Chapter 1

Probability

This chapter will provide the necessary background in probability theory that we need to formulate the models of investments we will discuss later in the course. Some of it should be familiar, some of it won't.

1.1 Basic probability theory

Consider an experiment such as tossing a coin or throwing a die. The set of all possible outcomes is called the **sample space**. We shall often, but not always, use S to denote sample spaces.

Example 1.1. Flipping a coin. The outcome is either H (for heads) or T (for tails). The sample space is

$$S = \{H, T\} \ .$$

Example 1.2. Flipping a coin twice. The outcome is either H followed by H, H followed by T, T followed by H, or T followed by T. The sample space is

$$S = \{HH, HT, TH, TT\} \ .$$

Let S be a sample space with n elements, say $S = \{1, \ldots, n\}$. Suppose we are given n real numbers p_i, $i = 1, \ldots, n$, such that

(i) $p_i \geq 0$ for every $i \in S$;

(ii) $\sum_{i=1}^{n} p_i = 1$.

We can then interpret p_i to be the likelihood of the outcome i, for any $i \in S$, and we shall say that the p_i's define a **probability measure** on S.

Example 1.3. In Example 1.1 above, we could have chosen $p_H = p_T = \frac{1}{2}$, or $p_H = \frac{1}{3}$, $p_T = \frac{2}{3}$.
In Example 1.2 above, we could have chosen $p_{HH} = p_{HT} = p_{TH} = p_{TT} = \frac{1}{4}$.

A subset of the sample space S is called an **event**. If A is an event, then we define the probability of A occurring, denoted by $P(A)$, by

$$P(A) = \sum_{i \in A} p_i \ .$$

Note that, in particular, $P(S) = \sum_{i \in S} p_i = 1$.

Example 1.4. Let a, b, c be three companies. Let (i, j, k) be the outcome that in 2010 company i makes more profit than company j and that company j makes more profit than company k. Then the sample space is

$$S = \{(a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a)\} .$$

Define a probability measure on S by letting

$$p_{(i,j,k)} = \frac{1}{6} \quad \text{for every } i,\ j,\ k.$$

Let A be the event that a makes most profit in 2010. Then $A = \{(a, b, c), (a, c, b)\}$ and $P(A) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Suppose that S is a sample space and that $A \subset S$ and $B \subset S$ are two events. Let us recall the following definitions.

- The **complement** of the event A, denoted by A' or $\overline{A}$, is given by

$$A' = \{\, s \in S \mid s \notin A \,\} .$$

- The **union** $A \cup B$ of the events A and B is given by

$$A \cup B = \{\, s \in S \mid s \in A \text{ or } s \in B \text{ or both} \,\} .$$

- The **intersection** $A \cap B$ of the events A and B is given by

$$A \cap B = \{\, s \in S \mid s \in A \text{ and } s \in B \,\} .$$

Note that

- $P(A \cup B)$ is the probability that at least one of A or B occurs;
- $P(A \cap B)$ is the probability that both A and B occur.

Example 1.5. Profits of companies (Example 1.4) continued. Let A be the event that company a makes most profit and let B be the event that company c makes least profit. Then

$$A = \{(a, b, c), (a, c, b)\} \quad \text{and} \quad B = \{(a, b, c), (b, a, c)\} .$$

We have $A \cup B = \{(a, b, c), (a, c, b), (b, a, c)\}$ and $A \cap B = \{(a, b, c)\}$. Moreover $P(A) = P(B) = \frac{2}{6} = \frac{1}{3}$, $P(A \cup B) = \frac{3}{6} = \frac{1}{2}$, and $P(A \cap B) = \frac{1}{6}$.

We now recall a useful result that relates $P(A \cap B)$ and $P(A \cup B)$.

Theorem 1.6. *For two events A and B*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) .$$

Example 1.7. In Example 1.5 above we have

$$P(A \cup B) = \frac{1}{2} = \frac{1}{3} + \frac{1}{3} - \frac{1}{6} = P(A) + P(B) - P(A \cap B) .$$

Example 1.8. Let the probability that the FTSE100 increases today be 0.52 and the probability that it increases tomorrow be 0.52 as well. Suppose that the probability it increases both today and tomorrow is 0.28. What is the probability that the FTSE100 increases neither today nor tomorrow?

Solution. Let A be the event that the FTSE100 increases today and let B be the event that the FTSE100 increases tomorrow. We know that $P(A) = P(B) = 0.52$, that $P(A \cap B) = 0.28$ and we want to find $P((A \cup B)')$. Now $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.76$, so $P((A \cup B)') = 1 - 0.76 = 0.24$ is the desired probability that the FTSE100 increases neither today nor tomorrow. $\square$

Recall that if A and B are events, then the **conditional probability** of A given B, denoted $P(A|B)$, is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} .$$

Example 1.9. Profits of companies (Example 1.5) continued. If A is the event that a makes most profit and B the event that c makes least profit, what is the probability that a makes most profit given that c makes least profit?

Solution.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/3} = \frac{1}{2}$$

is the desired probability that a makes most profit given that c makes least profit. $\square$

We say that two events A and B are **independent** if

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

or, equivalently, if

$$P(A \cap B) = P(A)P(B) .$$

Example 1.10. In Example 1.9 above, the two events A and B are dependent (that is, not independent), since $P(A|B) = \frac{1}{2} > P(A)$.

