Chapter 8

Law of Large Numbers

8.1 Law of Large Numbers for Discrete Random Variables

We are now in a position to prove our first fundamental theorem of probability. We have seen that an intuitive way to view the probability of a certain outcome is as the frequency with which that outcome occurs in the long run, when the experiment is repeated a large number of times. We have also defined probability mathematically as a value of a distribution function for the random variable representing the experiment. The Law of Large Numbers, which is a theorem proved about the mathematical model of probability, shows that this model is consistent with the frequency interpretation of probability. This theorem is sometimes called the *law of averages.* To find out what would happen if this law were not true, see the article by Robert M. Coates.[1]

Chebyshev Inequality

To discuss the Law of Large Numbers, we first need an important inequality called the *Chebyshev Inequality.*

Theorem 8.1 (Chebyshev Inequality) Let X be a discrete random variable with expected value $\mu = E(X)$, and let $\epsilon > 0$ be any positive real number. Then

$$P(|X - \mu| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2}\ .$$

Proof. Let $m(x)$ denote the distribution function of X. Then the probability that X differs from μ by at least ϵ is given by

$$P(|X - \mu| \geq \epsilon) = \sum_{|x-\mu|\geq\epsilon} m(x)\ .$$

[1] R. M. Coates, "The Law," *The World of Mathematics,* ed. James R. Newman (New York: Simon and Schuster, 1956.

We know that

$$V(X) = \sum_x (x-\mu)^2 m(x) \ ,$$

and this is clearly at least as large as

$$\sum_{|x-\mu|\geq\epsilon} (x-\mu)^2 m(x) \ ,$$

since all the summands are positive and we have restricted the range of summation in the second sum. But this last sum is at least

$$\begin{aligned}\sum_{|x-\mu|\geq\epsilon} \epsilon^2 m(x) &= \epsilon^2 \sum_{|x-\mu|\geq\epsilon} m(x) \\ &= \epsilon^2 P(|X-\mu|\geq\epsilon) \ .\end{aligned}$$

So,

$$P(|X-\mu|\geq\epsilon) \leq \frac{V(X)}{\epsilon^2} \ .$$

□

Note that X in the above theorem can be any discrete random variable, and ϵ any positive number.

Example 8.1 Let X by any random variable with $E(X) = \mu$ and $V(X) = \sigma^2$. Then, if $\epsilon = k\sigma$, Chebyshev's Inequality states that

$$P(|X-\mu|\geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2} \ .$$

Thus, for any random variable, the probability of a deviation from the mean of more than k standard deviations is $\leq 1/k^2$. If, for example, $k = 5$, $1/k^2 = .04$. □

Chebyshev's Inequality is the best possible inequality in the sense that, for any $\epsilon > 0$, it is possible to give an example of a random variable for which Chebyshev's Inequality is in fact an equality. To see this, given $\epsilon > 0$, choose X with distribution

$$p_X = \begin{pmatrix} -\epsilon & +\epsilon \\ 1/2 & 1/2 \end{pmatrix} \ .$$

Then $E(X) = 0$, $V(X) = \epsilon^2$, and

$$P(|X-\mu|\geq\epsilon) = \frac{V(X)}{\epsilon^2} = 1 \ .$$

We are now prepared to state and prove the Law of Large Numbers.

Law of Large Numbers

Theorem 8.2 (Law of Large Numbers) Let $X_1, X_2, \ldots, X_n$ be an independent trials process, with finite expected value $\mu = E(X_j)$ and finite variance $\sigma^2 = V(X_j)$. Let $S_n = X_1 + X_2 + \cdots + X_n$. Then for any $\epsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \to 0$$

as $n \to \infty$. Equivalently,

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) \to 1$$

as $n \to \infty$.

Proof. Since $X_1, X_2, \ldots, X_n$ are independent and have the same distributions, we can apply Theorem 6.9. We obtain

$$V(S_n) = n\sigma^2\ ,$$

and

$$V(\frac{S_n}{n}) = \frac{\sigma^2}{n}\ .$$

Also we know that

$$E(\frac{S_n}{n}) = \mu\ .$$

By Chebyshev's Inequality, for any $\epsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}\ .$$

Thus, for fixed ϵ,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \to 0$$

as $n \to \infty$, or equivalently,

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) \to 1$$

as $n \to \infty$. □

Law of Averages

Note that S_n/n is an average of the individual outcomes, and one often calls the Law of Large Numbers the "law of averages." It is a striking fact that we can start with a random experiment about which little can be predicted and, by taking averages, obtain an experiment in which the outcome can be predicted with a high degree of certainty. The Law of Large Numbers, as we have stated it, is often called the "Weak Law of Large Numbers" to distinguish it from the "Strong Law of Large Numbers" described in Exercise 15.

Consider the important special case of Bernoulli trials with probability p for success. Let $X_j = 1$ if the jth outcome is a success and 0 if it is a failure. Then $S_n = X_1 + X_2 + \cdots + X_n$ is the number of successes in n trials and $\mu = E(X_1) = p$. The Law of Large Numbers states that for any $\epsilon > 0$

$$P\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) \to 1$$

as $n \to \infty$. The above statement says that, in a large number of repetitions of a Bernoulli experiment, we can expect the proportion of times the event will occur to be near p. This shows that our mathematical model of probability agrees with our frequency interpretation of probability.

