to obtain

$$0 = - \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx + \int_{-\infty}^{\infty} e^{-x^2/2} dx.$$

This equality of integrals and (4) imply (see Remark 3.8)

$$\text{var}(Z) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = \sigma^2.$$

□

**Theorem 3.10.** The moment-generating function  $m_Z$  of  $Z \sim N(\mu, \sigma^2)$  is given by

$$m_Z(t) = e^{\mu t + \sigma^2 t^2 / 2}, \quad t \in \mathbb{R}.$$

*Proof.* The conclusion follows from the equality

$$tz - \frac{1}{2} \left( \frac{z - \mu}{\sigma} \right)^2 = \mu t + \frac{\sigma^2 t^2}{2} - \frac{1}{2} \left( \frac{z - (\sigma^2 t + \mu)}{\sigma} \right)^2$$

which is a consequence of

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

□

### 3.2.4 Gamma distribution and its particular cases

A continuous r.v.  $Y$  has **gamma distribution** when its density is of the form

$$f(y) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} & \text{if } y \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where  $\Gamma$  is the function defined in Appendix A.1. We use the notation  $Y \sim \text{Gamma}(\alpha, \beta)$ . Notice that

$$\begin{aligned} \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} dy &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-\beta y} dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{1}{\beta^\alpha} t^{\alpha-1} e^{-t} dt \\ &= 1. \end{aligned}$$

where we have used the change of variable  $t = \beta y$ . Thus (5) defines a pdf.

**Remark 3.11.** A particular case of the gamma distribution happens when  $\alpha = 1$ , this is the exponential distribution. Another two particular cases are given in the following definition.

**Definition 3.12.** Let  $k \in \mathbb{N}$  and  $\beta > 0$ .

(a) The r.v.  $Y$  whose pdf is given by

$$f(y) = \begin{cases} \frac{\beta^k}{(k-1)!} y^{k-1} e^{-\beta y} & \text{if } y \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

is said to have **Erlang distribution** ( $\text{Gamma}(k, \beta)$ ).

- (b) The r.v.  $Y$  has **chi-squared distribution** with  $k$  degrees of freedom, written as  $Y \sim \chi_k^2$ , when the pdf is of the form

$$f(y) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} y^{k/2-1} e^{-y/2} & \text{if } y \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

This is a  $\text{Gamma}(k/2, 1/2)$  distribution.

**Theorem 3.13.** The moment-generating function  $m_Y$  of  $Y \sim \text{Gamma}(\alpha, \beta)$  is given by

$$m_Y(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha}, \quad t < \beta.$$

*Proof.* To compute the corresponding integral, let the change of variable  $u = (\beta - t)y$ , thus

$$\begin{aligned} m_Y(t) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-(\beta-t)y} dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)(\beta-t)^\alpha} \int_0^\infty u^{\alpha-1} e^{-u} du \\ &= \frac{\beta^\alpha}{(\beta-t)^\alpha}, \end{aligned}$$

whenever  $\beta - t > 0$ . □

**Theorem 3.14.** If  $Y \sim \text{Gamma}(\alpha, \beta)$ , then  $E(Y) = \alpha/\beta$  and  $\text{var}(Y) = \alpha/\beta^2$ .

*Proof.* By differentiating  $m_Y$ , we have  $m_Y'(0) = \alpha/\beta$  and  $m_Y''(0) = \alpha(\alpha+1)/\beta^2$ . Then  $E(Y) = \alpha/\beta$  and

$$\begin{aligned} \text{var}(Y) &= E(Y^2) - (E(Y))^2 \\ &= \alpha(\alpha+1)/\beta^2 - \alpha^2/\beta^2 \\ &= \alpha/\beta^2. \end{aligned}$$
□

### 3.2.5 Beta distribution

Let  $\alpha > 0$  and  $\beta > 0$ . The r.v.  $Y$  follows a **beta distribution** with parameters  $(\alpha, \beta)$  whenever its pdf has the form

$$f(y) = \begin{cases} \frac{1}{B(\alpha, \beta)} y^{\alpha-1} (1-y)^{\beta-1} & \text{if } 0 \leq y \leq 1, \\ 0 & \text{otherwise,} \end{cases} \quad (8)$$

where  $B$  is the beta function—see Appendix A.2. We use the notation  $Y \sim \text{Beta}(\alpha, \beta)$ .

Unfortunately, there is not a closed-form expression for the moment-generating function of the beta distribution.

**Proposition 3.15.** Let  $Y \sim \text{Beta}(\alpha, \beta)$ . Then

$$E(Y) = \frac{\alpha}{\alpha + \beta} \quad \text{and} \quad \text{var}(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

*Proof.* It follows from direct calculations and the relation

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

proved in Theorem A.7. □

### 3.3 Exercises

Solve Exercises 4.59, 4.62, 4.63, 4.79, 4.88, 4.96, 4.105, 4.123, 4.133, 4.137, and 4.146 in Wackerly et al. [11]. The following websites could be useful

<https://homepage.divms.uiowa.edu/~mbognar/>

[https://college.cengage.com/nextbook/statistics/wackerly\\_966371/student/html/](https://college.cengage.com/nextbook/statistics/wackerly_966371/student/html/)

## 4 Multivariate distributions

### 4.1 Joint distribution functions

**Definition 4.1.** Let  $X_1, \dots, X_n$  be random variables defined on the same probability space  $(\Omega, \mathcal{F}, P)$ . The **joint distribution**  $F : \mathbb{R}^n \rightarrow [0, 1]$  of the random vector  $(X_1, \dots, X_n)$  is defined as

$$F(x_1, x_2, \dots, x_n) := P(X_1 \leq x_1, \dots, X_n \leq x_n), \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

The random vector  $\mathbf{X} = (X_1, \dots, X_n)$  is said to be **discrete** if the set of values of  $\mathbf{X}$  is finite or countably infinite. In such a case, the properties of  $\mathbf{X}$  are determined by the **joint probability mass function** (or simply **joint probability**)

$$p(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}), \quad \mathbf{x} \in R_{\mathbf{X}},$$

where  $\mathbf{x} = (x_1, \dots, x_n)$ . In particular,

$$P(\mathbf{X} \in A) = \sum_{\mathbf{x} \in A} p(\mathbf{x}), \quad A \subseteq R_{\mathbf{X}}.$$

On the other hand, if  $f : \mathbb{R}^n \rightarrow [0, \infty)$  satisfies

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f(s_1, \dots, s_n) ds_1 \dots ds_n,$$

then  $f$  is the **joint density function** of  $(X_1, \dots, X_n)$ . In this case,

$$P(a_1 \leq X_1 \leq b_1, \dots, a_n \leq X_n \leq b_n) = \int_{a_n}^{b_n} \dots \int_{a_1}^{b_1} f(s_1, \dots, s_n) ds_1 \dots ds_n.$$

As in the univariate case, we have that  $p$  and  $f$  are nonnegative and, further,

$$\sum_{\mathbf{x} \in R_{\mathbf{X}}} p(\mathbf{x}) = 1$$

and

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(s_1, \dots, s_n) ds_1 \dots ds_n = 1.$$

#### 4.1.1 Marginal and conditional distributions

Let  $(X_1, X_2)$  be a discrete random pair with joint probability function  $p$ . Then the **marginal probability mass functions** of  $X_1$  and  $X_2$  respectively are

$$p_1(x_1) := \sum_{x_2 \in R_{X_2}} p(x_1, x_2), \quad x_1 \in R_{X_1},$$

and

$$p_2(x_2) := \sum_{x_1 \in R_{X_1}} p(x_1, x_2), \quad x_2 \in R_{X_2}.$$

Analogously, if  $(Y_1, Y_2)$  be a continuous random pair with joint density  $f$ . Then the **marginal density functions** of  $Y_1$  and  $Y_2$  respectively are

$$f_1(y_1) := \int_{-\infty}^{\infty} f(y_1, y_2) dy_2, \quad y_1 \in \mathbb{R},$$

and

$$f_2(y_2) := \int_{-\infty}^{\infty} f(y_1, y_2) dy_1, \quad y_2 \in \mathbb{R}.$$



**Remark 4.2.** For  $n \geq 3$ , the definitions of marginal probability mass/density functions are analogous. For instance, if  $(Y_1, Y_2, Y_3)$  has joint density  $f$ , then

$$f_3(y_3) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2, y_3) dy_1 dy_2, \quad y_3 \in \mathbb{R}.$$

◇

Let  $(X_1, X_2)$  be a discrete random pair with probability function  $p$  and its marginals  $p_1$  and  $p_2$ . Recall the *multiplicative law* for the events  $A$  and  $B$

$$P(A \cap B) = P(A)P(B | A),$$

whenever  $P(A) > 0$ . Consider the events  $[X_1 = x_1]$  and  $[X_2 = x_2]$ , then

$$p(x_1, x_2) = p_1(x_1)P(X_2 = x_2 | X_1 = x_1)$$

whenever  $P(X_1 = x_1) > 0$ . This equality motivates the following definition

**Definition 4.3.** Let  $(X_1, X_2)$  be a discrete random pair with probability function  $p$  and marginals  $p_1$  and  $p_2$ . Then the **conditional probability function** of  $X_2$  given  $X_1$  is

$$p(x_2 | x_1) := \frac{p(x_1, x_2)}{p_1(x_1)}, \quad x_2 \in R_{X_2},$$

whenever  $p_1(x_1) > 0$ . Analogously for  $p(x_1 | x_2)$ .

**Definition 4.4.** Let  $(Y_1, Y_2)$  be a continuous random pair with density  $f$  and marginals  $f_1$  and  $f_2$ . Then the **conditional density** of  $Y_1$  given  $Y_2 = y_2$  is

$$f(y_1 | y_2) := \begin{cases} \frac{f(y_1, y_2)}{f_2(y_2)} & \text{if } f_2(y_2) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Analogously for  $f(y_2 | y_1)$ .