1.2 Random variables

Definition 1.11. A **random variable** is a quantity X determined by the outcome of an experiment. It is given by the following data:

(i) possible values $x_1, \ldots, x_n$ it can take on;

(ii) probabilities $p_1, \ldots, p_n$.

We interpret $p_i = P(X = x_i)$ to be the likelihood with which X takes the value x_i. The collection of the p_i's is referred to as the **probability distribution** of the random variable X.

Recall that if X is a random variable as defined above, then its **expectation**, denoted by $E(X)$, is given by

$$E(X) = \sum_{i=1}^{n} x_i p_i = \sum_{i=1}^{n} x_i P(X = x_i) .$$

Example 1.12. Suppose that a certain company

- makes $£1{,}000{,}000$ with probability $\frac{1}{4}$;
- loses $£500{,}000$ with probability $\frac{1}{4}$;
- makes $£2{,}000{,}000$ with probability $\frac{1}{2}$.

If X denotes the profit of the company in pounds, then X is a random variable with

$$\begin{aligned} x_1 &= 1000000 & p_1 &= \frac{1}{4} \\ x_2 &= -500000 & p_2 &= \frac{1}{4} \\ x_3 &= 2000000 & p_3 &= \frac{1}{2} \end{aligned}$$

In particular, the expected profit of the company in pounds is given by

$$E(X) = x_1p_1 + x_2p_2 + x_3p_3 = 1125000\,.$$

Definition 1.13. Let $p \in [0, 1]$. A random variable X is said to **Bernoulli(p)** (or simply Bernoulli) distributed if its possible values are 0 and 1, and if $P(X = 1) = p$ and $P(X = 0) = 1 - p$.

Note that if X is Bernoulli(p) distributed then $E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$.
Expectation is linear in the following sense:

Lemma 1.14. *Let X be a random variable and let a and b be constants. Then*

$$E(aX + b) = aE(X) + b\,.$$

Proof. Suppose that X takes on the value x_i with probability p_i. Then $aX + b$ is a random variable which takes on the values $ax_i + b$ with probability p_i. Thus

$$E(aX + b) = \sum_{i=1}^{n}(ax_i + b)p_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} p_i = aE(X) + b\,.$$

□

Repeated application of the lemma above yields the following important result.

Proposition 1.15. *If $X_1, \ldots, X_m$ are random variables and $\alpha_1, \ldots, \alpha_m$ are constants, then*

$$E(\sum_{j=1}^{m} \alpha_j X_j) = \sum_{j=1}^{m} \alpha_j E(X_j)\,.$$

Definition 1.16. The **variance** of a random variable X is given by

$$\mathrm{Var}(X) = E((X^2 - E(X))^2)\,.$$

The **standard deviation** of a random variable is given by

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}\,.$$

The following result is often useful.

Lemma 1.17. *If X is a random variable, then*

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 .$$

Proof. Using the linearity of the expectation we see that

$$\begin{aligned}\mathrm{Var}(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - 2E(X)E(X) + E(E(X)^2) \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X^2) - E(X)^2 .\end{aligned}$$

□

Example 1.18. Let X be Bernoulli(p) distributed. Then $E(X) = p$ and

$$E(X^2) = 1^2 p + 0^2(1-p) = p ,$$

so

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p) .$$

Unlike expectation, variance is not linear as the following result shows.

Lemma 1.19. *If X is a random variable and a and b are constants, then*

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X) .$$

Proof. This is a short calculation using the definition of variance and is left as an exercise. □

While variance is not linear in general, it is sometimes possible to conclude that the variance of a sum of two random variables is the sum of the variances of the random variables. Before stating this result we recall the following important concept.

Definition 1.20. Two random variables X and Y are said to be **independent** if

$$P(X = x_i, Y = y_i) = P(X = x_i)P(Y = y_i) .$$

Furthermore, a sequence of random variables $X_1, X_2, \ldots$ is said to be independent if X_i and X_j are independent whenever $i \neq j$.

The result alluded to earlier can now be formulated as follows.

Proposition 1.21. *If $X_1, X_2, \ldots, X_m$ is a sequence of independent random variables, then*

$$\mathrm{Var}\left(\sum_{j=1}^{m} X_j\right) = \sum_{j=1}^{m} \mathrm{Var}(X_j) .$$

Proof. This will follow from a more general result to be discussed shortly (see Proposition 1.27). □

Another concept that you have already encountered and that will play an important role later on is the following. A **random walk** is a sum of independent random variables. To be precise, suppose that $X_1, X_2, \ldots$ is a sequence of random variables with the following properties:

- $P(X_j = 1) = P(X_j = -1) = \frac{1}{2}$ for all j;
- the random variables $X_1, X_2, \ldots$ are independent.

We can think of X_j as the j-th step of the random walk. Now define:

$$S_n = \sum_{j=1}^{n} X_j \,.$$

Then S_n is a random walk.

Note that

$$E(X_j) = 1\frac{1}{2} + (-1)\frac{1}{2} = 0 \,,$$

so

$$E(S_n) = E(X_1) + \cdots E(X_n) = 0 \,.$$

In order to determine the variance of S_n note that

$$E(X_j^2) = 1^2\frac{1}{2} + (-1)^2\frac{1}{2} = 1 \,,$$

so

$$\mathrm{Var}(X_j) = E(X_j^2) - E(X_j)^2 = 1 - 0^2 = 1 \,,$$

and thus, by Proposition 1.21

$$\mathrm{Var}(S_n) = \mathrm{Var}(X_1) + \cdots \mathrm{Var}(X_n) = 1 + \cdots + 1 = n \,.$$

Recall that the **covariance** of two random variables X and Y is defined by

$$\mathrm{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) \,.$$

Note that $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ and that $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$. The following is a useful reformulation, the simple proof which is left as an exercise.

