14

The central limit theorem

The central limit theorem is a refinement of the law of large numbers. For a large number of independent identically distributed random variables $X_1, \ldots, X_n$, with finite variance, the average $\bar{X}_n$ approximately has a normal distribution, no matter what the distribution of the X_i is. In the first section we discuss the proper normalization of $\bar{X}_n$ to obtain a normal distribution in the limit. In the second section we will use the central limit theorem to approximate probabilities of averages and sums of random variables.

14.1 Standardizing averages

In the previous chapter we saw that the law of large numbers guarantees the convergence to μ of the average $\bar{X}_n$ of n independent random variables $X_1, \ldots, X_n$, all having the same expectation μ and variance σ^2. This convergence was illustrated by Figure 13.1. Closer examination of this figure suggests another phenomenon: for the two distributions considered (i.e., the $Gam(2,1)$ distribution and a bimodal distribution), the probability density function of $\bar{X}_n$ seems to become symmetrical and bell shaped around the expected value μ as n becomes larger and larger. However, the bell collapses into a single spike at μ. Nevertheless, by a proper normalization it is possible to stabilize the bell shape, as we will see.

In order to let the distribution of $\bar{X}_n$ settle down it seems to be a good idea to stabilize the expectation and variance. Since $\mathrm{E}\left[\bar{X}_n\right] = \mu$ for all n, only the variance needs some special attention. In Figure 14.1 we depict the probability density function of the centered average $\bar{X}_n - \mu$ of $Gam(2,1)$ random variables, multiplied by three different powers of n. In the left column we display the density of $n^{\frac{1}{4}}(\bar{X}_n - \mu)$, in the middle column the density of $n^{\frac{1}{2}}(\bar{X}_n - \mu)$, and in the right column the density of $n(\bar{X}_n - \mu)$. These figures suggest that $\sqrt{n}$ is the right factor to stabilize the bell shape.