**Remark 4.5.** With the conditional density  $f(\cdot | y_2)$  we can compute the **conditional distribution of  $Y_1$  given  $y_2$**

$$F(y_1 | y_2) = \int_{-\infty}^{y_1} f(s | y_2) ds, \quad y_1 \in \mathbb{R},$$

and the **conditional expectation of  $Y_1$  given  $y_2$**

$$E(Y_1 | y_2) = \int_{\mathbb{R}} y_1 f(y_1 | y_2) dy_1.$$

Similar expressions hold for discrete random variables.

◇

#### 4.1.2 Independent random variables

**Definition 4.6.** Let  $X_i$  have distribution  $F_i$ ,  $i = 1, \dots, n$ . Then  $X_1, \dots, X_n$  are **independent** iff

$$F(x_1, x_2, \dots, x_n) = F_1(x_1) \cdot F_2(x_2) \cdot \dots \cdot F_n(x_n) \quad \forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n,$$

where  $F$  is the joint distribution of the vector  $(X_1, \dots, X_n)$ .

An equivalent definition can be given in terms of densities or probability functions. Specifically, the random variables  $Y_1, \dots, Y_n$  are independent iff

$$f(y_1, \dots, y_n) = f_1(y_1) \cdot \dots \cdot f_n(y_n) \quad \forall (y_1, \dots, y_n) \in \mathbb{R}^n,$$

where  $f$  is the joint density of  $(Y_1, \dots, Y_n)$  and  $f_i$  is the marginal density of  $Y_i$  ( $1 \leq i \leq n$ ). Likewise, the random variables  $X_1, \dots, X_n$  are independent iff

$$p(x_1, \dots, x_n) = p_1(x_1) \cdot \dots \cdot p_n(x_n) \quad \forall (x_1, \dots, x_n) \in R_X,$$

where  $p$  is the joint probability of  $\mathbf{X} = (X_1, \dots, X_n)$  and  $p_i$  is the marginal probability of  $X_i$  ( $1 \leq i \leq n$ ).

## 4.2 Functions of random variables

### 4.2.1 Expectation

Let  $p$  be the joint probability of  $(X_1, \dots, X_n)$  and  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ . As in the univariate case, we can show that the expectation of  $g(X_1, \dots, X_n)$  is

$$E[g(\mathbf{X})] = \sum_{\mathbf{x} \in R_X} g(\mathbf{x})p(\mathbf{x}), \quad (9)$$

where  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbf{X} = (X_1, \dots, X_n)$ . The expectation (9) is well defined whenever the series (or sum) is absolutely convergent; in such a case, we also have

$$E[g(\mathbf{X})] = \sum_{x_n \in R_{X_n}} \dots \sum_{x_1 \in R_{X_1}} g(x_1, \dots, x_n)p(x_1, \dots, x_n),$$

and the sums can be computed in any order. A general version of Proposition 2.4(c) can be obtained with

$$g(x_1, \dots, x_n) = c_1 g_1(x_1) + \dots + c_n g_n(x_n).$$

**Theorem 4.7.** *Let  $(X_1, \dots, X_n)$  be a discrete random vector and  $g_i : \mathbb{R} \rightarrow \mathbb{R}$  ( $1 \leq i \leq n$ ). Then*

$$E[c_1 g_1(X_1) + \dots + c_n g_n(X_n)] = c_1 E[g_1(X_1)] + \dots + c_n E[g_n(X_n)],$$

*whenever all the expectations are finite.*

Next theorem is valid for general random variables; however we only prove it for the discrete case. A proof of the general statement can be found in Billingsley [3, p. 277].

**Theorem 4.8.** *Let  $X_1$  and  $X_2$  be discrete independent random variables. Then*

$$E[X_1 X_2] = E[X_1]E[X_2],$$

*whenever all the expectations are finite.*

*Proof.* Let  $p$  be the joint probability of  $(X_1, X_2)$ . Then

$$\begin{aligned} E[X_1 X_2] &= \sum_{x_1 \in R_{X_1}} \sum_{x_2 \in R_{X_2}} x_1 x_2 p(x_1, x_2) \\ &= \sum_{x_1 \in R_{X_1}} \sum_{x_2 \in R_{X_2}} x_1 x_2 p_1(x_1) p_2(x_2) \\ &= \left[ \sum_{x_1 \in R_{X_1}} x_1 p_1(x_1) \right] \left[ \sum_{x_2 \in R_{X_2}} x_2 p_2(x_2) \right] \\ &= E(X_1)E(X_2) \end{aligned}$$

where  $p_1$  and  $p_2$  are the marginals. □

The following lemma is also valid for general random variables—see Rosenthal [8, Proposition 3.2.3].

**Lemma 4.9.** *Let  $g_1, g_2 : \mathbb{R} \rightarrow \mathbb{R}$ . If  $X_1$  and  $X_2$  are independent, then  $g_1(X_1)$  and  $g_2(X_2)$  are independent.*

**Corollary 4.10.** *Let  $X_1$  and  $X_2$  be discrete independent random variables. If  $g_1, g_2 : \mathbb{R} \rightarrow \mathbb{R}$ , then*

$$E[g_1(X_1)g_2(X_2)] = E[g_1(X_1)]E[g_2(X_2)],$$

*whenever all the expectations are finite.*

As we mentioned above, the results in Theorem 4.8, Lemma 4.9 and Corollary 4.10 are also valid for continuous random variables. Further, these results still hold for  $n$  independent random variables.

**Theorem 4.11.** *Let  $X_1$  and  $X_2$  be independent random variables with moment-generating functions  $m_1$  and  $m_2$ , respectively. Then*

$$m_{X_1+X_2}(t) = m_1(t)m_2(t)$$

*whenever the above functions are well defined.*

*Proof.* The conclusion follows from the equality  $m_{X_1+X_2}(t) = E[e^{tX_1}e^{tX_2}]$  and Corollary 4.10. □

#### 4.2.2 Covariance

**Definition 4.12.** *Let  $Y_1$  and  $Y_2$  be random variables on the same probability space. Suppose the means  $\mu_1$  and  $\mu_2$  of  $Y_1$  and  $Y_2$ , respectively, are finite. The **covariance** of  $Y_1$  and  $Y_2$  is*

$$\text{cov}(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)]$$

*whenever the expectation in the right-hand side is finite. The **correlation coefficient**  $\rho$  of  $Y_1$  and  $Y_2$  is*

$$\rho := \frac{\text{cov}(Y_1, Y_2)}{\sigma_1\sigma_2}$$

*where  $\sigma_1$  and  $\sigma_2$  are the standard deviation of  $Y_1$  and  $Y_2$ , respectively. The random variables  $Y_1$  and  $Y_2$  are **uncorrelated** iff  $\text{cov}(Y_1, Y_2) = 0$ .*

A direct calculation shows the covariance is given by

$$\text{cov}(Y_1, Y_2) = E[Y_1Y_2] - \mu_1\mu_2 \tag{10}$$

**Corollary 4.13.** *If  $Y_1$  and  $Y_2$  are independent random variables, then*

$$\text{cov}(Y_1, Y_2) = 0.$$

*That is, independent random variables are uncorrelated.*

The converse of Corollary 4.13 does not hold in general as shown in Wackerly et al. [11, Example 5.24].

**Theorem 4.14.** Let  $X_i$  and  $Y_j$  be random variables for  $1 \leq i \leq n$  and  $1 \leq j \leq m$ . Then for constants  $a_i$  and  $b_j$ ,  $1 \leq i \leq n$  and  $1 \leq j \leq m$ ,

$$\text{cov} \left( \sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j \right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{cov}(X_i, Y_j),$$

provided that each covariance is finite.

**Corollary 4.15.** Let  $a_1, \dots, a_n$  be constants and let  $X_1, \dots, X_n$  be random variables. Then

$$\begin{aligned} \text{var} \left( \sum_{j=1}^n a_j X_j \right) &= \sum_{j=1}^n a_j^2 \text{var}(X_j) + \sum_{i \neq j} a_i a_j \text{cov}(X_i, X_j) \\ &= \sum_{j=1}^n a_j^2 \text{var}(X_j) + 2 \sum_{i < j} a_i a_j \text{cov}(X_i, X_j). \end{aligned}$$

**Remark 4.16.** Given the random variables  $X_1, \dots, X_n$ , its **covariance matrix**  $\Sigma$  is an  $n \times n$  symmetric matrix with entries

$$\Sigma_{ij} = \text{cov}(X_i, X_j), \quad 1 \leq i, j \leq n.$$

◇

**Remark 4.17.** Given the random variables  $X_1, \dots, X_n$ , its **joint moment-generating function** is defined as

$$m(t_1, \dots, t_n) = E[e^{t_1 X_1 + \dots + t_n X_n}]$$

whenever the expectation is finite. In matrix (vector) notation, put  $\mathbf{t} = (t_1, \dots, t_n)$  and  $\mathbf{X} = (X_1, \dots, X_n)$ , thus

$$m(\mathbf{t}) = E[e^{\mathbf{t}\mathbf{X}^\top}].$$

As in the univariate case, the moment-generating function determines the density. Further, we can compute  $E[X_i^k X_j^l]$  by letting  $(t_1, \dots, t_n) \rightarrow (0, \dots, 0)$  after computing the partial derivatives

$$\frac{\partial^{k+l} m}{\partial t_i^k \partial t_j^l}(t_1, \dots, t_n), \quad k, l = 0, 1, 2, \dots$$

◇

### 4.3 The bivariate normal distribution

Let  $\mu \in \mathbb{R}^n$  and let  $C$  be a positive definite and symmetric matrix. If  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  has density

$$f(\mathbf{x}) = \frac{1}{((2\pi)^n \det(C))^{1/2}} \exp \left( -\frac{1}{2}(\mathbf{x} - \mu)C^{-1}(\mathbf{x} - \mu)^\top \right), \quad \mathbf{x} \in \mathbb{R}^n,$$

then  $\mathbf{X}$  is said to have **multivariate normal distribution**. We use the notation

$$\mathbf{X} \sim N(\mu, C).$$

For simplicity, we only study the bivariate case. Let  $\sigma_1 > 0$ ,  $\sigma_2 > 0$ , and  $-1 < \rho < 1$ . Put

$$C = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Thus  $\det(C) = (1 - \rho^2)\sigma_1^2\sigma_2^2$  and

$$C^{-1} = \frac{1}{1 - \rho^2} \begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}.$$

