Chapter 7

Limit Theorems

Throughout this section we will assume a probability space $(\Omega, \mathscr{F}, P)$, in which is defined an infinite sequence of random variables (X_n) and a random variable X. The fact that for every infinite sequence of distributions it is possible to construct a probability space with a corresponding sequence of random variables is a non-trivial fact, whose proof is due to Kolmogorov (see for example Billingsley).

7.1 Convergence of sequences of random variables

For every point $\omega \in \Omega$, $(X_n(\omega))$ is a number sequence and $X(\omega)$ is a number. It might be that ω is such that the sequence $(X_n(\omega))$ converges to $X(\omega)$, but it might also be that this sequence does not converge at all, or that it does not converge to $X(\omega)$. The set

$$\left\{\omega \in \Omega \;:\; \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}$$

is an event; as such, it has a probability, which, in principle, could be either zero, one or any intermediate number. The following definition provides a terminology to one of those cases:

Definition 7.1 The sequence of random variable (X_n) is said to converge to the random variable X almost-surely *(כמעט תמיד) (or, w.p. 1) if*

$$P\left(\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

We write $X_n \xrightarrow{a.s} X$.

We can write this mode of convergence in more explicit form. The limit of $(X_n(\omega))$ exists and equals $X(\omega)$ if and only if

$$\forall \epsilon > 0,\ \exists N \in \mathbb{N},\ \forall n > N, \qquad |X_n(\omega) - X(\omega)| < \epsilon.$$

Note that we can replace this condition by the equivalent condition

$$\forall k \in \mathbb{N},\ \exists N \in \mathbb{N},\ \forall n > N, \qquad |X_n(\omega) - X(\omega)| \le \frac{1}{k}.$$

Equivalently, $X(\omega)$ is not the limit of $(X_n(\omega))$ if and only if

$$\exists k \in \mathbb{N},\ \forall N \in \mathbb{N},\ \exists n > N, \qquad |X_n(\omega) - X(\omega)| > \frac{1}{k}.$$

It follows that the condition $X_n \xrightarrow{\text{a.s}} X$ can be reformulated as

$$P\left(\left\{\omega\ :\ \exists k \in \mathbb{N},\ \forall N \in \mathbb{N},\ \exists n > N, \qquad |X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}\right) = 0,$$

or equivalently,

$$P\left(\bigcup_{k=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{n=N+1}^{\infty} \{\omega\ :\ |X_n(\omega) - X(\omega)| > 1/k\}\right) = 0.$$

This can further be written as

$$P\left(\bigcup_{k=1}^{\infty} \limsup_{n\to\infty} \{\omega\ :\ |X_n(\omega) - X(\omega)| > 1/k\}\right) = 0.$$

Note that the sequence of events $\{\omega\ :\ |X_n(\omega) - X(\omega)| > 1/k\}$ is increasing as a function of k, hence also their lim-sups. Thus, this equality is equal to

$$\lim_{k\to\infty} P\left(\limsup_{n\to\infty} \{\omega\ :\ |X_n(\omega) - X(\omega)| > 1/k\}\right) = 0.$$

But if the limit of an increasing non-negative sequence tends to zero, it must be that the sequence is identically zero, namely $X_n \xrightarrow{\text{a.s}} X$ if and only if

$$P\left(\limsup_{n\to\infty} \{\omega\ :\ |X_n(\omega) - X(\omega)| > 1/k\}\right) = 0 \qquad \forall k \in \mathbb{N}.$$

In words, $X_n \xrightarrow{\text{a.s}} X$ if for every k, the probability that $X_n(\omega)$ deviates from $X(\omega)$ by more than $1/k$ for infinitely many n's is zero.

Just like sequences of functions can converge to a limiting functions in more than one way (e.g., pointwise versus uniformly), so a sequence of random variables can converge to a limiting random variable in many different ways. Saying that X_n converges to X is like saying that the sequence of random variables $(X_n - X)$ converges to the (constant) random variable zero. The zero random variable has the property that its second moment is zero. This leads us to the following definition:

Definition 7.2 The sequence (X_n) is said to converge to X in the mean-square *(בתוחלת) if*

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^2\right] = 0.$$

We write $X_n \xrightarrow{m.s} X$.