Coin Tossing

Let us consider the special case of tossing a coin n times with S_n the number of heads that turn up. Then the random variable S_n/n represents the fraction of times heads turns up and will have values between 0 and 1. The Law of Large Numbers predicts that the outcomes for this random variable will, for large n, be near 1/2.

In Figure 8.1, we have plotted the distribution for this example for increasing values of n. We have marked the outcomes between .45 and .55 by dots at the top of the spikes. We see that as n increases the distribution gets more and more concentrated around .5 and a larger and larger percentage of the total area is contained within the interval $(.45, .55)$, as predicted by the Law of Large Numbers.

Die Rolling

Example 8.2 Consider n rolls of a die. Let X_j be the outcome of the jth roll. Then $S_n = X_1 + X_2 + \cdots + X_n$ is the sum of the first n rolls. This is an independent trials process with $E(X_j) = 7/2$. Thus, by the Law of Large Numbers, for any $\epsilon > 0$

$$P\left(\left|\frac{S_n}{n} - \frac{7}{2}\right| \geq \epsilon\right) \to 0$$

as $n \to \infty$. An equivalent way to state this is that, for any $\epsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \frac{7}{2}\right| < \epsilon\right) \to 1$$

as $n \to \infty$. □

Numerical Comparisons

It should be emphasized that, although Chebyshev's Inequality proves the Law of Large Numbers, it is actually a very crude inequality for the probabilities involved. However, its strength lies in the fact that it is true for any random variable at all, and it allows us to prove a very powerful theorem.

In the following example, we compare the estimates given by Chebyshev's Inequality with the actual values.

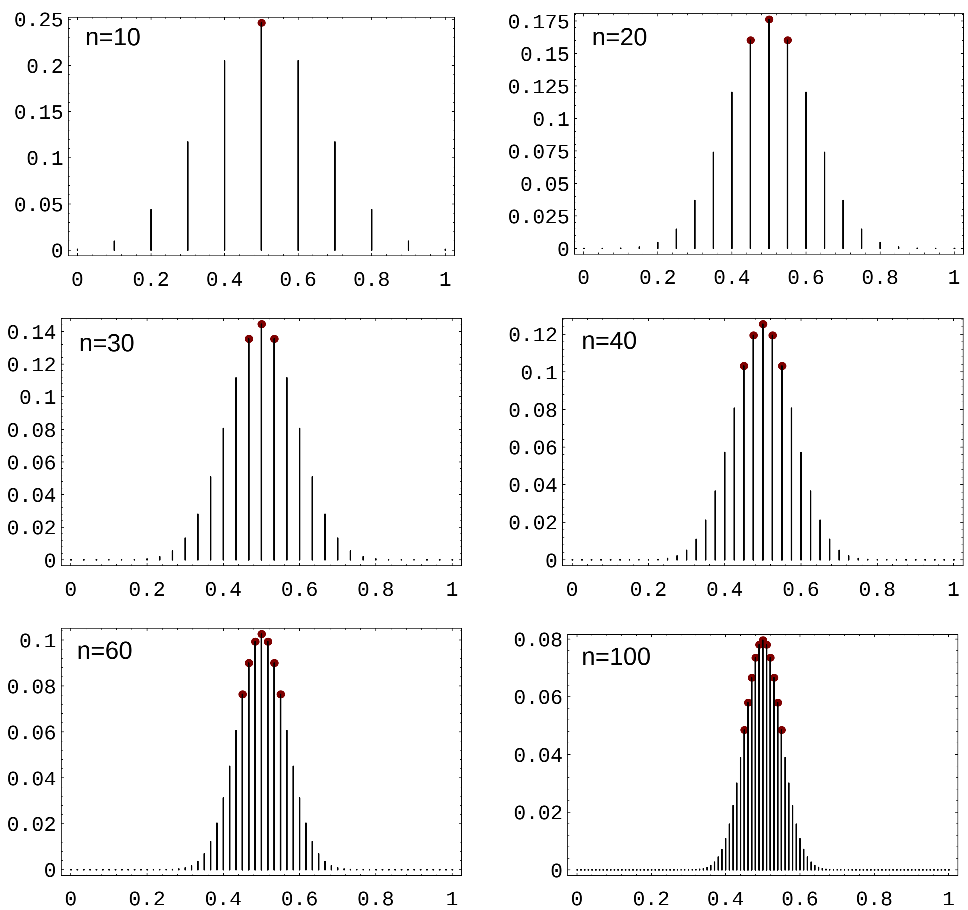

Figure 8.1: Bernoulli trials distributions.