Notice that

$$\begin{aligned} (\mathbf{x} - \mu)C^{-1}(\mathbf{x} - \mu)^\top &= \frac{1}{1 - \rho^2} \left[ \left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 - \frac{2\rho}{\sigma_1\sigma_2} (x_1 - \mu_1)(x_2 - \mu_2) + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \\ &= \frac{1}{1 - \rho^2} \left[ \left( \frac{x_1 - \mu_1}{\sigma_1} - \rho \frac{x_2 - \mu_2}{\sigma_2} \right)^2 + (1 - \rho^2) \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right], \quad (11) \end{aligned}$$

where  $\mu = (\mu_1, \mu_2)$ . Consider the change of variables

$$z_1 = \frac{x_1 - \mu_1}{\sigma_1} - \rho \frac{x_2 - \mu_2}{\sigma_2}, \quad z_2 = \frac{x_2 - \mu_2}{\sigma_2},$$

that is,

$$x_1 = \sigma_1 z_1 + \rho\sigma_1 z_2 + \mu_1, \quad x_2 = \sigma_2 z_2 + \mu_2, \quad (12)$$

whose Jacobian matrix is

$$\begin{bmatrix} \sigma_1 & \rho\sigma_1 \\ 0 & \sigma_2 \end{bmatrix}.$$

Then

$$\begin{aligned} \int \int_{\mathbb{R}^2} f &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \int \int_{\mathbb{R}^2} \exp \left[ \frac{-1}{2(1 - \rho^2)} (z_1^2 + (1 - \rho^2)z_2^2) \right] \sigma_1\sigma_2 dz_1 dz_2 \\ &= \left[ \frac{1}{\sqrt{1 - \rho^2} \sqrt{2\pi}} \int_{\mathbb{R}} \exp \left( -\frac{z_1^2}{2(1 - \rho^2)} \right) dz_1 \right] \left[ \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp \left( -\frac{z_2^2}{2} \right) dz_2 \right] \\ &= 1. \end{aligned}$$

**Theorem 4.18.** Let  $\sigma_1 > 0$ ,  $\sigma_2 > 0$ , and  $-1 < \rho < 1$ . Suppose  $\mathbf{X} \sim N(\mu, C)$  where  $\mu \in \mathbb{R}^2$  and

$$C = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Then the joint moment-generating function is

$$m(\mathbf{t}) = \exp \left( \mathbf{t}\mu^\top + \frac{1}{2} \mathbf{t}C\mathbf{t}^\top \right), \quad \mathbf{t} \in \mathbb{R}^2,$$

that is,

$$m(t_1, t_2) = \exp \left( t_1\mu_1 + t_2\mu_2 + \frac{1}{2} (\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2) \right).$$

*Proof.* By using the equality (11) and the change of variables (12), we have

$$\begin{aligned} E(e^{t_1 X_1 + t_2 X_2}) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int \int_{\mathbb{R}^2} e^{t_1 x_1 - \frac{1}{2(1-\rho^2)} \left( \frac{x_1 - \mu_1}{\sigma_1} - \rho \frac{x_2 - \mu_2}{\sigma_2} \right)^2 + t_2 x_2 - \frac{1}{2} \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2} dx_1 dx_2 \\ &= \frac{e^{t_1 \mu_1 + t_2 \mu_2}}{2\pi\sqrt{1-\rho^2}} \int \int_{\mathbb{R}^2} e^{t_1 \sigma_1 z_1 - \frac{z_1^2}{2(1-\rho^2)}} e^{(\rho t_1 \sigma_1 + t_2 \sigma_2) z_2 - \frac{1}{2} z_2^2} dz_1 dz_2. \end{aligned}$$

Consider now the identities

$$t_1 \sigma_1 z_1 - \frac{z_1^2}{2(1-\rho^2)} = \frac{1}{2}(1-\rho^2)t_1^2 \sigma_1^2 - \frac{1}{2} \left( \frac{z_1 - (1-\rho^2)t_1 \sigma_1}{\sqrt{1-\rho^2}} \right)^2$$

and

$$(\rho t_1 \sigma_1 + t_2 \sigma_2) z_2 - \frac{1}{2} z_2^2 = \frac{1}{2}(\rho t_1 \sigma_1 + t_2 \sigma_2)^2 - \frac{1}{2}(z_2 - \rho \sigma_1 t_1 - \sigma_2 t_2)^2$$

to obtain

$$E(e^{t_1 X_1 + t_2 X_2}) = \exp \left( t_1 \mu_1 + t_2 \mu_2 + \frac{1}{2}(\sigma_1^2 t_1^2 + 2\rho \sigma_1 \sigma_2 t_1 t_2 + \sigma_2^2 t_2^2) \right).$$

□

**Corollary 4.19.** Let  $\sigma_1 > 0$ ,  $\sigma_2 > 0$ , and  $-1 < \rho < 1$ . Suppose  $(X_1, X_2) \sim N(\mu, C)$  where  $\mu = (\mu_1, \mu_2)$  and

$$C = \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Then  $E(X_i) = \mu_i$ ,  $\text{var}(X_i) = \sigma_i^2$ ,  $i = 1, 2$ , and  $\text{cov}(X_1, X_2) = \rho \sigma_1 \sigma_2$ .

*Proof.* We only calculate the covariance, expectations and variances are similarly obtained (see Remark 4.17). Since

$$\frac{\partial^2 m}{\partial t_1 \partial t_2}(t_1, t_2) = m(t_1, t_2) [\rho \sigma_1 \sigma_2 + (\mu_1 + \sigma_1^2 t_1 + \rho \sigma_1 \sigma_2 t_2)(\mu_2 + \sigma_2^2 t_2 + \rho \sigma_1 \sigma_2 t_1)],$$

we have  $E(X_1 X_2) = \rho \sigma_1 \sigma_2 + \mu_1 \mu_2$ . Finally, by (10),  $\text{cov}(X_1, X_2) = \rho \sigma_1 \sigma_2$ . □

We can also show, by using (11), that the marginal density  $f_2(x_2)$  of  $X_2$  corresponds to an  $N(\mu_2, \sigma_2^2)$ . Likewise,  $X_1 \sim N(\mu_1, \sigma_1^2)$ .

#### 4.4 Conditional expectation and conditional variance\*

In Remark 4.5, we have introduced the *conditional expectation*  $E[Y_1 \mid y_2]$  given  $y_2$ ; we now extend that definition to  $E[g(Y_1) \mid y_2]$  as

$$E[g(Y_1) \mid y_2] := \int_{\mathbb{R}} g(y_1) f(y_1 \mid y_2) dy_1, \quad y_2 \in \mathbb{R},$$

where  $g : \mathbb{R} \rightarrow \mathbb{R}$  and  $f(\cdot \mid y_2)$  is given in Definition 4.4. The latter definition makes sense for continuous random variables whereas the conditional expectation for discrete random variables is

$$E[g(X_1) \mid x_2] := \sum_{x_1 \in R_{X_1}} g(x_1) p(x_1 \mid x_2), \quad x_2 \in R_{X_2}.$$

When  $E[g(Y_1) \mid y_2]$  is defined for each  $y_2$ , a function  $E[g(Y_1) \mid \cdot]$  is defined. In the discrete case, the function  $E[g(X_1) \mid \cdot]$  can be similarly defined.

**Definition 4.20.** Let  $Y_1$  and  $Y_2$  be random variables defined on  $\Omega$ . Suppose  $E[Y_1 | \cdot]$  is well defined. The random variable

$$E[Y_1 | Y_2] := E[Y_1 | \cdot] \circ Y_2,$$

defined on  $\Omega$ , is called the **conditional expectation of  $Y_1$  given  $Y_2$** .

**Theorem 4.21.** Let  $Y_1$  and  $Y_2$  be random variables defined on  $\Omega$ . Then

$$E(E[Y_1 | Y_2]) = E(Y_1)$$

provided that all the expectations exist.

Given the random variables  $Y_1$  and  $Y_2$ , the **conditional variance of  $Y_1$  given  $y_2$**  is

$$V[Y_1 | y_2] := E[Y_1^2 | y_2] - (E[Y_1 | y_2])^2.$$

Likewise, the **conditional variance of  $Y_1$  given  $Y_2$**  is the random variable

$$V[Y_1 | Y_2] := V[Y_1 | \cdot] \circ Y_2.$$

**Theorem 4.22.** Let  $Y_1$  and  $Y_2$  be random variables defined on  $\Omega$ . Then

$$\text{var}(E[Y_1 | Y_2]) + E(V[Y_1 | Y_2]) = \text{var}(Y_1)$$

provided that all the expectations exist.

## 4.5 Distribution of a function of random variables

Given the distribution or density of a random variable (vector)  $X$  and a transformation  $h(X)$ , we aim to find the distribution or density of  $Y = h(X)$ . We discuss three methods.

### 4.5.1 Direct method

This method is usually applied to continuous random variables. It relies on the properties of the integral. The key step consists in identifying the region  $h(X) \leq y$  in the space  $(x_1, \dots, x_n)$  as well as the support of the (joint) density of  $X$ .

**Example 4.23** (Ash [2]). Let  $X$  have density

$$f_X(x) := \begin{cases} \frac{1}{2} & \text{if } -1 \leq x < 0, \\ \frac{1}{2}e^{-x} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Find the distribution function  $F_Y$  and the density  $f_Y$  of  $Y = X^2$ .