In words: the sequence $(X_n - X)$ converges to zero in the mean-square if its second moments converge to zero.

A third mode of convergence hinges of the fact that we would relate the fact that (X_n) converges to X with the fact that for every $\epsilon > 0$, the probability that $|X_n - X| > \epsilon$ should tends to zero as $n \to \infty$:

Definition 7.3 The sequence of random variables (X_n) is said to converge to the random variable X in probability *(בהסתברות) if for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P(\{\omega \,:\, |X_n(\omega) - X(\omega)| > \epsilon\}) = 0.$$

We write $X_n \xrightarrow{\Pr} X$.

You might think that this coincides with the mode of convergence we have already defined—almost-sure convergence. We will see that this is not the case. Convergence in probability differs substantially from almost-sure convergence. Also, here too, we might replace $\epsilon > 0$ by $1/k$ for $k \in \mathbb{N}$.

Finally, we might say that the sequence (X_n) converges to X is the distribution of X_n converges to the distribution of X. In this case, we don't even need all the random variables to be defined on the same probability space; each variable could, in principle, belong to a "separate world".

Definition 7.4 The sequence of random variables (X_n) is said to converge to the random variable X in distribution *(בהתפלגות) if for every continuity point $a \in \mathbb{R}$ of*

F_X,

$$\lim_{n\to\infty} F_{X_n}(a) = F_X(a),$$

i.e., if the sequence of distribution functions of the X_n converges point-wise to the distribution function of X at all points where F_X is continuous. We write $X_n \xrightarrow{D} X$.

The first question to be addressed is whether there exists a hierarchy of modes of convergence. We want to know which modes of convergence imply which. The answer is that both almost-sure and mean-square convergence imply convergence in probability, which in turn implies convergence in distribution. On the other hand, almost-sure and mean-square convergence do not imply each other.

Proposition 7.1 *Almost-sure convergence implies convergence in probability.*

Proof: If $X_n \xrightarrow{\text{a.s}} X$, then

$$P\left(\limsup_{n\to\infty} \{\omega \,:\, |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0 \qquad \forall \epsilon > 0.$$

By the Fatou lemma,

$$\limsup_{n\to\infty} P(A_n) \le P\left(\limsup_{n\to\infty} A_n\right),$$

hence

$$\limsup_{n\to\infty} P(\{\omega \,:\, |X_n(\omega) - X(\omega)| > \epsilon\}) \le 0 \qquad \forall \epsilon > 0.$$

Since the sequence $P(\{\omega \,:\, |X_n(\omega) - X(\omega)| > \epsilon\})$ is non-negative, it follows as once that

$$\lim_{n\to\infty} P(\{\omega \,:\, |X_n(\omega) - X(\omega)| > \epsilon\}) = 0 \qquad \forall \epsilon > 0,$$

i.e., $X_n \xrightarrow{\text{Pr}} X$. ■

Proposition 7.2 *Mean-square convergence implies convergence in probability.*

Proof: This is an immediate consequence of the Markov inequality. Let $X_n \xrightarrow{\text{m.s}} X$, then for every $\epsilon > 0$,

$$P(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) = P(\{\omega : |X_n(\omega) - X(\omega)|^2 > \epsilon^2\}) \le \frac{\mathbb{E}|X_n - X|^2}{\epsilon^2},$$

i.e.,

$$\lim_{n\to\infty} P(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) = 0,$$

which implies that $X_n \xrightarrow{\text{Pr}} X$. ∎

Proposition 7.3 Mean-square convergence does not imply almost-sure convergence.

Proof: All we need is a counter example. Consider a family of independent Bernoulli variables X_n with atomistic distributions,

$$p_{X_n}(x) = \begin{cases} 1/n & x = 1 \\ 1 - 1/n & x = 0. \end{cases}$$

The larger n, the more it is likely that $X(\omega) = 0$. Thus, it seems sensible to guess that the sequence of random variables (X_n) converges to the (constant) random variable $X = 0$. The question is in what sense does this convergence occur.