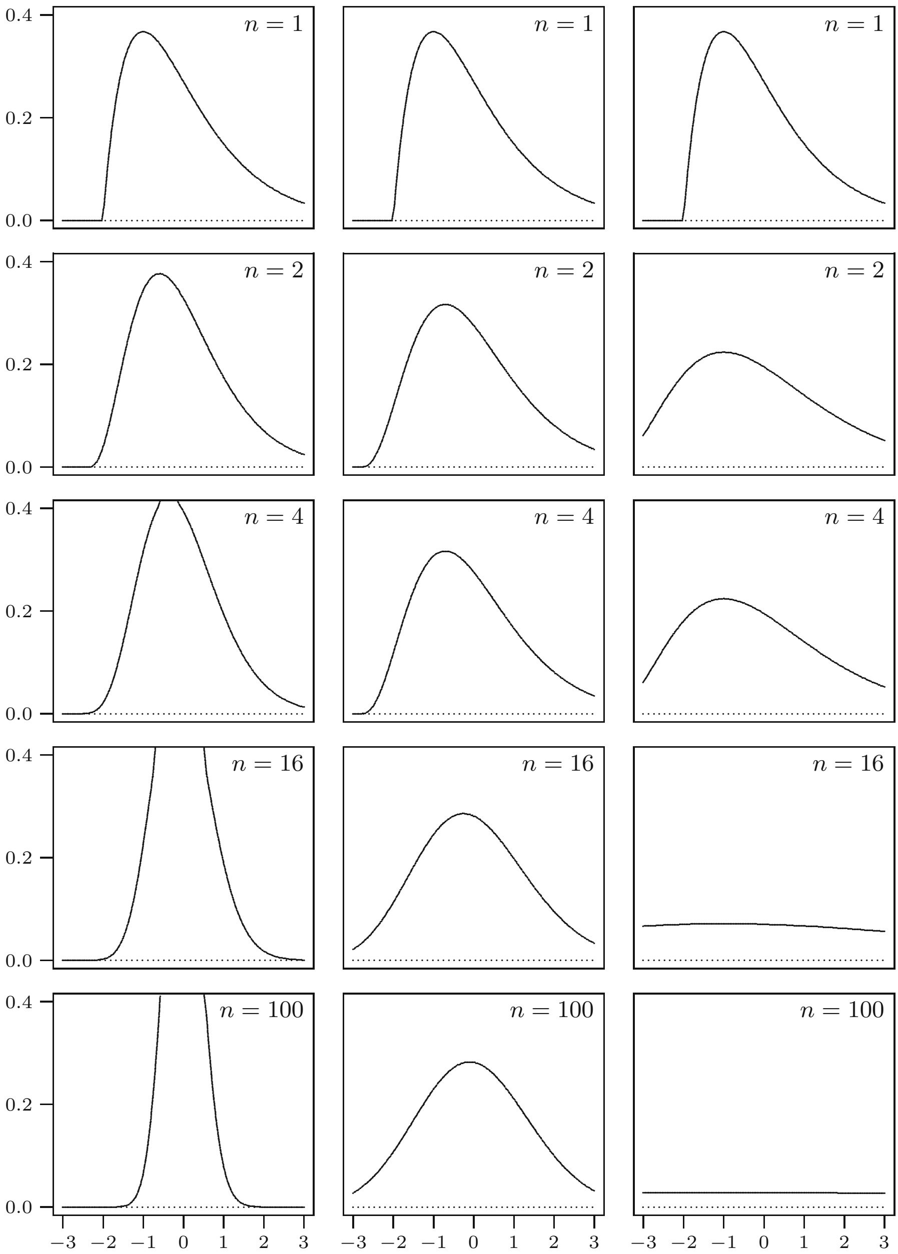

Fig. 14.1. Multiplying the difference $\bar{X}_n - \mu$ of n $Gam(2,1)$ random variables. Left column: $n^{\frac{1}{4}}(\bar{X}_n - \mu)$; middle column: $\sqrt{n}(\bar{X}_n - \mu)$; right column: $n(\bar{X}_n - \mu)$.

Indeed, according to the rule for the variance of an average (see page 182), we have $\mathrm{Var}\big(\bar{X}_n\big) = \sigma^2/n$, and therefore for any number C:

$$\mathrm{Var}\big(C(\bar{X}_n - \mu)\big) = \mathrm{Var}\big(C\bar{X}_n\big) = C^2\mathrm{Var}\big(\bar{X}_n\big) = C^2\frac{\sigma^2}{n}.$$

To stabilize the variance we therefore must choose $C = \sqrt{n}$. In fact, by choosing $C = \sqrt{n}/\sigma$, one *standardizes* the averages, i.e., the resulting random variable Z_n, defined by

$$Z_n = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma}, \quad n = 1, 2, \ldots,$$

has expected value 0 and variance 1. What more can we say about the distribution of the random variables Z_n?

In case $X_1, X_2, \ldots$ are independent $N(\mu, \sigma^2)$ distributed random variables, we know from Section 11.2 and the rule on expectation and variance under change of units (see page 98), that Z_n has an $N(0,1)$ distribution for all n. For the gamma and bimodal random variables from Section 13.1 we depicted the probability density function of Z_n in Figure 14.2. For both examples we see that the probability density functions of the Z_n seem to converge to the probability density function of the $N(0,1)$ distribution, indicated by the dotted line. The following amazing result states that this behavior generally occurs no matter what distribution we start with.

THE CENTRAL LIMIT THEOREM. Let $X_1, X_2, \ldots$ be any sequence of independent identically distributed random variables with finite positive variance. Let μ be the expected value and σ^2 the variance of each of the X_i. For $n \geq 1$, let Z_n be defined by

$$Z_n = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma};$$

then for any number a

$$\lim_{n\to\infty} F_{Z_n}(a) = \Phi(a),$$

where Φ is the distribution function of the $N(0,1)$ distribution. In words: the distribution function of Z_n converges to the distribution function Φ of the standard normal distribution.

Note that

$$Z_n = \frac{\bar{X}_n - \mathrm{E}\big[\bar{X}_n\big]}{\sqrt{\mathrm{Var}\big(\bar{X}_n\big)}},$$

which is a more direct way to see that Z_n is the average $\bar{X}_n$ standardized.