*Solution.* Consider the cases  $y < 0$ ,  $0 \leq y \leq 1$ , and  $y > 1$ . For the second case, observe that

$$\begin{aligned} P(Y \leq y) &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= P(-\sqrt{y} \leq X < 0) + P(0 \leq X \leq \sqrt{y}) \end{aligned}$$

and so

$$F_y(y) = \frac{1}{2} \sqrt{y} + \int_0^{\sqrt{y}} \frac{1}{2} e^{-x} dx.$$

The third case is analogous. Therefore

$$F_y(y) := \begin{cases} 0 & \text{if } y < 0, \\ \frac{1}{2} \sqrt{y} + \frac{1}{2}(1 - e^{-\sqrt{y}}) & \text{if } 0 \leq y \leq 1, \\ \frac{1}{2} + \frac{1}{2}(1 - e^{-\sqrt{y}}) & \text{if } y > 1, \end{cases}$$

and, by differentiation,

$$f_y(y) := \begin{cases} 0 & \text{if } y < 0, \\ \frac{1}{4\sqrt{y}} + \frac{1}{4\sqrt{y}} e^{-\sqrt{y}} & \text{if } 0 < y < 1, \\ \frac{1}{4\sqrt{y}} e^{-\sqrt{y}} & \text{if } y > 1. \end{cases}$$

◇

**Example 4.24** (Wackerly et al. [11]). Let  $X_1$  and  $X_2$  be independent and identically distributed random variables. Suppose  $X_i \sim \text{Unif}[0, 1]$  for  $i = 1, 2$ . Find the distribution of  $Y = X_1 + X_2$ .

*Solution.* Since  $X_1$  and  $X_2$  are independent, the joint density is

$$f(x_1, x_2) = f_1(x_1)f_2(x_2) = 1 \quad \forall (x_1, x_2) \in [0, 1] \times [0, 1].$$

Notice that  $0 \leq Y \leq 2$ . After plotting the regions  $x_1 + x_2 \leq y$  (for  $0 \leq y \leq 2$ ) and the support of  $(X_1, X_2)$ , we consider four cases  $y < 0$ ,  $0 \leq y < 1$ ,  $1 \leq y < 2$ , and  $y \geq 2$ .

For  $0 \leq y < 1$ ,

$$F_y(y) = \int_0^y \int_0^{y-x_2} 1 dx_1 dx_2 = \frac{y^2}{2}.$$

On the other hand, when  $1 \leq y < 2$ , it is convenient to calculate  $P(X_1 + X_2 \leq y)$  as follows

$$F_y(y) = 1 - \int_{y-1}^1 \int_{y-x_2}^1 1 dx_1 dx_2 = -1 + 2y - \frac{y^2}{2}.$$

Therefore,

$$F_y(y) := \begin{cases} 0 & \text{if } y < 0, \\ \frac{y^2}{2} & \text{if } 0 \leq y < 1, \\ -1 + 2y - \frac{y^2}{2} & \text{if } 1 \leq y < 2, \\ 1 & \text{if } y \geq 2. \end{cases}$$

◇

**Example 4.25.** Let  $X \sim N(0, 1)$ . Find the distribution of  $Y = \sigma X + \mu$ , where  $\mu \in \mathbb{R}$  and  $\sigma > 0$ .

*Solution.* Recall that  $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$  for  $x \in \mathbb{R}$ . Since  $Y(\omega) \leq y$  iff  $X(\omega) \leq (y - \mu)/\sigma$ , thus

$$F_Y(y) = F_X((y - \mu)/\sigma), \quad y \in \mathbb{R}.$$



Further, differentiate both sides to obtain

$$f_Y(y) = f_X((y - \mu)/\sigma) \frac{1}{\sigma} = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y - \mu)^2}.$$

Therefore,  $Y \sim N(\mu, \sigma^2)$ .

◇

### 4.5.2 Moment-generating functions

This method is an application of Remark B.14.

**Example 4.26.** Suppose  $X \sim N(\mu, \sigma^2)$ . Show that  $(X - \mu)/\sigma$  follows a standard normal distribution.

*Solution.* It follows from direct calculations and Remark B.14.

◇

**Example 4.27.** Suppose  $X_1$  and  $X_2$  are independent. Let  $X_1 \sim \text{Poi}(\lambda_1)$  and  $X_2 \sim \text{Poi}(\lambda_2)$ . Find the distribution of  $Y = X_1 + X_2$ . What is the distribution of  $Y$  if  $X_1 \sim N(\mu_1, \sigma_1^2)$  and  $X_2 \sim N(\mu_2, \sigma_2^2)$ ?

*Solution.* By Theorem 4.11,

$$m_Y(t) = m_{X_1}(t)m_{X_2}(t).$$

Thus

$$\begin{aligned} m_Y(n) &= e^{\lambda_1(e^t - 1)} e^{\lambda_2(e^t - 1)} \\ &= e^{(\lambda_1 + \lambda_2)(e^t - 1)} \end{aligned}$$

and hence  $X_1 + X_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$ . Similar arguments show that

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

when  $X_1 \sim N(\mu_1, \sigma_1^2)$  and  $X_2 \sim N(\mu_2, \sigma_2^2)$ , provided that  $X_1$  and  $X_2$  are independent.

◇

**Example 4.28.** Suppose  $X_1, \dots, X_n$  are independent and identically distributed (i.i.d.). Let  $X_1 \sim \text{Exp}(\beta)$ . Show that  $Y = X_1 + \dots + X_n$  follows an Erlang distribution.

*Solution.* By Theorem 4.11,  $m_Y(t) = \beta^n / (\beta - t)^n$ , that is,

$$m_Y(t) = \left(1 - \frac{t}{\beta}\right)^{-n}$$

which corresponds to a  $\text{Gamma}(n, \beta)$  distribution (see Theorem 3.13). By Definition 3.12(a),  $\text{Gamma}(n, \beta)$  is also known as Erlang distribution.

◇

### 4.5.3 Jacobians

Given the (joint) density  $f_X$ , this method aims to find the density of  $Y = h(X)$ .

Let us consider first the univariate case. Suppose that  $h$  is strictly increasing on the support of  $f_X$ . Then the inverse  $h^{-1}$  exists and

$$\begin{aligned} F_Y(y) &= P(h(X) \leq y) \\ &= P(X \leq h^{-1}(y)) \\ &= F_X(h^{-1}(y)). \end{aligned}$$

Therefore, by differentiating with respect to  $y$ , we have

$$f_Y(y) = f_X(h^{-1}(y)) \frac{dh^{-1}}{dy}(y),$$

whenever  $h^{-1}$  is differentiable. The above arguments also hold when  $h$  is strictly decreasing ( $y_1 < y_2$  imply  $h(y_1) > h(y_2)$ ) but the formula becomes

$$f_Y(y) = -f_X(h^{-1}(y)) \frac{dh^{-1}}{dy}(y).$$

For either case,

$$f_Y(y) = f_X(h^{-1}(y)) \left| \frac{dh^{-1}}{dy}(y) \right|,$$

whenever  $h$  is strictly monotone and  $h^{-1}$  is differentiable.

**Example 4.29.** Let  $X$  have density

$$f_X(x) := \begin{cases} 2x & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the density of  $Y = -aX + b$ , where  $a > 0$ .

*Solution.* Let  $h(x) = -ax + b$ . Then  $h^{-1}(y) = (b - y)/a$  and

$$\left| \frac{dh^{-1}}{dy}(y) \right| = \frac{1}{a}.$$

Hence

$$f_Y(y) := \begin{cases} \frac{2}{a^2}(b - y) & \text{if } b - a \leq y \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

It should be noted that the support of  $f_Y$  is the image  $h([0, 1])$ . ◇

### The multivariate case

Let  $f_X$  be the joint density of  $\mathbf{X} = (X_1, X_2)$ . Suppose that  $h : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is a 1-1 mapping on the support of  $f_X$  and, further,

$$h^{-1}(y_1, y_2) = (h_1^{-1}(y_1, y_2), h_2^{-1}(y_1, y_2))$$

is differentiable. Consider the random vector  $\mathbf{Y} = h(\mathbf{X})$ . Then

$$f_Y(y_1, y_2) = f_X(h_1^{-1}(y_1, y_2), h_2^{-1}(y_1, y_2)) \cdot |\det(Dh^{-1}(y_1, y_2))|$$

where  $Dh^{-1}$  is the Jacobian matrix

$$\begin{bmatrix} \frac{\partial h_1^{-1}}{\partial y_1} & \frac{\partial h_1^{-1}}{\partial y_2} \\ \frac{\partial h_2^{-1}}{\partial y_1} & \frac{\partial h_2^{-1}}{\partial y_2} \end{bmatrix}.$$

**Example 4.30** ( $t$  distribution, Ash [2]). Suppose  $X_1$  and  $X_2$  are independent. Let  $X_1 \sim N(0, 1)$  and  $X_2 \sim \chi_k^2$ . Find the density of  $\frac{X_1}{\sqrt{X_2}/\sqrt{k}}$ .

*Solution.* Let  $h_1(x_1, x_2) = \sqrt{k}x_1/\sqrt{x_2}$  and  $h_2(x_1, x_2) = x_2$  (this is common trick!) for  $x_1 \in \mathbb{R}$  and  $x_2 > 0$ . Thus

$$h^{-1}(y_1, y_2) = (y_1 \sqrt{y_2}/\sqrt{k}, y_2), \quad y_1 \in \mathbb{R}, y_2 > 0,$$

and

$$Dh^{-1} = \begin{bmatrix} \sqrt{y_2/k} & y_1/(2\sqrt{ky_2}) \\ 0 & 1 \end{bmatrix}.$$

Since  $X_1$  and  $X_2$  are independent, its joint density is

$$f_X(x_1, x_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2} \cdot \frac{1}{2^{k/2}\Gamma(k/2)} x_2^{k/2-1} e^{-x_2/2}, \quad x_1 \in \mathbb{R}, x_2 > 0.$$

Then

$$f_Y(y_1, y_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_1^2 y_2}{2k}} \cdot \frac{1}{2^{k/2}\Gamma(k/2)} y_2^{k/2-1} e^{-y_2/2} \cdot \sqrt{y_2/k}, \quad y_1 \in \mathbb{R}, y_2 > 0.$$

The density of  $Y_1$  is given by

$$\begin{aligned} \int_0^\infty f_Y(y_1, y_2) dy_2 &= \frac{1}{\sqrt{2\pi k} \Gamma(k/2) 2^{k/2}} \int_0^\infty y_2^{\frac{k+1}{2}-1} e^{-\frac{y_2}{2}(1+\frac{y_1^2}{k})} dy_2 \\ &= \left(1 + \frac{y_1^2}{k}\right)^{-\frac{k+1}{2}} \frac{1}{\sqrt{k\pi} \Gamma(k/2)} \int_0^\infty z^{\frac{k+1}{2}-1} e^{-z} dz \\ &= \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi} \Gamma(\frac{k}{2})} \left(1 + \frac{y_1^2}{k}\right)^{-\frac{k+1}{2}}, \quad y_1 \in \mathbb{R}, \end{aligned}$$

where we have set the change of variable  $z = \frac{y_2}{2}(1 + \frac{y_1^2}{k})$ . Therefore, the density of  $T := \frac{X_1}{\sqrt{X_2}/\sqrt{k}}$  is

$$\frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi} \Gamma(\frac{k}{2})} \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}, \quad t \in \mathbb{R}. \quad (13)$$

The r.v.  $T$  is said to have  $t$  **distribution** (or **Student's  $t$ -distribution**) with  $k$  **degrees of freedom**.  $\diamond$

**Example 4.31** ( $F$  distribution, Ash [2]). Suppose  $X_1$  and  $X_2$  are independent. Let  $X_1 \sim \chi_m^2$  and  $X_2 \sim \chi_n^2$ . Find the densities of  $\frac{X_1}{X_2}$  and  $\frac{X_1/m}{X_2/n}$ .