First, we show that $X_n \xrightarrow{\text{m.s}} X$. Indeed,

$$\mathbb{E}[|X_n - X|^2] = \mathbb{E}[X_n^2] = \frac{1}{n} \to 0.$$

On the other hand, (X_n) does not converge to X almost-surely. Since for $\epsilon = 1/2$,

$$\sum_{n=1}^{\infty} P(\{\omega : |X_n(\omega) - X(w)| > 1/2\}) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty,$$

it follows from the second lemma of Borel-Cantelli that

$$P\left(\limsup_{n\to\infty}(\{\omega : |X_n(\omega) - X(w)| > 1/2\}\right) = 1.$$

∎

Proposition 7.4 Almost-sure convergence does not imply mean-square convergence.

Proof: Again, we construct a counter example, this time taking

$$p_{X_n}(x) = \begin{cases} 1/n^2 & x = n^3 \\ 1 - 1/n^2 & x = 0. \end{cases},$$

Once again, the larger n, the more it is likely that $X(\omega) = 0$.

We immediately see that X_n does not converge to X in the mean-square, since

$$\mathbb{E}|X_n - X|^2 = \mathbb{E}[X_n^2] = \frac{n^6}{n^2} \to \infty.$$

Yet, $X_n \xrightarrow{\text{a.s}} X$. For every $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

hence by the first lemma of Borel-Cantelli,

$$P\left(\limsup_{n\to\infty}(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0.$$

■

Corollary 7.1 Convergence in probability does not imply neither almost-sure convergence not convergence in the mean-square.

Proof: Suppose, for example, that convergence in probability implies almost-sure convergence. This would mean that convergence in the mean-square implies almost-sure convergence, which contradicts the last proposition. ■

Finally, we show that convergence in probability implies convergence in distribution, hence both almost-sure convergence and convergence in the mean-square imply convergence in distribution.

Proposition 7.5 *Convergence in probability implies convergence in distribution.*

Proof: Let $a \in R$ be given, and set $\epsilon > 0$. On the one hand

$$\begin{aligned} F_{X_n}(a) &= P\left(X_n \le a\right) \\ &+ P\left(X_n \le a, X \le a+\epsilon\right) + P\left(X_n \le a, X > a+\epsilon\right) \\ &= P\left(X_n \le a | X \le a+\epsilon\right) P\left(X \le a+\epsilon\right) + P\left(X_n \le a, X > a+\epsilon\right) \\ &\le P\left(X \le a+\epsilon\right) + P\left(X_n < X-\epsilon\right) \\ &\le F_X(a+\epsilon) + P\left(|X_n - X| > \epsilon\right), \end{aligned}$$

where we have used the fact that if $A \subset B$ then $P(A) \le P(B)$. By a similar argument

$$\begin{aligned} F_X(a-\epsilon) &= P\left(X \le a-\epsilon, X_n \le a\right) + P\left(X \le a-\epsilon, X_n > a\right) \\ &= P\left(X \le a-\epsilon | X_n \le a\right) P\left(X_n \le a\right) + P\left(X \le a-\epsilon, X_n > a\right) \\ &\le P\left(X_n \le a\right) + P\left(X < X_n - \epsilon\right) \\ &\le F_{X_n}(a) + P\left(|X_n - X| > \epsilon\right), \end{aligned}$$

Thus, we have obtained that

$$F_X(a-\epsilon) - P\left(|X_n - X| > \epsilon\right) \le F_{X_n}(a) \le F_X(a+\epsilon) + P\left(|X_n - X| > \epsilon\right).$$

Taking now $n \to \infty$ we have

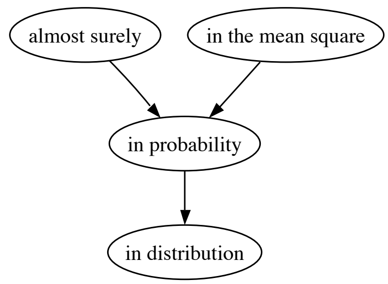
$$F_X(a-\epsilon) \le \liminf_{n\to\infty} F_{X_n}(a) \le \limsup_{n\to\infty} F_{X_n}(a) \le F_X(a+\epsilon).$$

Finally, since this inequality holds for any $\epsilon > 0$ we conclude that provided that a is a continuity point of F_X,

$$\lim_{n\to\infty} F_{X_n}(a) = F_X(a).$$

■

To conclude, the various modes of convergence satisfy the following scheme:



✎ *Exercise 7.1* Prove that if X_n converges in distribution to a constant c, then X_n converges in probability to c.