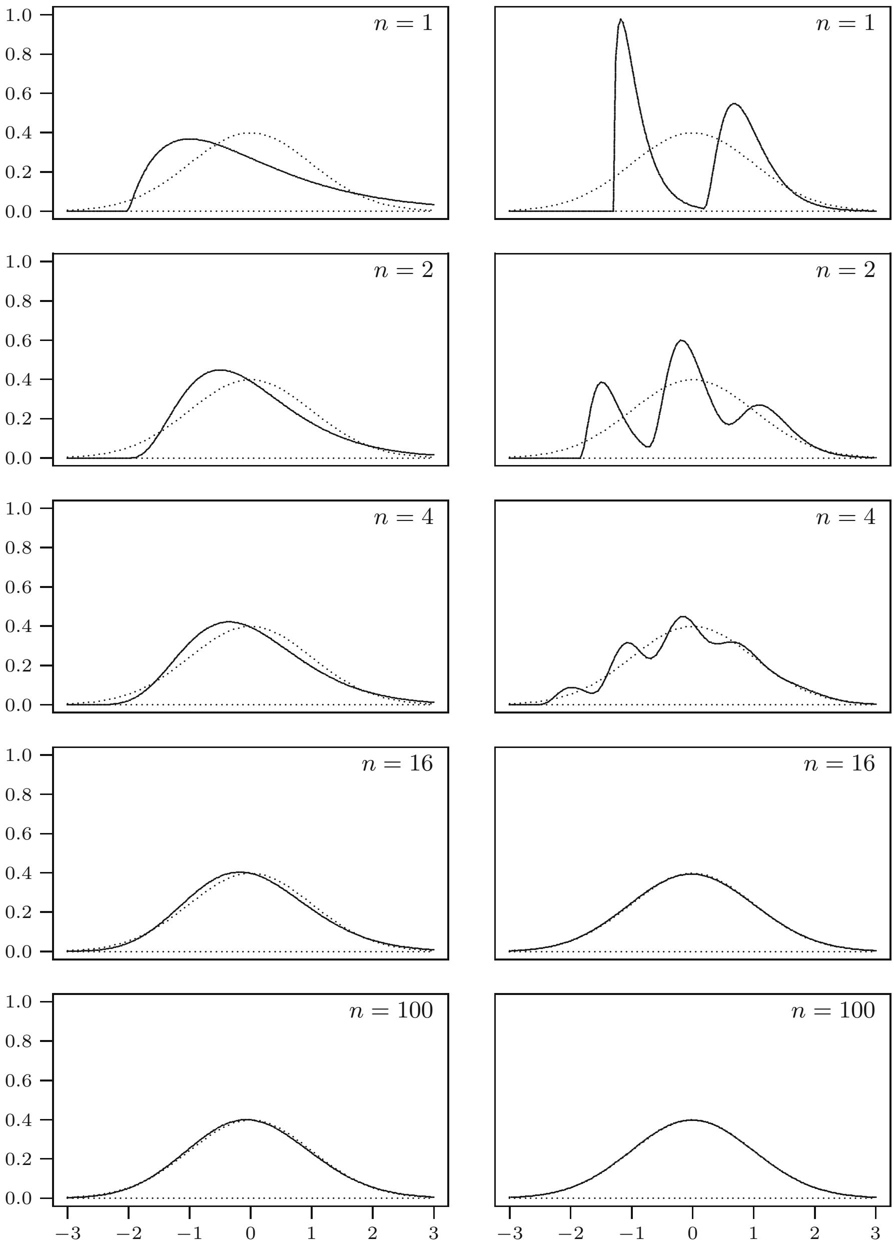

Fig. 14.2. Densities of standardized averages Z_n. Left column: from a gamma density; right column: from a bimodal density. Dotted line: $N(0,1)$ probability density.

One can also write Z_n as a standardized sum

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}. \tag{14.1}$$

In the next section we will see that this last representation of Z_n is very helpful when one wants to approximate probabilities of sums of independent identically distributed random variables.