Lemma 1.22. *Let X and Y be two random variables. Then*

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) \,.$$

Covariance turns out to be linear in each of its arguments. We shall consider a special case first.

Lemma 1.23. *Let X_1, X_2 and Y be three random variables. Then*

$$\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y) \,.$$

Proof. Using the previous lemma and the linearity of expectation we see that

$$\begin{aligned}
\mathrm{Cov}(X_1 + X_2, Y) &= E((X_1 + X_2)Y) - E(X_1 + X_2)E(Y) \\
&= E(X_1Y + X_2Y) - (E(X_1) + E(X_2))E(Y) \\
&= E(X_1Y) + E(X_2Y) - E(X_1)E(Y) - E(X_2)E(Y) \\
&= E(X_1Y) - E(X_1)E(Y) + E(X_2Y) - E(X_2)E(Y) \\
&= \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y) \,.
\end{aligned}$$

□

Repeated application of the previous lemma, together with the fact that $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ yields the following result.

Proposition 1.24. *Let X_i, $i = 1, 2, \ldots, m$ and Y_j, $j = 1, 2, \ldots n$ be two sequences of random variables. Then*

$$\mathrm{Cov}\left(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{Cov}(X_i, Y_j).$$

Proof. See Exercise 3 on Coursework 1. □

The **correlation** of two random variables X and Y is given by

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}\,.$$

It is a non-trivial fact that

$$-1 \leq \mathrm{Cor}(X, Y) \leq 1\,.$$

We now recall that two random variables X and Y are said to **uncorrelated** if

$$\mathrm{Cov}(X, Y) = 0\,.$$

Lemma 1.25. *If two random variables X and Y are independent, then they are uncorrelated.*

Proof. Suppose that X and Y are independent and that X takes on values $x_1, \ldots, x_m$ while Y takes on values $y_1, \ldots, y_n$. Then

$$\begin{aligned}
E(XY) &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j P(X = x_i, Y = y_j) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j P(X = x_i)P(Y = y_j) \\
&= \sum_{i=1}^{m} x_i P(X = x_i) \sum_{j=1}^{n} y_j P(Y = y_j) \\
&= E(X)E(Y)\,.
\end{aligned}$$

Thus $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$. □

Note that the converse of the lemma is false, that is, X and Y uncorrelated does not imply that X and Y are independent, as the following example shows.

Example 1.26. Let X be a random variable taking on values $1, 0, -1$ and let Y be a random variable taking on values $1, 0$. Suppose that

$$P(X = 1, Y = 1) = P(X = -1, Y = 1) = P(X = 0, Y = 0) = \frac{1}{3}$$

and that the remaining joint probabilities are all 0. Thus

$$P(X = 1) = P(X = 0) = P(X = -1) = \frac{1}{3}$$

and

$$P(Y = 1) = \frac{2}{3}, \quad P(Y = 0) = \frac{1}{3}\,.$$

Now

$$E(X) = 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + (-1) \cdot \frac{1}{3} = 0\,,$$

while

$$E(XY) = 1 \cdot 1 \cdot \frac{1}{3} + (-1) \cdot 1 \cdot \frac{1}{3} + 0 \cdot 0 \cdot \frac{1}{3} = 0\,.$$

Thus $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0$, so X and Y are uncorrelated. However, X and Y are not independent since

$$P(X = 0, Y = 0) = \frac{1}{3} \neq \frac{1}{9} = P(X = 0)P(Y = 0)\,.$$

Proposition 1.27. *Given a sequence $X_1, X_2, \ldots, X_n$ of random variables we have*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)\,.$$

If the random variables $X_1, X_2, \ldots, X_n$ are mutually uncorrelated (that is X_i and X_j are uncorrelated whenever $i \neq j$), then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Cov}(X_i, X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i)\,.$$

Proof. Follows from Proposition 1.24. □

1.3 Continuous random variables

Recall that a **continuous random variable** X takes values in $\mathbb{R}$ and is associated with a **probability density function** (abbreviated 'pdf'), that is, an integrable function f_X on $\mathbb{R}$ such that

(i) $f_X(x) \geq 0$, for every $x \in \mathbb{R}$;

(ii) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

If a and b are real numbers with $a \leq b$ we interpret $\int_a^b f_X(x)\,dx$ to be the likelihood that X takes on values between a and b, that is,

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

Recall that if X is a continuous random variable the **expectation** $E(X)$ of X is given by

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

while the **variance** of X is given by

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - \left(\int_{-\infty}^{\infty} x f_X(x)\,dx\right)^2\,.$$

whenever these quantities are finite.

If X is a random variable and n is a positive integer, we say that the n-th moment of X exists if

$$E(|X|^n) < \infty\,.$$

If the n-th moment exists we call $E(X^n)$ the n-th **moment** of X. In particular, the first moment of a random variable equals its expectation (if it exists).