*Solution.* Let  $h_1(x_1, x_2) = x_1/x_2$  and  $h_2(x_1, x_2) = x_2$ , for  $x_1 > 0$  and  $x_2 > 0$ . We can show that the density of  $Y_1 = \frac{X_1}{X_2}$  is

$$\int_0^\infty f_{X_1}(y_1 y_2) f_{X_2}(y_2) y_2 dy_2.$$

After replacing the densities given in Definition 3.12(b) and letting  $z = y_2(1 + y_1)/2$ , we obtain

$$f_{Y_1}(z) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \frac{z^{\frac{m}{2}-1}}{(1+z)^{\frac{m+n}{2}}}, \quad z > 0.$$

Finally, let  $W := \frac{X_1/m}{X_2/n}$ . Then  $W = (n/m)Y_1$  and hence

$$f_W(w) = \frac{m}{n} f_{Y_1}(mw/n), \quad w > 0.$$

The r.v.  $W$  is said to follow an  $F$  **distribution with  $m$  and  $n$  degrees of freedom**.  $\diamond$

## 4.6 Exercises

Solve Exercises 4.111, 4.112, 5.11, 5.22, 5.26, 5.55, 5.65, 5.75, 5.86, 5.89, 5.91, 5.139, 5.142(a), 5.167, 6.7, 6.18, 6.40 and 6.64 in Wackerly et al. [11].

## 5 Limit theorems

### 5.1 Convergence of sequences of random variables

Throughout this section  $X_n, X : \Omega \rightarrow \mathbb{R}$  ( $n = 1, 2, \dots$ ) denote random variables on some probability space  $(\Omega, \mathcal{S}, \mathbb{P})$ .

**Definition 5.1.** *The sequence  $\{X_n\}$  converges*

(a) **almost surely** to  $X$ , denoted  $X_n \xrightarrow{a.s.} X$ , iff

$$\mathbb{P} \left( \{w \in \Omega \mid \lim_{n \rightarrow \infty} X_n(w) = X(w)\} \right) = 1,$$

(b) **in distribution** to  $X$ , denoted  $X_n \xrightarrow{d} X$ , iff

$$\lim_{n \rightarrow \infty} F_n(x) = F(x),$$

for every  $x$  at which  $F$  is continuous, where  $F_n$  and  $F$  are the distribution functions of  $X_n$  and  $X$ , respectively.

(c) **to  $X$  in the mean of order  $p$** , denoted  $X_n \xrightarrow{L^p} X$ , iff

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^p) = 0,$$

whenever  $p \in [1, \infty)$ ,  $\mathbb{E}(|X|^p) < \infty$ , and  $\mathbb{E}(|X_n|^p) < \infty$  ( $n = 1, 2, \dots$ ).

### 5.2 Laws of large numbers

**Theorem 5.2** (Weak Law of Large Numbers). *Let  $\{X_n\}$  be a sequence of independent random variables such that  $\mathbb{E}(X_n) = \mu$  and  $\text{var}(X_n) \leq M$  for every  $n \geq 1$ . Then*

$$\bar{X}_n := \frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{L^2} \mu.$$

*Proof.* Observe that  $\mathbb{E}[(\bar{X}_n - \mu)^2] = \text{var}(\bar{X}_n) \leq n^{-1}M$  which implies the conclusion.  $\square$

**Remark 5.3.** The conclusion of Theorem 5.2 is equivalent to

$$\mathbb{E}[\bar{X}_n^2] \rightarrow \mu^2.$$

**Theorem 5.4** (Strong Law of Large Numbers). *Let  $\{X_n\}$  be a sequence of i.i.d. random variables such that  $\mathbb{E}(|X_1|) < \infty$ . Then*

$$\bar{X}_n := \frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{a.s.} \mathbb{E}(X_1).$$

*Proof.* See Billingsley [3, Theorem 22.1] or Rosenthal [8, Theorem 5.4.4].  $\square$

**Example 5.5** (Monte Carlo Simulation). Suppose we want to compute the integral  $I = \int_0^1 g(x)dx$ . Note that

$$I = E[g(U)]$$

where  $U \sim \text{Unif}[0, 1]$ . On the other hand, let  $U_i$  ( $i = 1, \dots, n$ ) be i.i.d. uniform random variables on the interval  $[0, 1]$  and consider

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n g(U_i)$$

If  $g(x) = \sin(x^2)$ , then the random variable

$$\bar{Y}_n(\omega) = \frac{1}{n} \sum_{i=1}^n \sin(U_i^2)$$

can be simulated, for  $n = 100000$ , with the MATLAB code

```
mean(sin(rand(1,100000).^2))
```

We can observe that  $\bar{Y}_n(\omega)$  approximates 0.31, that is,

$$\int_0^1 \sin(x^2)dx \approx 0.31.$$

By the Strong Law of Large Numbers, we conclude that  $\bar{Y}_n$  indeed converges almost surely to  $I$ .  $\diamond$

**Example 5.6.** In order to illustrate the Weak Law of Large Numbers, in contrast to the latter example, we need to simulate  $\bar{X}_n(\omega)$  for different values of  $\omega$ . Consider the following matrix of simulated values

$$\begin{bmatrix} X_1(\omega_1) & X_2(\omega_1) & \dots & X_n(\omega_1) \\ X_1(\omega_2) & X_2(\omega_2) & \dots & X_n(\omega_2) \\ \vdots & \vdots & \ddots & \vdots \\ X_1(\omega_m) & X_2(\omega_m) & \dots & X_n(\omega_m) \end{bmatrix}.$$

Then  $\bar{X}_n(\omega_j)$  is the average of the the  $j$ -th row,  $1 \leq j \leq m$ , and  $m$  values of the r.v.  $\bar{X}_n$  can be arranged in a vector

$$\begin{bmatrix} \bar{X}_n(\omega_1) \\ \bar{X}_n(\omega_2) \\ \vdots \\ \bar{X}_n(\omega_m) \end{bmatrix}.$$

As mentioned in Remark 5.3, we need to show that  $E[\bar{X}_n^2] \rightarrow E(X_1)^2$ .

Consider the following concrete example. Let  $X_1, X_2, X_3, \dots$  be a sequence of i.i.d. random variables, where  $X_1 \sim \text{Exp}(\beta)$ . The second moment  $E[\bar{X}_n^2]$  of  $\bar{X}_n$ , with  $m$  different values of  $\omega$ , can be simulated in MATLAB as follows

```
MeanExp = 2;           % beta=0.5
n= 10000;              % number of rvs
m= 50;                 % number of w's
```

```

X = exprnd(MeanExp,m,n); % (m,n) matrix
overlineXnOmega = mean(X,2); % (m,1) vector
SecondMOverlineXn = mean(overlineXnOmega.^2)

```

We observe that  $E[\overline{X}_n^2]$  approximates  $1/\beta^2$  and hence  $\text{var}(\overline{X}_n) \approx 0$ .  $\diamond$

### 5.3 The central limit theorem

**Theorem 5.7** (Central Limit Theorem (CLT)). *Let  $\{X_k\}$  be a sequence of i.i.d. random variables with finite mean  $\mu$  and finite variance  $\sigma^2$ . Then*

$$\frac{S_n - n\mu}{\sigma \sqrt{n}} \xrightarrow{d} Z,$$

where  $S_n := X_1 + \dots + X_n$  and  $Z$  is a standard normal distributed random variable.

**Remark 5.8.** Recall that any real sequence  $y_n \rightarrow y$  implies  $(1 + n^{-1}y_n)^n \rightarrow e^y$ . This fact will be used in the following proof.  $\diamond$

*Sketch of the proof.* Notice that

$$\frac{S_n - n\mu}{\sigma \sqrt{n}} = \frac{1}{\sqrt{n}}(\tilde{X}_1 + \dots + \tilde{X}_n), \quad (14)$$

where

$$\tilde{X}_k := \frac{X_k - \mu}{\sigma}, \quad k = 1, \dots, n.$$

Moreover, for each  $k$ ,  $\mathbb{E}(\tilde{X}_k) = 0$  and  $\mathbb{E}(\tilde{X}_k^2) = 1$ , then the moment-generating function of  $\tilde{X}_k$  can be written as

$$m_{\tilde{X}_k}(t) = 1 + \frac{t^2}{2} + h(t), \quad t \in \mathbb{R},$$

see (2). Since  $m_{\tilde{X}_k}(0) = 1$ , we can use the L'Hôpital's rule to compute

$$\lim_{t \rightarrow 0} \frac{m_{\tilde{X}_k}(t) - 1}{t^2} = \lim_{t \rightarrow 0} \frac{m_{\tilde{X}_k}''(t)}{2} = \frac{1}{2}.$$

and hence

$$\lim_{t \rightarrow 0} h(t)/t^2 = 0. \quad (15)$$

By (14), the moment-generating function of  $(\sigma \sqrt{n})^{-1}(S_n - n\mu)$  is

$$\left[ m_{\tilde{X}_k}(t/\sqrt{n}) \right]^n = \left[ 1 + \frac{t^2}{2n} + h(t/\sqrt{n}) \right]^n, \quad t \in \mathbb{R}.$$