Since

$$\bar{X}_n = \frac{\sigma}{\sqrt{n}} Z_n + \mu,$$

it follows that $\bar{X}_n$ approximately has an $N(\mu, \sigma^2/n)$ distribution; see the change-of-units rule for normal random variables on page 106. This explains the symmetrical bell shape of the probability densities in Figure 13.1.

Remark 14.1 (Some history). Originally, the central limit theorem was proved in 1733 by De Moivre for independent $Ber(\frac{1}{2})$ distributed random variables. Lagrange extended De Moivre's result to $Ber(p)$ random variables and later formulated the central limit theorem as stated above. Around 1901 a first rigorous proof of this result was given by Lyapunov. Several versions of the central limit theorem exist with weaker conditions than those presented here. For example, for applications it is interesting that it is not necessary that all X_i have the same distribution; see Ross [26], Section 8.3, or Feller [8], Section 8.4, and Billingsley [3], Section 27.

14.2 Applications of the central limit theorem

The central limit theorem provides a tool to approximate the probability distribution of the average or the sum of independent identically distributed random variables. This plays an important role in applications, for instance, see Sections 23.4, 24.1, 26.2, and 27.2. Here we will illustrate the use of the central limit theorem to approximate probabilities of averages and sums of random variables in three examples. The first example deals with an average; the other two concern sums of random variables.

Did we have bad luck?

In the example in Section 13.3 averages of independent $Gam(2, 1)$ distributed random variables were simulated for $n = 1, \ldots, 500$. In Figure 13.2 the realization of $\bar{X}_n$ for $n = 400$ is 1.99, which is almost exactly equal to the expected value 2. For $n = 500$ the simulation was 2.06, a little bit farther away. Did we have bad luck, or is a value 2.06 or higher not unusual? To answer this question we want to compute $\mathrm{P}(\bar{X}_n \geq 2.06)$. We will find an approximation of this probability using the central limit theorem.

Note that

$$\begin{aligned}
\mathrm{P}\big(\bar{X}_n \geq 2.06\big) &= \mathrm{P}\big(\bar{X}_n - \mu \geq 2.06 - \mu\big) \\
&= \mathrm{P}\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \geq \sqrt{n}\frac{2.06 - \mu}{\sigma}\right) \\
&= \mathrm{P}\left(Z_n \geq \sqrt{n}\frac{2.06 - \mu}{\sigma}\right).
\end{aligned}$$

Since the X_i are $Gam(2,1)$ random variables, $\mu = \mathrm{E}[X_i] = 2$ and $\sigma^2 = \mathrm{Var}(X_i) = 2$. We find for $n = 500$ that

$$\begin{aligned}
\mathrm{P}\big(\bar{X}_{500} \geq 2.06\big) &= \mathrm{P}\left(Z_{500} \geq \sqrt{500}\frac{2.06 - 2}{\sqrt{2}}\right) \\
&= \mathrm{P}(Z_{500} \geq 0.95) \\
&= 1 - \mathrm{P}(Z_{500} < 0.95).
\end{aligned}$$

It now follows from the central limit theorem that

$$\mathrm{P}\big(\bar{X}_{500} \geq 2.06\big) \approx 1 - \Phi(0.95) = 0.1711.$$

This is close to the exact answer 0.1710881, which was obtained using the probability density of $\bar{X}_n$ as given in Section 13.1.

Thus we see that there is about a 17% probability that the average $\bar{X}_{500}$ is at least 0.06 above 2. Since 17% is quite large, we conclude that the value 2.06 is not unusual. In other words, we did not have bad luck; $n = 500$ is simply not large enough to be that close. Would 2.06 be unusual if $n = 5000$?

QUICK EXERCISE 14.1 Show that $\mathrm{P}\big(\bar{X}_{5000} \geq 2.06\big) \approx 0.0013$, using the central limit theorem.