✎ *Exercise 7.2* Prove that if X_n converges to X in probability then it has a subsequence that converges to X almost-surely.

7.2 The weak law of large numbers

Theorem 7.1 (Weak law of large numbers (החוק החלש של המספרים הגדולים)) *Let X_n be a sequence of independent identically-distributed random variables on a probability space $(\Omega, \mathscr{F}, P)$ and let $\mu = \mathbb{E}[X_i]$. Define the sequence of running averages,*

$$S_n = \frac{X_1 + \cdots + X_n}{n}.$$

Then, S_n converges to μ in probability, i.e., for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\{\omega \,:\, |S_n(\omega) - \mu| > \epsilon\}\right) = 0.$$

Comment: Take the particular case where $X_1, X_2, \ldots$ are i.i.d., $A \subset \mathbb{R}$ and

$$Y_i = I_{X_i \in A} = \begin{cases} 1 & X_i(\omega) \in A \\ 0 & X_i(\omega) \notin A \end{cases}.$$

Then,

$$S_n = \frac{1}{n}\sum_{i=1}^{n} Y_i = \text{fraction of times } X_i(\omega) \in A \text{ for } 1 \le i \le n.$$

The weak law of large numbers states that the fraction of times the outcome is in a given set converges in probability to $E[Y_1]$, which is the probability of this set, $P_{X_1}(A)$, namely, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(\{\omega : |S_n(\omega) - P_{X_1}(A)| > \epsilon\}) = 0.$$

Proof: We will prove the weak law under the additional assumption that the random variables have finite variance $\sigma^2 = \text{Var}[X_i]$. Then, the weak law of large numbers is an immediate consequence of the Chebyshev inequality: by the additivity of the expectation and the variance (for independent random variables),

$$\mathbb{E}[S_n] = \mu \qquad \text{and} \qquad \text{Var}[S_n] = \frac{\sigma^2}{n}.$$

Then,

$$P(\{\omega : |S_n(\omega) - \mu| > \epsilon\}) \leq \frac{\text{Var}[S_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2},$$

which tends to zero as $n \to \infty$. ■

Comment: the first proof is due to Jacob Bernoulli (1713), who proved it for the particular case of binomial variables.

7.3 The strong law of large numbers

Our next limit theorem is the strong law of large number, which states that the running average of a sequence of i.i.d. variables converges to the mean almost-surely (thus strengthening the weak law of large numbers, which only provides convergence in probability).

Theorem 7.2 (Strong law of large numbers (החוק החזק של המספרים הגדולים)) *Let X_n be a sequence of independent, identically distributed random variables with finite expectation $\mu = \mathbb{E}[X_i]$. Define the sequence of running averages,*

$$S_n = \frac{X_1 + \cdots + X_n}{n}.$$

Then, S_n converges to μ almost-surely, i.e., for every $\epsilon > 0$,

$$P(\{\omega : \lim_{n\to\infty} S_n(\omega) = \mu\}) = 1.$$

Proof: We have to prove that for every $\epsilon > 0$,

$$P\left(\limsup_{n\to\infty}\{\omega : |S_n(\omega) - \mu| > \epsilon\}\right) = 0,$$

which by the lemma of first Borel-Cantelli lemma holds if

$$\sum_{n=1}^{\infty} P\left(\{\omega : |S_n(\omega) - \mu| > \epsilon\}\right) < \infty.$$

We will prove the theorem under the additional assumption that the random variables are bounded, i.e., there is $M < \infty$ such that $|X_i| \le M$.