Example 1.28. Let $A > 0$ and let X be a random variable such that

$$f_X(t) = \frac{A}{\pi^2 A^2 + t^2}\,.$$

If this is the case we say that X has a **Cauchy distribution**. Note that $f_X(x) \geq 0$ for any real x. Moreover, using the substitution $x = \pi A \tan\theta$, we see that

$$\begin{aligned}\int_{-\infty}^{\infty} f_X(x)\,dx = \int_{-\infty}^{\infty} \frac{A}{\pi^2 A^2 + x^2}\,dx &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{A}{\pi^2 A^2 + \pi^2 A^2 \tan^2\theta}\pi A \sec^2\theta\,d\theta \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\pi A^2 \sec^2\theta}{\pi^2 A^2 \sec^2\theta}\,d\theta = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{\pi}\,d\theta = \pi\frac{1}{\pi} = 1\,,\end{aligned}$$

so f_X is a proper pdf.

It turns out that the second moment of the Cauchy distribution does not exist. To see this note that

$$\lim_{x\to\infty} x^2 f_X(x) = \lim_{x\to\infty} \frac{Ax^2}{\pi^2 A + x^2} = \lim_{x\to\infty} \frac{A}{\frac{\pi^2 A}{x^2} + 1} = \frac{A}{1} = A\,,$$

so

$$E(|X|^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \infty\,,$$

that is, the second moment of X does not exist. In fact, the first moment does not exist either. In order to see this note that

$$\lim_{x\to\infty} \frac{x\frac{A}{\pi^2 A^2 + x^2}}{\frac{A}{x}} = 1\,,$$

so

$$E(|X|) = \int_{-\infty}^{\infty} |x|\, f_X(x)\,dx = 2\int_0^{\infty} x\frac{A}{\pi^2 A^2 + x^2}\,dx \approx \int_0^{\infty} \frac{A}{x}\,dx = \infty\,,$$

because $\int \frac{1}{x}\,dx = \log x \to \infty$ as $x \to \infty$.

We now recall the following important concept.

Definition 1.29. The **characteristic function** G_X of a random variable X is defined for any $\alpha \in \mathbb{R}$ by

$$G_X(\alpha) = E(e^{i\alpha X}) = \int_{-\infty}^{\infty} e^{i\alpha x} f_X(x)\,dx\,.$$

Knowing the characteristic function of a random variable makes it possible to calculate its moments, as the following result shows.

Lemma 1.30. *Given a random variable X with characteristic function G_X we have for any $\alpha \in \mathbb{R}$*

$$G_X(\alpha) = 1 + \sum_{k=1}^{\infty} \frac{(i\alpha)^k}{k!} E(X^k)\,.$$

Proof. Observe that

$$G_X(\alpha) = \int_{-\infty}^{\infty} e^{i\alpha x} f_X(x)\,dx = \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{(i\alpha x)^k}{k!} f_X(x)\,dx$$
$$= \sum_{k=0}^{\infty} \frac{(i\alpha)^k}{k!} \int_{-\infty}^{\infty} x^k f_X(x)\,dx = 1 + \sum_{k=1}^{\infty} \frac{(i\alpha)^k}{k!} E(X^k)\,.$$

□

Example 1.31. Suppose that $X \sim \mathrm{Uniform}(0,1)$, that is, X is a random variable with

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$G_X(\alpha) = \int_{-\infty}^{\infty} e^{i\alpha x} f_X(x)\,dx = \int_0^1 e^{i\alpha x} = \left[\frac{e^{i\alpha x}}{i\alpha}\right]_{x=0}^{x=1} = \frac{e^{i\alpha} - 1}{i\alpha}\,.$$

Now,

$$G_X(\alpha) = \frac{1}{i\alpha}\left(\sum_{k=0}^{\infty} \frac{(i\alpha)^k}{k!} - 1\right) = \frac{1}{i\alpha} \sum_{k=1}^{\infty} \frac{(i\alpha)^k}{k!} = \sum_{k=0}^{\infty} \frac{(i\alpha)^k}{(k+1)!} = 1 + \sum_{k=1}^{\infty} \frac{(i\alpha)^k}{k!} \frac{1}{k+1}\,.$$

Using the previous lemma we see that

$$E(X^k) = \frac{1}{k+1}\,.$$

Note that we could also have calculated the moments directly:

$$E(X^k) = \int_{-\infty}^{\infty} x^k f_X(x)\,dx = \int_0^1 x^k\,dx = \left[\frac{1}{k+1} x^{k+1}\right]_{x=0}^{x=1} = \frac{1}{k+1}\,.$$

Recall that the **joint probability distribution function** of two continuous random variables X and Y is denoted by $f_{X,Y}$ and has the property that for every $a \le b$ and every $c \le d$

$$P(a \le X \le b, c \le Y \le d) = \int_c^d \int_a^b f_{X,Y}(x,y)\,dx\,dy\,.$$

In particular

$$P(a \le X \le b, -\infty \le Y \le \infty) = \int_{-\infty}^{\infty} \int_a^b f_{X,Y}(x,y)\,dx\,dy\,,$$

so

$$P(a \le X \le b) = \int_a^b \int_{\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx\,,$$

hence

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,.$$

Similarly

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,.$$

It turns out that two continuous random variables are independent if and only if their joint probability distribution function factorises. To be precise:

Lemma 1.32. *Two continuous random variables X and Y are independent if and only if*

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \forall x, y \in \mathbb{R}\,.$$

1.4 Gaussian or normal random variables

A random variable X is said to be **Gaussian** or **normal** with mean μ and variance σ^2 where $\mu \in \mathbb{R}$ and $\sigma > 0$ if X has pdf

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

If X is normal with parameters μ and σ^2 we write $X \sim N(\mu, \sigma)$. It turns out that

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

Furthermore, it turns out that the sum of two independent Gaussian random variables is again a Gaussian random variable. To be precise we have the following important result.