Rounding amounts to the nearest integer

In Exercise 13.2 an accountant wanted to simplify his bookkeeping by rounding amounts to the nearest integer, and you were asked to use Chebyshev's inequality to compute an upper bound for the probability

$$p = \mathrm{P}(|X_1 + X_2 + \cdots + X_{100}| > 10)$$

that the cumulative rounding error $X_1 + X_2 + \cdots + X_{100}$ exceeds € 10. This upper bound equals 1/12. In order to know the exact value of p one has to determine the distribution of the sum $X_1 + \cdots + X_{100}$. This is difficult, but the central limit theorem is a handy tool to get an approximation of p. Clearly,

$$p = \mathrm{P}(X_1 + \cdots + X_{100} < -10) + \mathrm{P}(X_1 + \cdots + X_{100} > 10).$$

Standardizing as in (14.1), for the second probability we write, with $n = 100$

$$\begin{aligned}
\mathrm{P}(X_1 + \cdots + X_n > 10) &= \mathrm{P}(X_1 + \cdots + X_n - n\mu > 10 - n\mu) \\
&= \mathrm{P}\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} > \frac{10 - n\mu}{\sigma\sqrt{n}}\right) \\
&= \mathrm{P}\left(Z_n > \frac{10 - n\mu}{\sigma\sqrt{n}}\right).
\end{aligned}$$

The X_i are $U(-0.5, 0.5)$ random variables, $\mu = \mathrm{E}[X_i] = 0$, and $\sigma^2 = \mathrm{Var}(X_i) = 1/12$, so that

$$\mathrm{P}(X_1 + \cdots + X_{100} > 10) = \mathrm{P}\left(Z_{100} > \frac{10 - 100 \cdot 0}{\sqrt{1/12}\sqrt{100}}\right) = \mathrm{P}(Z_{100} > 3.46)\,.$$

It follows from the central limit theorem that

$$\mathrm{P}(Z_{100} > 3.46) \approx 1 - \Phi(3.46) = 0.0003.$$

Similarly,

$$\mathrm{P}(X_1 + \cdots + X_{100} < -10) \approx \Phi(-3.46) = 0.0003.$$

Thus we find that $p = 0.0006$.

Normal approximation of the binomial distribution

In Section 4.3 we considered the (fictitious) situation that you attend, completely unprepared, a multiple-choice exam consisting of 10 questions. We saw that the probability you will pass equals

$$\mathrm{P}(X \geq 6) = 0.0197,$$

where X—being the sum of 10 independent $Ber(\frac{1}{4})$ random variables—has a $Bin(10, \frac{1}{4})$ distribution. As we saw in Chapter 4 it is rather easy, but tedious, to calculate $\mathrm{P}(X \geq 6)$. Although n is small, we investigate what the central limit theorem will yield as an approximation of $\mathrm{P}(X \geq 6)$. Recall that a random variable with a $Bin(n, p)$ distribution can be written as the sum of n independent $Ber(p)$ distributed random variables $R_1, \ldots, R_n$. Substituting $n = 10$, $\mu = p = 1/4$, and $\sigma^2 = p(1-p) = 3/16$, it follows from the central limit theorem that

$$\begin{aligned}
\mathrm{P}(X \geq 6) &= \mathrm{P}(R_1 + \cdots + R_n \geq 6) \\
&= \mathrm{P}\left(\frac{R_1 + \cdots + R_n - n\mu}{\sigma\sqrt{n}} \geq \frac{6 - n\mu}{\sigma\sqrt{n}}\right) \\
&= \mathrm{P}\left(Z_{10} \geq \frac{6 - 2\frac{1}{2}}{\sqrt{\frac{3}{16}}\sqrt{10}}\right) \\
&\approx 1 - \Phi(2.56) = 0.0052.
\end{aligned}$$

The number 0.0052 is quite a poor approximation for the true value 0.0197. Note however, that we could also argue that

$$\begin{aligned}
\mathrm{P}(X \geq 6) &= \mathrm{P}(X > 5) \\
&= \mathrm{P}(R_1 + \cdots + R_n > 5) \\
&= \mathrm{P}\left(Z_{10} \geq \frac{5 - 2\frac{1}{2}}{\sqrt{\frac{3}{16}}\sqrt{10}} \right) \\
&\approx 1 - \Phi(1.83) = 0.0336,
\end{aligned}$$

which gives an approximation that is too large! A better approach lies somewhere in the middle, as the following quick exercise illustrates.