Set

$$Y_k = \frac{X_k - \mu}{2M},$$

The reverse relation is $X_k = 2MY_k + \mu$. Clearly $\mathbb{E}[Y_k] = 0$, and by the boundedness assumption, $|Y_k| \le 1$. Hence, we can use Hoeffding's inequality and get that for every n and a

$$P\left(\sum_{k=1}^{n} Y_k \ge a\right) \le \exp\left(-\frac{a^2}{2n}\right)$$

By our definition of Y_k,

$$\begin{aligned}
\{\omega : |S_n(\omega) - \mu| > \epsilon\} &= \left\{\omega : \left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| > \epsilon\right\} \\
&= \left\{\omega : \left|\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)\right| > \epsilon\right\} \\
&= \left\{\omega : \left|\frac{2M}{n}\sum_{k=1}^{n} Y_k\right| > \epsilon\right\} \\
&\subset \left\{\omega : \frac{2M}{n}\sum_{k=1}^{n} Y_k \ge \epsilon\right\} \cup \left\{\omega : \frac{2M}{n}\sum_{k=1}^{n} Y_k \le -\epsilon\right\} \\
&= \left\{\omega : \sum_{k=1}^{n} Y_k \ge \frac{n\epsilon}{2M}\right\} \cup \left\{\omega : -\sum_{k=1}^{n} Y_k \ge -\frac{n\epsilon}{2M}\right\}.
\end{aligned}$$

By Hoeffding's inequality (once for (Y_k) and once for $(-Y_k)$),

$$P\left(\{\omega : |S_n(\omega) - \mu| > \epsilon\}\right) \le 2\exp\left(-\frac{n\epsilon^2}{8M^2}\right),$$

which is indeed summable.

∎

7.4 The central limit theorem

Theorem 7.3 (Central limit theorem (משפט הגבול המרכזי)) Let (X_n) *be a sequence of i.i.d. random variables with* $\mathbb{E}[X_i] = 0$ *and* $\text{Var}[X_i] = 1$*. Then, the sequence of random variables*

$$S_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

converges in distribution to a random variable $X \sim \mathcal{N}(0,1)$*. That is, for every* $a \in \mathbb{R}$*,*

$$\lim_{n\to\infty} P(S_n \le a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-y^2/2}\, dy.$$

Comments:

① If $\mathbb{E}[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2$ then the same applies for

$$S_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}.$$

② The central limit theorem (CLT) is about a running average rescaled by a factor of $\sqrt{n}$. If we denote by Y_n the running average,

$$Y_n = \frac{X_1 + \cdots + X_n}{n},$$

then the CLT states that

$$P\left(Y_n \le \frac{a}{\sqrt{n}}\right) \sim \Phi(a),$$

i.e., it provides an estimate of the distribution of Y_n at distances $O(n^{-1/2})$ from its mean. It is a theorem about *small deviations* from the mean. There exist more sophisticated theorems about the distribution of Y_n far from the mean, part of the so-called theory of *large deviations*.

③ There are many variants of this theorem.

Proof: We will use the following fact, which we won't prove: if the sequence of moment generating functions $M_{X_n}(t)$ of a sequence of random variables (X_n) converges for every t to the moment generating function $M_X(t)$ of a random variable X, then X_n converges to X in distribution. In other words,

$$M_{X_n}(t) \to M_X(t) \text{ for all } t \text{ implies that } X_n \xrightarrow{D} X.$$

Thus, we need to show that the moment generating functions of the S_n's tends as $n \to \infty$ to $\exp(t^2/2)$, which is the moment generating function of a standard normal variable.

Recall that the PDF of a sum of two random variables is the convolution of their PDF, but the moment generating function of their sum is the product of the their moment generating function. Inductively,

$$M_{X_1+X_2+\ldots,+X_n}(t) = \prod_{i=1}^{n} M_{X_i}(t) = [M_{X_1}(t)]^n,$$

where we have used the fact that they are i.i.d., Now, if a random variable Y has a moment generating function M_Y, then

$$M_{Y/a}(t) = \int_{\mathbb{R}} e^{ty} f_{Y/a}(y)\, dy,$$

but since $f_{Y/a}(y) = a\, f_Y(ay)$ we get that

$$M_{Y/a}(t) = a \int_{\mathbb{R}} e^{ty} f_Y(ay)\, dy = \int_{\mathbb{R}} e^{aty/a} f_Y(ay)\, d(ay) = M_Y(t/a),$$