Lemma 1.33. *Suppose that $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$. If X_1 and X_2 are independent then*

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

If $\mu = 0$ and $\sigma = 1$, we say that X is **standard normal**, in which case

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Note that every normal random variable can be transformed to a standard normal random variable as follows.

Lemma 1.34. *If $X \sim \mathrm{N}(\mu, \sigma^2)$ then*

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

Proof. See Remark 1.40. □

The **cumulative distribution function** of a standard normal random variable is defined to be

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt .$$

Note that if $X \sim \mathrm{N}(0,1)$, then $\Phi(x) = P(X \leq x)$. Note also that

$$\Phi(x) = P(X \leq x) = P(X \geq -x) = 1 - P(X \leq -x) = 1 - \Phi(-x). \tag{1.1}$$

Moreover, if $a \leq b$ then

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = \Phi(b) - \Phi(a).$$

The function Φ cannot be expressed in terms of elementary functions. In this module we will use tables to determine the values of Φ. Using relation (1.1) we see that $\Phi(x)$ only needs to be tabulated for $x > 0$. In the table distributed in the lectures, $\Phi(x)$ is tabulated for arguments x, correct to 2 decimal places. If higher precision is required **linear interpolation** can be used. This is done as follows.

Suppose that $\underline{x} < x < \overline{x}$, where $\underline{x}$ and $\overline{x}$ are the nearest tabulated arguments of Φ. Then a good approximation to $\Phi(x)$ is given by

$$\Phi(x) \approx \frac{\overline{x} - x}{\overline{x} - \underline{x}} \Phi(\underline{x}) + \frac{x - \underline{x}}{\overline{x} - \underline{x}} \Phi(\overline{x}).$$

Example 1.35. Let $x = 1.116$. Then $\underline{x} = 1.11$ and $\overline{x} = 1.12$ and a good approximation of $\Phi(x)$ is

$$\Phi(x) \approx 0.4\Phi(1.11) + 0.6\Phi(1.12) = 0.4 \cdot 0.8665 + 0.6 \cdot 0.8686 = 0.8678\,.$$

Example 1.36. IQ scores of 11 year olds are normally distributed with mean value 100 and standard deviation 14.2. What is the probability that a randomly chosen 11 year has IQ more than 130?

Solution. Let X be the IQ of a randomly chosen 11 year old. We know that $X \sim \mathrm{N}(\mu, \sigma^2)$, where $\mu = 100$ and $\sigma^2 = (14.2)^2$. Now

$$P(X > 130) = P\left(\frac{X-\mu}{\sigma} > \frac{130-\mu}{\sigma}\right) = P\left(\frac{X-\mu}{\sigma} > 2.113\right) = 1 - \Phi(2.113)\,.$$

In order to determine $\Phi(2.113)$ we use linear interpolation. The nearest tabulated arguments are $\underline{x} = 2.11$ and $\overline{x} = 2.12$ and a good approximation to $\Phi(2.113)$ is

$$\Phi(2.113) \approx 0.7\Phi(2.11) + 0.3\Phi(2.12) = 0.7 \cdot 0.9826 + 0.3 \cdot 0.9830 = 0.9827\,.$$

Thus, the desired probability is

$$P(X > 130) = 1 - \Phi(2.113) = 0.017\,.$$

□

Note 1.37. You only need to use linear interpolation if you are specifically asked to do so.

Our next task is to derive the transformation formula for the probability distribution functions of continuous random variables. Before doing so recall that if X is a continuous random variable with pdf f_X, its **cumulative distribution function**, denoted F_X and abbreviated cdf, is defined to be

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt\,.$$

Note that by the fundamental theorem of calculus

$$\frac{d}{dx}F_X(x) = f_X(x)\,.$$

Suppose now that X is a continuous random variable and g a real valued function. We are now going to answer the question how the pdf of the random variable $Y = g(X)$ is related to that of X.

Theorem 1.38 (Transformation Formula)**.** *Let X be a continuous random variable and let $Y = g(X)$, where g is a differentiable function which is*

(i) either strictly monotonically increasing (so $g'(x) > 0\ \forall x \in \mathbb{R}$)

(ii) or strictly monotonically decreasing (so $g'(x) < 0\ \forall x \in \mathbb{R}$).

Then

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| & \text{for all } y \text{ for which } g^{-1}(y) \text{ exists} \\ 0 & \text{for all other } y \end{cases}$$

Proof. We give the proof for strictly monotonically increasing g only. The other case is similar. We start by calculating the cdf of Y:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(y)\,dy = P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))\,.$$

Now

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(g^{-1}(y)) = F_X'(g^{-1}(y))\frac{d}{dy}g^{-1}(y)\,,$$

by the chain rule. Observe that since g is monotonically increasing, so is g^{-1}. Thus, for all y for which $g^{-1}(y)$ exists

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

since

$$\frac{d}{dy}g^{-1}(y) > 0\,.$$

On the other hand, $F_Y(y)$ is either 0 or 1 for all y for which $g^{-1}(y)$ does not exist, so $f_Y(y) = 0$ in this case. $\square$

Example 1.39. Suppose that $X \sim \mathrm{N}(\mu, \sigma^2)$ and let a and b be real constants with $a \neq 0$. Let $Y = aX + b$. What is the pdf of Y?