Quick exercise 14.2 Apply the central limit theorem to find 0.0143 as an approximation to $\mathrm{P}\big(X \geq 5\frac{1}{2}\big)$. Since $\mathrm{P}(X \geq 6) = \mathrm{P}\big(X \geq 5\frac{1}{2}\big)$, this also provides an approximation of $\mathrm{P}(X \geq 6)$.

How large should n be?

In view of the previous examples one might raise the question of how large n should be to have a good approximation when using the central limit theorem. In other words, how fast is the convergence to the normal distribution? This is a difficult question to answer in general. For instance, in the third example one might initially be tempted to think that the approximation was quite poor, but after taking the fact into account that we approximate a discrete distribution by a continuous one we obtain a considerable improvement of the approximation, as was illustrated in Quick exercise 14.2. For another example, see Figure 14.2. Here we see that the convergence is slightly faster for the bimodal distribution than for the $Gam(2, 1)$ distribution, which is due to the fact that the $Gam(2, 1)$ is rather asymmetric.

In general the approximation might be poor when n is small, when the distribution of the X_i is asymmetric, bimodal, or discrete, or when the value a in

$$\mathrm{P}\big(\bar{X}_n > a\big)$$

is far from the center of the distribution of the X_i.

14.3 Solutions to the quick exercises

14.1 In the same way we approximated $\mathrm{P}\big(\bar{X}_n \geq 2.06\big)$ using the central limit theorem, we have that

$$\mathrm{P}\big(\bar{X}_n \geq 2.06\big) = \mathrm{P}\left(Z_n \geq \sqrt{n}\frac{2.06 - \mu}{\sigma} \right).$$

With $\mu = 2$ and $\sigma = \sqrt{2}$, we find for $n = 5000$ that

$$\mathrm{P}\left(\bar{X}_{5000} \geq 2.06\right) = \mathrm{P}(Z_{5000} \geq 3),$$

which is approximately equal to $1 - \Phi(3) = 0.0013$, thanks to the central limit theorem. Because we think that 0.13% is a small probability, to find 2.06 as a value for $\bar{X}_{5000}$ would mean that you really had bad luck!

14.2 Similar to the computation $\mathrm{P}(X \geq 6)$, we have

$$\begin{aligned}
\mathrm{P}\left(X \geq 5\frac{1}{2}\right) &= \mathrm{P}\left(R_1 + \cdots + R_{10} \geq 5\frac{1}{2}\right) \\
&= \mathrm{P}\left(Z_{10} \geq \frac{5\frac{1}{2} - 2\frac{1}{2}}{\sqrt{\frac{3}{16}\sqrt{10}}}\right) \\
&\approx 1 - \Phi(2.19) = 0.0143.
\end{aligned}$$

We have seen that using the central limit theorem to approximate $\mathrm{P}(X \geq 6)$ gives an underestimate of this probability, while using the central limit theorem to $\mathrm{P}(X > 5)$ gives an overestimation. Since $5\frac{1}{2}$ is "in the middle," the approximation will be better.

14.4 Exercises

14.1 Let $X_1, X_2, \ldots, X_{144}$ be independent identically distributed random variables, each with expected value $\mu = \mathrm{E}[X_i] = 2$, and variance $\sigma^2 = \mathrm{Var}(X_i) = 4$. Approximate $\mathrm{P}(X_1 + X_2 + \cdots + X_{144} > 144)$, using the central limit theorem.

14.2 ⊡ Let $X_1, X_2, \ldots, X_{625}$ be independent identically distributed random variables, with probability density function f given by

$$f(x) = \begin{cases} 3(1-x)^2 & \text{for } 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Use the central limit theorem to approximate $\mathrm{P}(X_1 + X_2 + \cdots + X_{625} < 170)$.