Observe that

$$\frac{t^2}{2n} + h(t/\sqrt{n}) = \frac{1}{n} \left( \frac{1}{2} + \frac{h(t/\sqrt{n})}{t^2/n} \right) t^2$$

and, by (15),

$$\lim_{n \rightarrow \infty} \left( \frac{1}{2} + \frac{h(t/\sqrt{n})}{t^2/n} \right) t^2 = \frac{t^2}{2}.$$

Then—see Remark 5.8—the sequence of moment-generating functions of  $(\sigma \sqrt{n})^{-1}(S_n - n\mu)$  converges to  $e^{-t^2/2}$  which is the moment-generating function of a standard normal r.v. This implies that the distributions of  $(\sigma \sqrt{n})^{-1}(S_n - n\mu)$  converge to a standard normal distribution.  $\square$

**Example 5.9.** We can use the following MATLAB code to compare the distribution  $F$  of a  $N(0, 1)$  r.v. with the distributions of  $\frac{S_n - n\mu}{\sigma \sqrt{n}}$ , where  $X_1 \sim \text{Poi}(\lambda)$ .

```
lambda=3;
sigma=sqrt(lambda);
n= 50;           %number of rvs
m= 1000;         %number of  $\omega$ 's

Xmatrix = poissrnd(lambda,m,n);
Sn = sum(Xmatrix,2);
Yn = (Sn - n*lambda)./(sigma*sqrt(n));
cdfplot(Yn)
hold on

x = linspace(min(Yn),max(Yn));
plot(x,normcdf(x,0,1))
```

$\diamond$

**Example 5.10.** We now consider  $\text{Gamma}(\alpha, \beta)$  i.i.d. random variables and plot the histogram of  $\frac{S_n - n\mu}{\sigma \sqrt{n}}$  along with the density of a normal distribution.

```
alpha=3;
beta=1;
mu=alpha/beta;
sigma=sqrt(alpha)/beta;
n= 100;           %number of rvs
m= 1000;         %number of  $\omega$ 's

Xmatrix = gamrnd(alpha,1/beta,m,n);
Sn = sum(Xmatrix,2);
Yn = (Sn - n*mu)./(sigma*sqrt(n));
histfit(Yn)
```

$\diamond$

**Remark 5.11.** If the random variables  $X_k$  ( $k = 1, 2, \dots$ ) in the CLT are Beornulli (with parameter  $p$ ), then the sum  $S_n$  has binomial distribution and

$$\frac{S_n - np}{\sqrt{npq}} \xrightarrow{d} Z.$$

This special case is known as **de Moivre–Laplace Theorem**.  $\diamond$



## 5.4 Exercises

5.1 Use stochastic simulation to approximate the value of

$$I = 4 \int_0^1 \sqrt{1 - x^2} dx.$$

Based on a geometric interpretation of the integral, what is the value of  $I$ ?

5.2 Illustrate the de Moivre–Laplace Theorem by simulation.

## A Gamma and beta functions

### A.1 The gamma function

**Definition A.1.** For each positive real number  $x$ , the **gamma function** is given by

$$\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt. \quad (16)$$

To see why the improper Riemann integral (16) exists, notice first that

$$t^{x-1} e^{-t} \leq t^{x-1} \quad 0 < t \leq 1.$$

On the other hand, for each  $x > 0$ ,  $t^{x-1} e^{-t/2} \rightarrow 0$  as  $t \rightarrow \infty$ , thus

$$t^{x-1} e^{-t} \leq M_x e^{-t/2}, \quad t \geq 1,$$

for some constant  $M_x$ . Therefore the integral (16) is finite for each  $x > 0$ .

**Lemma A.2.** Let  $p, q \in (1, \infty)$ . If  $\frac{1}{p} + \frac{1}{q} = 1$ , then for each  $\alpha, \beta \geq 0$ ,

$$\alpha\beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q}. \quad (17)$$

*Proof.* Notice that  $h : [0, \infty) \rightarrow \mathbb{R}$  given by

$$h(x) = (1-t) + tx - x^t, \quad 0 < t < 1,$$

attains its minimum at  $x = 1$ . Then

$$0 \leq (1-t) + tx - x^t, \quad \forall x > 0.$$

In particular, for  $x = a/b$ ,

$$a^t b^{1-t} \leq ta + (1-t)b.$$

The conclusion of the lemma follows from the change of variables  $t = 1/p$ ,  $\alpha = a^t$ , and  $\beta = b^{1-t}$ .  $\square$

**Lemma A.3** (Hölder's inequality). Let  $f, g : [a, b] \rightarrow \mathbb{R}$  be continuous functions. Then

$$\left| \int_a^b fg \right| \leq \left( \int_a^b |f|^p \right)^{1/p} \left( \int_a^b |g|^q \right)^{1/q}. \quad (18)$$

*Proof.* If  $f \equiv 0$  or  $g \equiv 0$ , then the inequality trivially holds. When  $f$  and  $g$  are not identically zero, we have

$$\|f\|_p := \left( \int_a^b |f|^p \right)^{1/p} > 0$$

and

$$\|g\|_q := \left( \int_a^b |g|^q \right)^{1/q} > 0.$$

By Lemma A.2

$$\frac{|f(t)|}{\|f\|_p} \frac{|g(t)|}{\|g\|_q} \leq \frac{1}{p} \left( \frac{|f(t)|}{\|f\|_p} \right)^p + \frac{1}{q} \left( \frac{|g(t)|}{\|g\|_q} \right)^q \quad \forall t \in [a, b].$$

By integrating the latter inequality,

$$\begin{aligned} \frac{1}{\|f\|_p \|g\|_q} \int_a^b |f(t)g(t)| dt &\leq \frac{1}{p(\|f\|_p)^p} \int_a^b |f(t)|^p dt + \frac{1}{q(\|g\|_q)^q} \int_a^b |g(t)|^q dt \\ &= \frac{1}{p} + \frac{1}{q} \\ &= 1, \end{aligned}$$

and hence the required inequality follows.  $\square$

Hölder's inequality holds for *integrable* functions (not necessarily continuous) and also for *improper* integrals, see Aliprantis and Burkinshaw [1, Theorem 31.3].

**Theorem A.4.** *The gamma function satisfies*

- (a)  $\Gamma(x+1) = x\Gamma(x)$  for every  $x > 0$ ,
- (b)  $\Gamma(n+1) = n!$  for each  $n = 1, 2, 3, \dots$ , and
- (c)  $\log(\Gamma)$  is convex on the interval  $(0, \infty)$ .

*Proof.* (a) Integration by parts yields

$$\begin{aligned} \Gamma(x+1) &= \int_0^\infty t^x e^{-t} dt \\ &= 0 + x \int_0^\infty t^{x-1} e^{-t} dt \\ &= x\Gamma(x). \end{aligned}$$

- (b) From (16), we see that  $\Gamma(1) = 1$ . By (a),

$$\begin{aligned} \Gamma(2) &= 1, \\ \Gamma(3) &= 2 \cdot 1, \\ \Gamma(4) &= 3 \cdot 2 \cdot 1. \end{aligned}$$

By induction, we conclude  $\Gamma(n+1) = n!$ .

- (c) Let  $p, q \in (1, \infty)$  satisfy  $\frac{1}{p} + \frac{1}{q} = 1$ . It suffices to show that

$$\log \Gamma\left(\frac{x}{p} + \frac{y}{q}\right) \leq \frac{\log \Gamma(x)}{p} + \frac{\log \Gamma(y)}{q} \quad \forall x, y \geq 0. \quad (19)$$

Thus

$$\begin{aligned} \Gamma\left(\frac{x}{p} + \frac{y}{q}\right) &= \int_0^\infty t^{\frac{x}{p} + \frac{y}{q} - 1} e^{-t} dt \\ &= \int_0^\infty \left[ t^{\frac{x}{p} - \frac{1}{p}} e^{-\frac{t}{p}} \right] \left[ t^{\frac{y}{q} - \frac{1}{q}} e^{-\frac{t}{q}} \right] dt \\ &\leq \left[ \int_0^\infty t^{x-1} e^{-t} dt \right]^{1/p} \left[ \int_0^\infty t^{y-1} e^{-t} dt \right]^{1/q} \quad \text{by Hölder's inequality} \\ &= [\Gamma(x)]^{1/p} [\Gamma(y)]^{1/q}. \end{aligned}$$

Then (4) follows because  $\log(\cdot)$  is increasing.  $\square$

The following theorem gives three properties that characterize the gamma function. A proof of Theorem A.5, also known as Bohr-Mollerup Theorem, can be found in Rudin [10, Teorema 8.19].

**Theorem A.5.** *Let  $G : (0, \infty) \rightarrow (0, \infty)$  satisfy*

- (a)  $G(x+1) = xG(x)$  for every  $x > 0$ ,
- (b)  $G(1) = 1$ , and
- (c)  $\log(G)$  is convex on the interval  $(0, \infty)$ .