Solution. Write $g(x) = ax + b$. Then $Y = g(X)$. Note that $g'(x) = a$, so g is strictly monotonically increasing if $a > 0$ and strictly monotonically decreasing if $a < 0$. Now, if $y = ax + b$, then $x = (y - b)/a$, so

$$g^{-1}(y) = \frac{y-b}{a}\,,$$

and

$$\frac{d}{dy}g^{-1}(y) = \frac{1}{a}\,.$$

Moreover, since $X \sim \mathrm{N}(\mu, \sigma^2)$,

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

Using the Transformation Formula we see that

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{((y-b)/a-\mu)^2}{2\sigma^2}\right)\left|\frac{1}{a}\right| = \frac{1}{\sqrt{2\pi}\,|a|\,\sigma}\exp\left(-\frac{((y-a\mu-b)^2}{2a^2\sigma^2}\right)\,.$$

Thus

$$Y \sim \mathrm{N}(a\mu + b, a^2\sigma^2)\,.$$

$\square$

Remark 1.40. The above example also shows that if $X \sim \mathrm{N}(\mu, \sigma^2)$, then

$$\frac{1}{\sigma}X + \left(-\frac{\mu}{\sigma}\right) \sim \mathrm{N}\left(\frac{1}{\sigma}\mu + \left(-\frac{\mu}{\sigma}\right), \frac{1}{\sigma^2}\sigma^2\right)\,,$$

that is

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

1.5 Lognormal random variables

Lognormal random variables are a particular type of continuous random variables that occur in a number of practical applications.

Definition 1.41. A random variable Y is said to be **lognormal** with parameters μ and σ^2 where $\mu \in \mathbb{R}$ and $\sigma > 0$, if

$$\log Y \sim \mathrm{N}(\mu, \sigma^2)\,.$$

We write

$$Y \sim \mathrm{LogNormal}(\mu, \sigma^2)\,.$$

Note that if $Y \sim \mathrm{LogNormal}(\mu, \sigma^2)$, then $Y = \exp(X)$, where $X \sim \mathrm{N}(\mu, \sigma^2)$.

We will now use the Transformation Formula to determine the pdf of a lognormal random variable.

Proposition 1.42. *If* $Y \sim \mathrm{LogNormal}(\mu, \sigma^2)$ *then*

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right) & \textit{if } y > 0; \\ 0 & \textit{if } y \leq 0. \end{cases}$$

Proof. Write $g(x) = e^x$. Then $Y = g(X)$. Now, if $y = e^x$, then $x = \log y$, so

$$g^{-1}(y) = \log y \quad \text{for } y > 0\,,$$

and

$$\frac{d}{dy} g^{-1}(y) = \frac{1}{y} \quad \text{for } y > 0\,.$$

Moreover, since $X \sim \mathrm{N}(\mu, \sigma^2)$,

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

Using the Transformation Formula we see that if $y > 0$ then

$$f_Y(y) = f_X(\log y)\frac{1}{y} = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right)\,,$$

while $f_Y(y) = 0$ if $y \leq 0$. □

In Coursework 2 you will be asked to prove the following useful results.

Proposition 1.43. *Let* $Y \sim \mathrm{LogNormal}(\mu, \sigma^2)$. *Then*

$$E(Y) = \exp(\mu + \frac{1}{2}\sigma^2) \quad \textit{and} \quad \mathrm{Var}(Y) = \exp(2\mu + \sigma^2)(e^{\sigma^2} - 1)\,.$$

Proof. See Exercise 4, Coursework 2. □

Note that we can calculate the cdf of a lognormally distributed random variable using the table for Φ.

Example 1.44. Suppose $Y \sim \mathrm{LogNormal}(\mu, \sigma^2)$ with $\mu = 0.20$ and $\sigma = 0.50$. Determine y such that $P(Y \leq y) = 0.95$.

Solution. Note that $P(Y \le y) = P(\log Y \le \log y)$ where $\log Y \sim \mathrm{N}(\mu, \sigma^2)$. Thus

$$0.95 = P(Y \le y) = P(\log Y \le \log y) = P\left(\frac{\log Y - \mu}{\sigma} \le \frac{\log y - \mu}{\sigma}\right) = \Phi\left(\frac{\log y - \mu}{\sigma}\right).$$

From the table for Φ we see that

$$\frac{\log y - \mu}{\sigma} = 1.645\,,$$

so

$$y = \exp(\mu + 1.645\sigma) = 2.78\,.$$

□

1.6 The IID lognormal model

This is our first model of stock market prices. Let $S(n)$ denote the price of some product at the end of n time periods, where n is a non-negative integer. The model assumes that $S(n)/S(n-1)$, where $n \in \mathbb{N}$, is a sequence of independent identically distributed random variables with common distribution

$$\frac{S(n)}{S(n-1)} \sim \mathrm{LogNormal}(\mu, \sigma^2)\,.$$

Note that the model makes an assumption about the relative price changes $S(n)/S(n-1)$ from one time period to the next. In practice one is interested in the distribution of the relative price change $S(n)/S(0)$.