from which we deduce that

$$M_{S_n}(t) = \left[M_{X_1}\left(\frac{t}{\sqrt{n}} \right) \right]^n.$$

Take the logarithm of both sides, and write the left hand side explicitly,

$$\log M_{S_n}(t) = n \log \int_{\mathbb{R}} e^{tx/\sqrt{n}} f_{X_1}(x)\, dx.$$

Taylor expanding the exponential about $t = 0$ we have,

$$\begin{aligned}
\log M_{S_n}(t) &= n \log \int_{\mathbb{R}} \left(1 + \frac{tx}{\sqrt{n}} + \frac{t^2x^2}{2n} + \frac{t^3x^3}{6n^{3/2}} e^{\xi x/\sqrt{n}} \right) f_{X_1}(x)\, dx \\
&= n \log \left(1 + 0 + \frac{t^2}{2n} + O(n^{-3/2}) \right) \\
&= n \left(\frac{t^2}{2n} + O(n^{-3/2}) \right) \to \frac{t^2}{2}.
\end{aligned}$$

■

Example: Suppose that an experimentalist wants to measure some quantity. He knows that due to various sources of errors, the result of every single measurement is a random variable, whose mean μ is the correct answer, and the variance of his measurement is σ^2. He therefore performs independent measurements and averages the results. How many such measurements does he need to perform to be sure, within 95% certainty, that his estimate does not deviate from the true result by $\sigma/4$?

The question we're asking is how large should n be in order for the inequality

$$P\left(\mu - \frac{\sigma}{4} \le \frac{1}{n}\sum_{k=1}^{n} X_k \le \mu + \frac{\sigma}{4}\right) \ge 0.95$$

to hold. This is equivalent to asking what should n be for

$$P\left(-\frac{\sqrt{n}}{4} \le \frac{1}{\sqrt{n}}\sum_{k=1}^{n} \frac{X_k - \mu}{\sigma} \le \frac{\sqrt{n}}{4}\right) \ge 0.95.$$

By the central limit theorem the right hand side is, for large n, approximately

$$\frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{n}/4} e^{-y^2/2}\, dy,$$

which turns out to be larger than 0.95 for ≥ 62.

The problem with this argument that it uses the assumption that "n is large", but it is not clear what large is. Is $n = 62$ sufficiently large for this argument to hold? This problem could have been solved without this difficulty but resorting instead to the Chebyshev inequality:

$$P\left(-\frac{\sqrt{n}}{4} \le \frac{1}{\sqrt{n}}\sum_{k=1}^{n} \frac{X_k - \mu}{\sigma} \le \frac{\sqrt{n}}{4}\right) = 1 - P\left(\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n} \frac{X_k - \mu}{\sigma}\right| \ge \frac{\sqrt{n}}{4}\right)$$
$$\ge 1 - \frac{16}{n},$$

and the right hand side is larger than 0.95 if

$$n \ge \frac{16}{0.05} = 320.$$

▲▲▲

Example: The number of students X who are going to fail in the exam is a Poisson variable with mean 100, i.e, $X \sim \text{Poi}(100)$. I am going to admit that the exam was too hard if more than 120 student fail. What is the probability for it to happen?

We know the exact answer,

$$P(X \geq 120) = e^{-100} \sum_{k=120}^{\infty} \frac{100^k}{k!},$$

which is a quite useless expression. Let's base our estimate on the central limit theorem as follows: a Poisson variable with mean 100 can be expressed as the sum of one hundred independent variables $X_k \sim \text{Poi}(1)$ (the sum of independent Poisson variables is again a Poisson variable), that is $X = \sum_{k=1}^{100} X_k$. Now,

$$P(X \geq 120) = P\left(\frac{1}{\sqrt{100}} \sum_{k=1}^{100} \frac{X_k - 1}{1} \geq \frac{20}{10}\right),$$

which by the central limit theorem equals approximately,

$$P(X \geq 120) \approx \frac{1}{\sqrt{2\pi}} \int_2^{\infty} e^{-y^2/2}\, dy \approx 0.228.$$

▲▲▲

Example: Let us examine numerically a particular example. Let $X_i \sim \mathcal{E}xp(1)$ be independent exponential variable and set