*Then  $G(x) = \Gamma(x)$  for every  $x > 0$ .*

## A.2 The beta function

**Definition A.6.** *For each pair  $(x, y)$  of positive numbers, define the **beta function** as*

$$B(x, y) := \int_0^1 t^{x-1} (1-t)^{y-1} dt. \quad (20)$$

**Theorem A.7.** *The following equality holds for each pair  $(x, y)$  of positive numbers*

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \quad (21)$$

*Proof.* Put  $x = 1$  in (20) to have

$$B(1, y) = \frac{1}{y}. \quad (22)$$

Notice also that

$$\begin{aligned} B(x+1, y) &= \int_0^1 t^x (1-t)^{y-1} dt \\ &= \int_0^1 \left( \frac{t}{1-t} \right)^x (1-t)^{x+y-1} dt \\ &= \frac{t^x (1-t)^y}{x+y} \Big|_1^0 + \int_0^1 \frac{x}{x+y} (1-t)^{y-1} t^{x-1} dt, \end{aligned}$$

hence

$$B(x+1, y) = \frac{x}{x+y} B(x, y). \quad (23)$$

In addition, for each  $y$

$$\begin{aligned} B\left(\frac{x_1}{p} + \frac{x_2}{q}, y\right) &= \int_0^1 t^{\frac{x_1}{p} + \frac{x_2}{q} - 1} (1-t)^{y-1} dt \\ &= \int_0^1 t^{\frac{x_1-1}{p}} (1-t)^{\frac{y-1}{p}} t^{\frac{x_2-1}{q}} (1-t)^{\frac{y-1}{q}} dt \\ &\leq \left[ \int_0^1 t^{x_1-1} (1-t)^{y-1} dt \right]^{\frac{1}{p}} \left[ \int_0^1 t^{x_2-1} (1-t)^{y-1} dt \right]^{\frac{1}{q}}, \end{aligned}$$

the latter inequality follows from Hölder's inequality. Equivalently,

$$\log B\left(\frac{x_1}{p} + \frac{x_2}{q}, y\right) \leq \frac{\log B(x_1, y)}{p} + \frac{\log B(x_2, y)}{q},$$

which implies that  $\log B(\cdot, y)$  is convex.

Let  $G : (0, \infty) \rightarrow (0, \infty)$  be given by

$$G(x) := \frac{\Gamma(x+y)}{\Gamma(y)} B(x, y).$$

Equations (22) and (23) imply

(a)  $G(x+1) = xG(x)$  for every  $x > 0$ ,

(b)  $G(1) = 1$ ,

further, since  $\log B(\cdot, y)$  is convex, we conclude that

(c)  $\log(G)$  is convex on  $(0, \infty)$ .

By Theorem A.5,

$$\frac{\Gamma(x+y)}{\Gamma(y)} B(x, y) = \Gamma(x),$$

which is equivalent to (20). □

### A.3 The Gaussian integral

The integral in the following theorem is known as **Gaussian integral**.

**Theorem A.8.**  $\int_{-\infty}^{\infty} e^{-s^2} ds = \sqrt{\pi}$ .

*Proof.* Consider the integral (20) and set  $t = \sin^2 \theta$ , thus

$$\begin{aligned} B(x, y) &= \int_0^1 t^{x-1} (1-t)^{y-1} dt \\ &= \int_0^{\pi/2} (\sin^2 \theta)^{x-1} (1 - \sin^2 \theta)^{y-1} 2 \sin \theta \cos \theta d\theta \\ &= 2 \int_0^{\pi/2} (\sin \theta)^{2x-1} (\cos \theta)^{2y-1} d\theta. \end{aligned}$$

We now use Theorem A.7 with  $x = y = \frac{1}{2}$  to obtain

$$\frac{[\Gamma(1/2)]^2}{\Gamma(1)} = 2 \int_0^{\pi/2} 1 d\theta.$$

Then

$$\Gamma(1/2) = \sqrt{\pi}. \tag{24}$$

On the other hand, the change of variable  $t = s^2$  in the gamma function yields

$$\begin{aligned} \Gamma(x) &= \int_0^{\infty} t^{x-1} e^{-t} dt \\ &= 2 \int_0^{\infty} s^{2x-1} e^{-s^2} ds. \end{aligned}$$

In particular, for  $x = \frac{1}{2}$ , we have  $\Gamma(1/2) = 2 \int_0^{\infty} e^{-s^2} ds$ . Finally, from (24), we conclude that

$$\int_{-\infty}^{\infty} e^{-s^2} ds = \sqrt{\pi}.$$

□

## B Probability spaces

### B.1 The family of events

In this appendix,  $\Omega$  denotes a nonempty set called **sample space**.

**Definition B.1.** A family  $\mathcal{F}$  of subsets of  $\Omega$  is called a  $\sigma$ -**field** if

- (1)  $\emptyset \in \mathcal{F}$ ,
- (2)  $A \in \mathcal{F} \implies A^c \in \mathcal{F}$ .
- (3)  $A_n \in \mathcal{F}, n \in \mathbb{N}, \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ .

The members of the  $\sigma$ -field  $\mathcal{F}$  (also called  $\sigma$ -**algebra**) are called **events**.

Due to property (3) in the definition of  $\sigma$ -field,  $\mathcal{F}$  is said to be *closed under countable unions*. In fact,  $\mathcal{F}$  is also closed under *finite unions* because

$$A_1 \cup \dots \cup A_n = A_1 \cup \dots \cup A_n \cup \emptyset \cup \emptyset \cup \dots.$$

Further,  $\mathcal{F}$  is closed under countable intersections since

$$\bigcap_{n \in \mathbb{N}} B_n = \left( \bigcup_{n \in \mathbb{N}} B_n^c \right)^c.$$

**Example B.2.** The smallest  $\sigma$ -field of a sample space  $\Omega$  is

$$\mathcal{F} = \{\emptyset, \Sigma\}.$$

On the other hand, the largest  $\sigma$ -field of  $\Omega$  is the *power set* that consists of all the subsets of  $\Omega$ . When  $\Omega$  is a finite or countably infinite set, the usual  $\sigma$ -field for  $\Omega$  is the power set.  $\diamond$

It can be shown that the intersection of  $\sigma$ -fields of  $\Omega$  is also a  $\sigma$ -field (Exercise A.??), thus the following definition makes sense.

**Definition B.3.** Let  $\mathcal{C}$  be a nonempty collection of subsets of  $\Omega$ . The  $\sigma$ -**field generated by**  $\mathcal{C}$  is the smallest  $\sigma$ -field in  $\Omega$  containing  $\mathcal{C}$ , i.e.,

$$\sigma(\mathcal{C}) := \bigcap \{ \mathcal{F} \mid \mathcal{C} \subseteq \mathcal{F}, \mathcal{F} \text{ is a } \sigma\text{-field} \}.$$

**Example B.4.** Let  $\Omega = \mathbb{R}$ . Consider the collection  $\mathcal{C}$  of bounded open intervals

$$\mathcal{C} = \{(a, b) \mid a, b \in \mathbb{R}, a < b\}.$$

The  $\sigma$ -field  $\sigma(\mathcal{C})$  is known as the **Borel  $\sigma$ -field** of  $\mathbb{R}$ , also denoted  $\mathcal{B}(\mathbb{R})$ . The events of  $\mathcal{B}(\mathbb{R})$  are called **Borel events** or **Borel sets**. We now present some events of the Borel  $\sigma$ -field  $\mathcal{B}(\mathbb{R})$ :

(a) Intervals of the form  $(a, \infty)$  are Borel sets since

$$(a, \infty) = \bigcup_{n \in \mathbb{N}} (a, a + 1).$$

Similarly for the event  $(-\infty, b)$ .

(b) Intervals of the form  $(-\infty, a]$  are also events since

$$(-\infty, a] = (a, \infty)^c.$$

Similarly for the event  $[b, \infty)$ .

(c) Intervals of the form  $[a, b)$ , with  $a < b$ , because

$$[a, b) = [(-\infty, a) \cup [b, \infty)]^c.$$

Similarly,  $(a, b]$  and  $[a, b]$  are also events.

Countable unions and intersections of the above intervals are also Borel sets. However, not every subset of  $\mathbb{R}$  is an event of the Borel  $\sigma$ -field (see, for instance, Rana [6, p. 113]).  $\diamond$

**Example B.5.** Given a sample space  $\Omega$  and a  $\sigma$ -field  $\mathcal{F}$ , we can endow any nonempty event  $S$  in  $\Omega$  with the *induced*  $\sigma$ -field

$$\{S \cap A \mid A \in \mathcal{F}\}.$$

As a particular case, the induced Borel  $\sigma$ -field of  $[0, 1]$ , denoted  $\mathcal{B}([0, 1])$ , consists of all the Borel subsets of  $[0, 1]$ , that is,

$$\mathcal{B}([0, 1]) = \{A \in \mathcal{B}(\mathbb{R}) \mid A \subseteq [0, 1]\}.$$

$\diamond$

**Definition B.6.** The events  $E_1, E_2, E_3, \dots$  in  $\mathcal{F}$  are **disjoint** if

$$E_j \cap E_i = \emptyset \quad \text{for each } i \neq j.$$

## B.2 Probability measures

**Definition B.7.** Let  $\mathcal{F}$  be a  $\sigma$ -field of subsets of  $\Omega$ . A **probability measure** is any function  $P : \mathcal{F} \rightarrow [0, 1]$  such that

(a)  $P(\Omega) = 1$ ,  $P(\emptyset) = 0$ , and

(b)  $P$  is **countably additive** on  $\mathcal{F}$ , that is, for disjoint events  $E_1, E_2, E_3, \dots$  in  $\mathcal{F}$ ,

$$\bigcup_{j=1}^{\infty} E_j \in \mathcal{F} \quad \Rightarrow \quad \mu(\bigcup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} \mu(E_j).$$

We can observe that any probability measure is finitely additive. Thus

$$B = (B \setminus A) \cup (B \cap A) \quad \Rightarrow \quad P(B \setminus A) = P(B) - P(B \cap A).$$

In particular,  $P(A^c) = 1 - P(A)$ . Notice also that,  $A \subseteq B$  implies  $P(A) \leq P(B)$ . With similar arguments we can prove the following properties of  $P$ .

**Proposition B.8.** Let  $\mathcal{F}$  be a  $\sigma$ -field and  $P : \mathcal{F} \rightarrow \mathbb{R}$ . Then for any events  $A_1, A_2, \dots, A_n$  in  $\mathcal{F}$ ,

(a)  $P(A_2 - A_1) = P(A_2) - P(A_2 \cap A_1)$ ,

$$(b) P(A_1 \cup A_2 \cup \dots \cup A_n) \leq \sum_{j=1}^n P(A_j).$$

One of the most important probability measures is the so-called **Lebesgue measure** which extends the *length* of intervals to a wider family of sets. Lebesgue's measure  $\lambda$  assigns a real number  $\lambda(B)$  to each Borel set  $B \in \mathcal{B}(\mathbb{R})$ .