Lemma 1.45. *If $S(n)$ is given by the IID lognormal model, then*

$$\frac{S(n)}{S(0)} \sim \mathrm{LogNormal}(n\mu, n\sigma^2) \quad \textit{for any } n \in \mathbb{N}\,.$$

Proof. For $i \in \mathbb{N}$ write

$$Y_i = \frac{S(i)}{S(i-1)}\,.$$

Then

$$\frac{S(n)}{S(0)} = \frac{S(1)}{S(0)}\frac{S(2)}{S(1)}\frac{S(3)}{S(2)} \cdots \frac{S(n)}{S(n-1)} = Y_1 Y_2 Y_3 \cdots Y_n\,,$$

so

$$\log \frac{S(n)}{S(0)} = \sum_{i=1}^{n} \log Y_i\,.$$

But since by our assumption on the model the random variables $\log Y_1, \log Y_2, \ldots, \log Y_n$ are independent with

$$\log Y_i \sim \mathrm{N}(\mu, \sigma^2) \quad \text{for } i = 1, \ldots, n,$$

we see, using Lemma 1.33, that

$$\log \frac{S(n)}{S(0)} \sim \mathrm{N}(n\mu, n\sigma^2)\,.$$

Thus

$$\frac{S(n)}{S(0)} \sim \mathrm{LogNormal}(n\mu, n\sigma^2)\,.$$

□

Example 1.46. Assume that the price of a product at the end of week n is given by the IID lognormal model with parameters $\mu = 0.0165$ and $\sigma = 0.0730$. What is the probability that

(a) the price increases over the first week?

(b) the price increases in each of the first two weeks?

(c) the price is higher at the end of week 2 than at the start?

Solution. Let $S(n)$ denote the price of the product at the end of week n.

(a) The desired probability is $P(S(1) > S(0))$. But

$$P(S(1) > S(0)) = P\left(\frac{S(1)}{S(0)} > 1\right) = P\left(\log\frac{S(1)}{S(0)} > 0\right),$$

where

$$\log\frac{S(1)}{S(0)} \sim \mathrm{N}(\mu, \sigma^2).$$

Thus

$$\begin{aligned} P(S(1) > S(0)) = P\left(\log\frac{S(1)}{S(0)} > 0\right) &= P\left(\frac{\log\frac{S(1)}{S(0)} - \mu}{\sigma} > -\frac{\mu}{\sigma}\right) \\ &= 1 - \Phi\left(-\frac{\mu}{\sigma}\right) = \Phi\left(\frac{\mu}{\sigma}\right) = \Phi(0.23) = 0.5910. \end{aligned}$$

Thus, the probability that the price increases over the first week is 0.5910.

(b) Let $p = 0.5910$ be the probability that the price increases over the first week. Since the random variables $S(n)/S(n-1)$ are independent, the probability that the price increases in each of the first weeks is $p^2 = 0.3493$.

(c) The desired probability is $P(S(2) > S(0))$. But

$$P(S(2) > S(0)) = P\left(\frac{S(2)}{S(0)} > 1\right) = P\left(\log\frac{S(2)}{S(0)} > 0\right),$$

where, by Lemma 1.45,

$$\log\frac{S(2)}{S(0)} \sim \mathrm{N}(2\mu, 2\sigma^2).$$

Thus

$$\begin{aligned} P(S(2) > S(0)) = P\left(\log\frac{S(2)}{S(0)} > 0\right) &= P\left(\frac{\log\frac{S(2)}{S(0)} - 2\mu}{\sqrt{2}\sigma} > -\frac{2\mu}{\sqrt{2}\sigma}\right) \\ &= 1 - \Phi\left(-\frac{\sqrt{2}\mu}{\sigma}\right) = \Phi\left(\frac{\sqrt{2}\mu}{\sigma}\right) = \Phi(0.32) = 0.6255. \end{aligned}$$

Thus, the probability that the price is higher at the end of week 2 than at the start is 0.6255.

□

1.7 The Central Limit Theorem

One of the reasons that Gaussian random variables occur so often is the Central Limit Theorem (CLT), one of the gems of Probability Theory. To motivate the formulation of the CLT suppose for the moment that $X_1, X_2, \ldots, X_n$ are independent $\mathrm{N}(\mu, \sigma^2)$-distributed random variables. Then, by Lemma 1.33,

$$X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)\,,$$

and so

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \sim \mathrm{N}(0,1)\,.$$

The CLT generalises this fact to sequences of independent random variables that need not have Gaussian distributions, but it only holds in the limit as $n \to \infty$. Here is a precise formulation.

Theorem 1.47 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be independent identically distributed random variables with mean $E(X_i) = \mu$ and variance $\mathrm{Var}(X_i) = \sigma^2$, where $\mu \in \mathbb{R}$ and $\sigma > 0$. Define $S_n = \sum_{i=1}^n X_i$. Then*

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le x\right) = \Phi(x) \quad (\forall x \in \mathbb{R})\,.$$

The conclusion of the CLT is often informally expressed as '$\frac{S_n - n\mu}{\sqrt{n}\sigma}$ converges to $\mathrm{N}(0,1)$'. The proof of the CLT relies on the following three auxiliary results.