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - 1).$$

A sum of n independent exponential variables has distribution $\mathcal{G}amma(n, 1)$, i.e., its pdf is

$$\frac{x^{n-1} e^{-x}}{\Gamma(n)}.$$

The density for this sum shifted by n is

$$\frac{(x+n)^{n-1} e^{-(x+n)}}{\Gamma(n)},$$

with $x > -n$ and after dividing by $\sqrt{n}$,

$$f_{S_n}(x) = \sqrt{n}\, \frac{(\sqrt{n}x + n)^{n-1} e^{-(\sqrt{n}x+n)}}{\Gamma(n)},$$

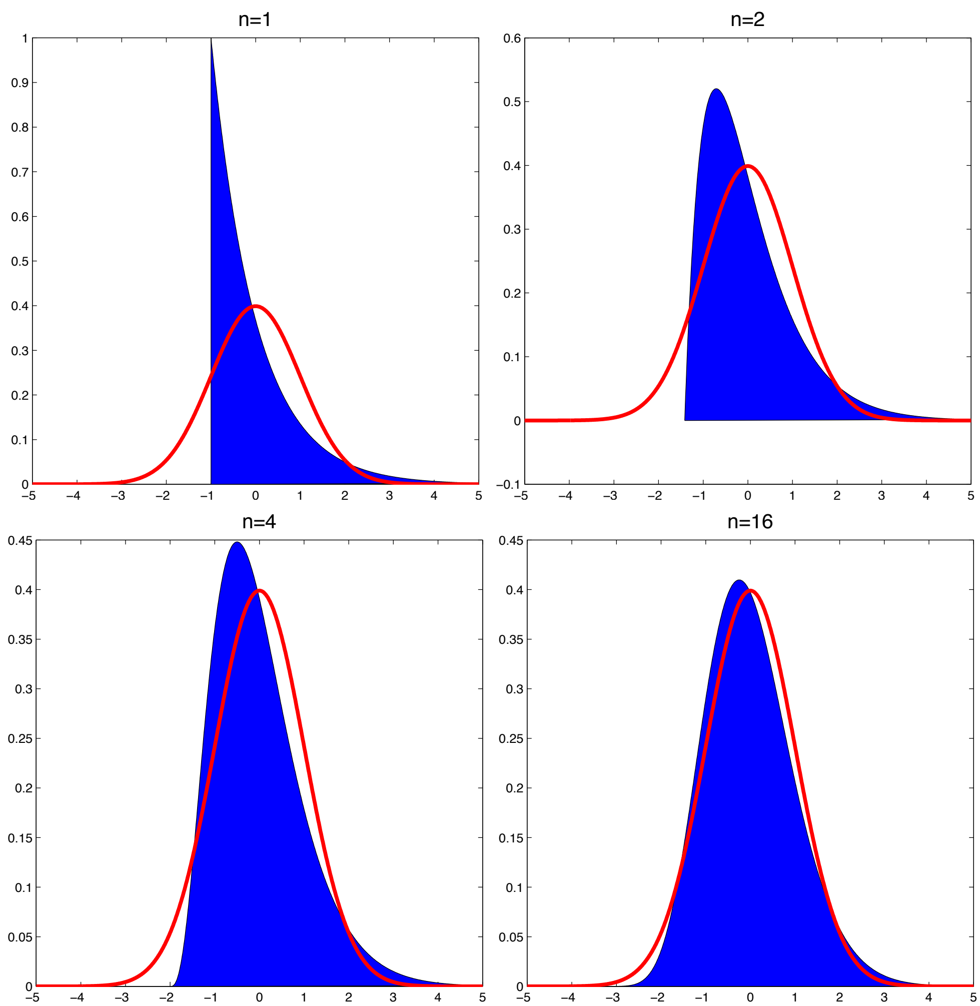


Figure 7.1: The approach of a normalized sum of $1, 2, 4$ and 16 exponential random variables to the normal distribution.

with $x > -\sqrt{n}$. See Figure 7.1 for a visualization of the approach of the distribution of S_n toward the standard normal distribution.

▲ ▲ ▲