A **probability space** is a triple  $(\Omega, \mathcal{F}, P)$  where  $\mathcal{F}$  is a  $\sigma$ -field of subsets of  $\Omega$  and  $P : \mathcal{F} \rightarrow [0, 1]$  is a probability measure.

Some examples of probability spaces are

(a)  $(N, 2^N, P)$  where  $N$  is finite, the  $\sigma$ -field  $2^N$  is the *power set* of  $N$ , and  $P$  is any pmf on  $N$ .

(b) In part (a),  $N$  can be replaced by a countably infinite set.

(c)  $([0, 1], \mathcal{B}([0, 1]), \lambda)$  where  $\lambda$  is the Lebesgue's measure.

### B.3 General random variables

**Definition B.9.** Let  $(\Omega, \mathcal{F}, P)$  a probability space. A **random variable** is any *measurable* function  $X : \Omega \rightarrow \mathbb{R}$ , i.e.,

$$X^{-1}((-\infty, a]) \in \mathcal{F} \quad \forall a \in \mathbb{R}.$$

Recall that  $X \leq a$  is a shorter notation for the event

$$X^{-1}((-\infty, a]) = \{\omega \in \Omega \mid X(\omega) \leq a\}.$$

The **probability distribution**  $F_X : \mathbb{R} \rightarrow \mathbb{R}$  of  $X$  is defined as

$$F_X(x) := P(X \leq x).$$

If there is no confusion, we simply write  $F$  instead of  $F_X$ .

A proof of the following theorem can be found in Resnick [7, pp. 33-34].

**Theorem B.10.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and a random variable  $X : \Omega \rightarrow \mathbb{R}$ . Then its probability distribution  $F : \mathbb{R} \rightarrow \mathbb{R}$  satisfies the following properties

- (a) if  $x < y$ , then  $F(x) \leq F(y)$ ,
- (b)  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow \infty} F(x) = 1$ ,
- (c)  $F$  is right-continuous at any  $a \in \mathbb{R}$ , that is,  $\lim_{x \downarrow a} F(x) = F(a)$ .

If  $F : \mathbb{R} \rightarrow \mathbb{R}$  satisfies properties (a), (b), and (c) of Theorem B.10, then  $F$  is called a **distribution function**.

A converse of Theorem B.10 is useful to simulate random variables by means of the so-called *inverse transformation (or inversion) method*. This method uses the so-called **quantile function**  $Q$  given in the following theorem.

**Theorem B.11.** Let  $U$  be a random variable with uniform distribution on the interval  $[0, 1]$ . Given a distribution function  $F$ , define the quantile function

$$Q(u) := \inf F^{-1}([u, 1]) = \inf\{x \in \mathbb{R} \mid u \leq F(x)\}, \quad u \in (0, 1). \quad (25)$$

Then  $Q \circ U$  has probability distribution  $F$ .



*Proof.* To see that  $Q \circ U$  indeed has distribution  $F$ , we note the following properties for each  $u \in (0, 1)$  and  $a \in \mathbb{R}$ .

- (a) From the definition of  $Q$ , we conclude that  $Q$  is nondecreasing.
- (b)  $Q(u) \in F^{-1}([u, 1])$ , that is,  $Q(u) = \min F^{-1}([u, 1])$ . To prove this assertion, observe that for each  $n \in \mathbb{N}$ , there exists  $x_n \in F^{-1}([u, 1])$  such that

$$Q(u) \leq x_n < Q(u) + \frac{1}{n}.$$

Then  $F(x_n) \geq u$  for each  $n$  and  $x_n \downarrow Q(u)$ . Since  $F$  is right-continuous, in particular at  $Q(u)$ ,

$$\begin{aligned} u &\leq \lim_{n \rightarrow \infty} F(x_n) \\ &= F(Q(u)), \end{aligned}$$

that is,  $Q(u) \in F^{-1}([u, 1])$ .

- (c)  $Q(F(a)) \leq a$ . This inequality follows from the definition (25) of  $Q(F(a))$  because

$$a \in \{x \in \mathbb{R} \mid F(a) \leq F(x)\}.$$

- (d)  $Q^{-1}((-\infty, a]) = (0, F(a)]$  or, equivalently,  $Q(u) \leq a$  if and only if  $u \leq F(a)$ .

Let us show first that  $Q(u) \leq a$  implies  $u \leq F(a)$ . Notice that (b) allows us to assert the existence of some  $x_u$  such that  $Q(u) = x_u$  and

$$F(x_u) \geq u.$$

Then  $x_u \leq a$  and, by the monotonicity of  $F$ ,  $F(x_u) \leq F(a)$ . But  $u \leq F(x_u)$ , hence  $u \leq F(a)$ .

Conversely,  $u \leq F(a)$  implies  $Q(u) \leq a$ . Indeed, by part (a) and (c),

$$Q(u) \leq Q(F(a)) \leq a.$$

Finally, the theorem follows from (d) because

$$\begin{aligned} P(Q \circ U \leq a) &= P(Q[U(\omega)] \leq a) \\ &= P(U(\omega) \leq F(a)) \\ &= F(a). \end{aligned}$$

□

**Remark B.12.** The function  $Q$  is sometimes called the **quantile function** of  $F$ . If there exists the inverse function  $F^{-1}$  of  $F$ , then  $Q = F^{-1}$ .

**Theorem B.13.** Let  $F : \mathbb{R} \rightarrow \mathbb{R}$  be a distribution function. Then there exists a probability space  $(\Omega, \mathcal{F}, P)$  and a random variable  $X : \Omega \rightarrow \mathbb{R}$  such that

$$P(X \leq a) = F(a) \quad a \in \mathbb{R}.$$

*Proof.* Let  $\Omega = (0, 1)$ ,  $\mathcal{B}((0, 1))$ , and  $P = \lambda$ . Given  $F$ , consider the corresponding quantile function  $Q$ . Define

$$X = Q \circ U,$$

where  $U : (0, 1) \rightarrow (0, 1)$  is the uniform random variable. Then, by Theorem B.11,  $X$  has probability distribution  $F$ .  $\square$

**Remark B.14.** Given a random variable  $X$ , the **moment-generating function** is

$$m_X(t) = E(e^{tX})$$

whenever it is finite on a neighborhood of the origin. If  $Y$  is another random variable and

$$m_X(t) = m_Y(t)$$

on a neighborhood of the origin, then  $F_X(x) = F_Y(x)$  for every  $x \in \mathbb{R}$ . A proof of this fact can be found in Billingsley [3, Section 30]. Another useful result is the following: suppose  $m_X$  is the moment-generating function of  $X$ , then

$$m_{aX+b}(t) = e^{bt} m_X(at)$$

for  $a, b \in \mathbb{R}$ .  $\diamond$

## B.4 A note on simulation of random variables

When the distribution function  $F$  has inverse, Theorem B.13 can be useful to simulate values of random variables. The procedure relies on the generation of the so-called *random numbers* which are a computational model of the uniform distribution on the interval  $(0, 1)$ .

**Example B.15.** Let  $Y \sim \text{Exp}(\beta)$ . Then the restriction of  $F_Y(y) = 1 - e^{-\beta y}$  to the non-negative reals has inverse. Therefore

$$X := -\frac{1}{\beta} \log(1 - U)$$

has exponential distribution, where  $U \sim \text{Unif}([0, 1])$ .  $\diamond$

We now consider an example of a discrete distribution which does not have inverse.

**Example B.16.** Consider a random variable  $X$  that takes the values 1, 2, and 3 with probabilities 0.5, 0.3, and 0.2, respectively. Then the quantile function is

$$Q(w) = \begin{cases} 1 & \text{if } 0 < w \leq 0.5, \\ 2 & \text{if } 0.5 < w \leq 0.8, \\ 3 & \text{if } 0.8 < w < 1, \end{cases}$$

and  $Q(U)$  has distribution  $F_X$ .  $\diamond$

**Example B.17.** Consider the r.v.  $Y \sim \text{Gamma}(n, \lambda)$ , that is,

$$F_Y(x) = \int_0^x \frac{\lambda e^{-\lambda y} (\lambda y)^{n-1}}{(n-1)!} dy, \quad 0 \leq x < \infty.$$

Instead of looking for the inverse  $F_Y^{-1}$ , we recall that

$$X := -\frac{1}{\lambda} \log(U_1 \cdots U_n)$$

has gamma distribution, whenever  $U_1, \dots, U_n$  are i.i.d. uniform random variables on  $(0, 1)$ .  $\diamond$

An introduction to stochastic simulation can be found, for instance, in Ross [9].

## References

- [1] C. D. ALIPRANTIS AND O. BURKINSHAW, *Principles of real analysis*, Academic Press, Inc., San Diego, CA, third ed., 1998.
- [2] R. B. ASH, *Statistical Inference: A Concise Course*, Dover, 2011. Online version, Lectures on Statistics, <https://faculty.math.illinois.edu/~r-ash/Stat.html>.
- [3] P. BILLINGSLEY, *Probability and measure*, John Wiley & Sons, Inc., New York, third ed., 1995.
- [4] M. LEFEBVRE, *Applied probability and statistics*, Springer, New York, 2006.
- [5] V. M. PANARETOS, *Statistics for mathematicians. A rigorous first course*, Compact Textbooks in Mathematics, Birkhäuser/Springer, [Cham], 2016.
- [6] I. K. RANA, *An introduction to measure and integration*, vol. 45 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, second ed., 2002.
- [7] S. I. RESNICK, *A probability path*, Birkhäuser Boston, Inc., Boston, MA, 1999.
- [8] J. S. ROSENTHAL, *A first look at rigorous probability theory*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, second ed., 2006.
- [9] S. M. ROSS, *Simulation, Statistical Modeling and Decision Science*, Academic Press, Inc., San Diego, CA, second ed., 1997.
- [10] W. RUDIN, *Principles of mathematical analysis*, International Series in Pure and Applied Mathematics, McGraw-Hill Book Co., New York-Auckland-Düsseldorf, third ed., 1976.
- [11] D. D. WACKERLY, W. MENDENHALL III, AND R. L. SCHEAFFER, *Mathematical Statistics with Applications*, Duxbury Advanced Series, seventh edition ed., 2008.