Lemma 1.48. *Let Y_n be a sequence of random variables and let Z be a random variable. Suppose that their characteristic functions satisfy*

$$\lim_{n\to\infty} G_{Y_n}(\alpha) = G_Z(\alpha) \quad (\forall \alpha \in \mathbb{R})\,.$$

Then

$$\lim_{n\to\infty} P(Y_n \le x) = P(Z \le x) \quad (\forall x \in \mathbb{R})\,.$$

Lemma 1.49. *If $X_1, X_2, \ldots, X_n$ are independent random variables and $S_n = X_1 + \cdots + X_n$, then*

$$G_{S_n}(\alpha) = G_{X_1}(\alpha) G_{X_2}(\alpha) \cdots G_{X_n}(\alpha) \quad (\forall \alpha \in \mathbb{R})\,.$$

Lemma 1.50. *Let Z be a random variable. Then*

$$Z \sim \mathrm{N}(0,1) \quad \textit{if and only if} \quad G_Z(\alpha) = \exp(-\frac{\alpha^2}{2})\,.$$

Proof of the CLT. We shall first prove a special case. Suppose for the moment that $\mu = 0$ and $\sigma = 1$. In order to prove the CLT in this case we need to show that

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n) \le x\right) = \Phi(x)\,.$$

Let

$$Y_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)\,.$$

We have

$$\begin{aligned}
G_{Y_n}(\alpha) &= E(e^{i\alpha Y_n}) \\
&= E(e^{i\alpha \frac{1}{\sqrt{n}}(X_1+\cdots+X_n)}) \\
&= G_{S_n}\left(\frac{\alpha}{\sqrt{n}}\right) && \text{where } S_n = X_1+\cdots+X_n \\
&= G_{X_1}\left(\frac{\alpha}{\sqrt{n}}\right) G_{X_2}\left(\frac{\alpha}{\sqrt{n}}\right)\cdots G_{X_n}\left(\frac{\alpha}{\sqrt{n}}\right) && \text{by Lemma 1.49} \\
&= G_{X_1}\left(\frac{\alpha}{\sqrt{n}}\right)^n,
\end{aligned}$$

where the last equality holds since the X_i's are identically distributed. Moreover,

$$\begin{aligned}
G_{X_1}\left(\frac{\alpha}{\sqrt{n}}\right) &= 1 + \sum_{k=1}^{\infty} \frac{\left(\frac{i\alpha}{\sqrt{n}}\right)^k}{k!} E(X_1^k) \\
&= 1 + \frac{i\alpha}{\sqrt{n}} E(X_1) + \frac{1}{2}\left(\frac{i\alpha}{\sqrt{n}}\right)^2 E(X_1^2) + \cdots \\
&= 1 - \frac{1}{2}\frac{\alpha^2}{n} + \cdots
\end{aligned}$$

since, by hypothesis, $E(X_1) = 0$ and $E(X_1^2) = \mathrm{Var}(X_1) + E(X_1)^2 = 1$. Thus

$$G_{Y_n}(\alpha) = G_{X_1}\left(\frac{\alpha}{\sqrt{n}}\right)^n = \left(1 - \frac{1}{2}\frac{\alpha^2}{n} + \cdots\right)^n = \exp\left(n \log\left(1 - \frac{1}{2}\frac{\alpha^2}{n} + \cdots\right)\right). \tag{1.2}$$

In order to proceed we need the Taylor-Maclaurin expansion of $\log(1-x)$, which we now quickly derive. Note that

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$$

thus

$$\underbrace{\int_0^x \frac{1}{1-t}\,dt}_{=-\log(1-x)} = x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \frac{1}{4}x^4 + \cdots$$

so

$$\log(1-x) = -x - \frac{1}{2}x^2 - \frac{1}{3}x^3 - \frac{1}{4}x^4 - \cdots$$

Using the first term of the above expansion of $\log(1-x)$ in equation 1.2 we see that

$$G_{Y_n}(\alpha) = \exp\left(n\left(-\frac{1}{2}\frac{\alpha^2}{n} - \cdots\right)\right) = \exp\left(-\frac{\alpha^2}{2} - \cdots\right),$$

so

$$\lim_{n\to\infty} G_{Y_n}(\alpha) = \exp\left(-\frac{\alpha^2}{2}\right).$$

Let now $Z \sim \mathrm{N}(0,1)$. By Lemma 1.50,

$$\exp\left(-\frac{\alpha^2}{2}\right) = G_Z(\alpha),$$

so combining the last two equations gives

$$\lim_{n\to\infty} G_{Y_n}(\alpha) = G_Z(\alpha)\,,$$

and Lemma 1.48 now implies

$$\lim_{n\to\infty} P(Y_n \le x) = P(Z \le x) = \Phi(x)\,,$$

that is, we have proved the CLT in the special case where $\mu = 0$ and $\sigma = 1$.

Now we prove the CLT when μ and $\sigma > 0$ are arbitrary, using the fact the CLT is true for $\mu = 0$ and $\sigma = 1$. For $i \in \mathbb{N}$ call

$$W_i = \frac{X_i - \mu}{\sigma}\,.$$

Then

$$E(W_i) = \frac{\mu - \mu}{\sigma} = 0$$

and

$$\mathrm{Var}(W_i) = \frac{1}{\sigma^2}\mathrm{Var}(X_i) = \frac{1}{\sigma^2}\sigma^2 = 1\,.$$

Since the W_i's are independent random variables with mean 0 and variance 1 we can now apply the special case of the CLT we have just established to conclude that

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}(W_1 + \cdots + W_n) \le x\right) = \Phi(x)\,,$$

so

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}\sigma}\left((X_1 - \mu) + \cdots + (X_n - \mu)\right) \le x\right) = \Phi(x)\,,$$

hence

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}\sigma}(X_1 + \cdots + X_n - n\mu) \le x\right) = \Phi(x)\,,$$

thus

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le x\right) = \Phi(x)\,,$$

and the proof of the general case of the CLT is finished. □