14.3 ⊞ In Exercise 13.4 **a** you were asked to use Chebyshev's inequality to determine how large n should be (how many people should be interviewed) so that the probability that $\bar{X}_n$ is within 0.2 of the "true" p is at least 0.9. Here p is the proportion of the voters in Florida who will vote for G (and $1 - p$ is the proportion of the voters who will vote for B). How large should n at least be according to the central limit theorem?

14.4 ⊡ In the single-server queue model from Section 6.4, T_i is the time between the arrival of the $(i-1)$th and ith customers. Furthermore, one of the model assumptions is that the T_i are independent, $Exp(0.5)$ distributed random variables. In Section 11.2 we saw that the probability $\mathrm{P}(T_1 + \cdots + T_{30} \leq 60)$ of the 30th customer arriving within an hour at the well is equal to 0.542. Find the normal approximation of this probability.

14.5 ⊞ Let X be a $Bin(n,p)$ distributed random variable. Show that the random variable

$$\frac{X - np}{\sqrt{np(1-p)}}$$

has a distribution that is approximately standard normal.

14.6 ⊡ Again, as in the previous exercise, let X be a $Bin(n,p)$ distributed random variable.

a. An exact computation yields that $\mathrm{P}(X \leq 25) = 0.55347$, when $n = 100$ and $p = 1/4$. Use the central limit theorem to give an approximation of $\mathrm{P}(X \leq 25)$ and $\mathrm{P}(X < 26)$.

b. When $n = 100$ and $p = 1/4$, then $\mathrm{P}(X \leq 2) = 1.87 \cdot 10^{-10}$. Use the central limit theorem to give an approximation of this probability.

14.7 Let $X_1, X_2, \ldots, X_n$ be n independent random variables, each with expected value μ and finite positive variance σ^2. Use Chebyshev's inequality to show that for any $a > 0$ one has

$$\mathrm{P}\left(\left|n^{\frac{1}{4}} \frac{\bar{X}_n - \mu}{\sigma}\right| \geq a\right) \leq \frac{1}{a^2\sqrt{n}}.$$

Use this fact to explain the occurrence of a single spike in the left column of Figure 14.1.

14.8 Let $X_1, X_2, \ldots$ be a sequence of independent $N(0,1)$ distributed random variables. For $n = 1, 2, \ldots$, let Y_n be the random variable, defined by

$$Y_n = X_1^2 + \cdots + X_n^2.$$

a. Show that $\mathrm{E}\left[X_i^2\right] = 1$.

b. One can show—using integration by parts—that $\mathrm{E}\left[X_i^4\right] = 3$. Deduce from this that $\mathrm{Var}\left(X_i^2\right) = 2$.

c. Use the central limit theorem to approximate $\mathrm{P}(Y_{100} > 110)$.

14.9 ⊞ A factory produces links for heavy metal chains. The research lab of the factory models the length (in cm) of a link by the random variable X, with expected value $\mathrm{E}[X] = 5$ and variance $\mathrm{Var}(X) = 0.04$. The length of a link is defined in such a way that the length of a chain is equal to the sum of the lengths of its links. The factory sells chains of 50 meters; to be on the safe side 1002 links are used for such chains. The factory guarantees that the chain is not shorter than 50 meters. If by chance a chain is too short, the customer is reimbursed, and a new chain is given for free.

a. Give an estimate of the probability that for a chain of at least 50 meters more than 1002 links are needed. For what percentage of the chains does the factory have to reimburse clients and provide free chains?

b. The sales department of the factory notices that it has to hand out a lot of free chains and asks the research lab what is wrong. After further investigations the research lab reports to the sales department that the expectation value 5 is incorrect, and that the correct value is 4.99 (cm). Do you think that it was necessary to report such a minor change of this value?

14.10 Chebyshev's inequality was used in Exercise 13.5 to determine how many times n one needs to measure a sample to be 90% sure that the average of the measurements is within half a degree of the actual melting point c of a new material.

a. Use the normal approximation to find a less conservative value for n.

b. Only in case the random errors U_i in the measurements have a normal distribution the value of n from **a** is "exact," in all other cases an approximation. Explain this.