Example 8.3 Let $X_1, X_2, \ldots, X_n$ be a Bernoulli trials process with probability .3 for success and .7 for failure. Let $X_j = 1$ if the jth outcome is a success and 0 otherwise. Then, $E(X_j) = .3$ and $V(X_j) = (.3)(.7) = .21$. If

$$A_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is the *average* of the X_i, then $E(A_n) = .3$ and $V(A_n) = V(S_n)/n^2 = .21/n$. Chebyshev's Inequality states that if, for example, $\epsilon = .1$,

$$P(|A_n - .3| \geq .1) \leq \frac{.21}{n(.1)^2} = \frac{21}{n} \ .$$

Thus, if $n = 100$,

$$P(|A_{100} - .3| \geq .1) \leq .21 \ ,$$

or if $n = 1000$,

$$P(|A_{1000} - .3| \geq .1) \leq .021 \ .$$

These can be rewritten as

$$\begin{aligned} P(.2 < A_{100} < .4) &\geq .79 \ , \\ P(.2 < A_{1000} < .4) &\geq .979 \ . \end{aligned}$$

These values should be compared with the actual values, which are (to six decimal places)

$$\begin{aligned} P(.2 < A_{100} < .4) &\approx .962549 \\ P(.2 < A_{1000} < .4) &\approx 1 \ . \end{aligned}$$

The program **Law** can be used to carry out the above calculations in a systematic way. □

Historical Remarks

The Law of Large Numbers was first proved by the Swiss mathematician James Bernoulli in the fourth part of his work *Ars Conjectandi* published posthumously in 1713.[2] As often happens with a first proof, Bernoulli's proof was much more difficult than the proof we have presented using Chebyshev's inequality. Chebyshev developed his inequality to prove a general form of the Law of Large Numbers (see Exercise 12). The inequality itself appeared much earlier in a work by Bienaymé, and in discussing its history Maistrov remarks that it was referred to as the Bienaymé-Chebyshev Inequality for a long time.[3]

In *Ars Conjectandi* Bernoulli provides his reader with a long discussion of the meaning of his theorem with lots of examples. In modern notation he has an event

[2] J. Bernoulli, *The Art of Conjecturing IV,* trans. Bing Sung, Technical Report No. 2, Dept. of Statistics, Harvard Univ., 1966

[3] L. E. Maistrov, *Probability Theory: A Historical Approach,* trans. and ed. Samual Kotz, (New York: Academic Press, 1974), p. 202

that occurs with probability p but he does not know p. He wants to estimate p by the fraction $\bar{p}$ of the times the event occurs when the experiment is repeated a number of times. He discusses in detail the problem of estimating, by this method, the proportion of white balls in an urn that contains an unknown number of white and black balls. He would do this by drawing a sequence of balls from the urn, replacing the ball drawn after each draw, and estimating the unknown proportion of white balls in the urn by the proportion of the balls drawn that are white. He shows that, by choosing n large enough he can obtain any desired accuracy and reliability for the estimate. He also provides a lively discussion of the applicability of his theorem to estimating the probability of dying of a particular disease, of different kinds of weather occurring, and so forth.

In speaking of the number of trials necessary for making a judgement, Bernoulli observes that the "man on the street" believes the "law of averages."

> Further, it cannot escape anyone that for judging in this way about any event at all, it is not enough to use one or two trials, but rather a great number of trials is required. And sometimes the stupidest man—by some instinct of nature *per se* and by no previous instruction (this is truly amazing)— knows for sure that the more observations of this sort that are taken, the less the danger will be of straying from the mark.[4]

But he goes on to say that he must contemplate another possibility.

> Something futher must be contemplated here which perhaps no one has thought about till now. It certainly remains to be inquired whether after the number of observations has been increased, the probability is increased of attaining the true ratio between the number of cases in which some event can happen and in which it cannot happen, so that this probability finally exceeds any given degree of certainty; or whether the problem has, so to speak, its own asymptote—that is, whether some degree of certainty is given which one can never exceed.[5]

Bernoulli recognized the importance of this theorem, writing:

> Therefore, this is the problem which I now set forth and make known after I have already pondered over it for twenty years. Both its novelty and its very great usefullness, coupled with its just as great difficulty, can exceed in weight and value all the remaining chapters of this thesis.[6]

Bernoulli concludes his long proof with the remark:

> Whence, finally, this one thing seems to follow: that if observations of all events were to be continued throughout all eternity, (and hence the ultimate probability would tend toward perfect certainty), everything in

[4] Bernoulli, op. cit., p. 38.
[5] ibid., p. 39.
[6] ibid., p. 42.

the world would be perceived to happen in fixed ratios and according to a constant law of alternation, so that even in the most accidental and fortuitous occurrences we would be bound to recognize, as it were, a certain necessity and, so to speak, a certain fate.

I do now know whether Plato wished to aim at this in his doctrine of the universal return of things, according to which he predicted that all things will return to their original state after countless ages have past.[7]

Exercises

1 A fair coin is tossed 100 times. The expected number of heads is 50, and the standard deviation for the number of heads is $(100 \cdot 1/2 \cdot 1/2)^{1/2} = 5$. What does Chebyshev's Inequality tell you about the probability that the number of heads that turn up deviates from the expected number 50 by three or more standard deviations (i.e., by at least 15)?

2 Write a program that uses the function binomial(n, p, x) to compute the exact probability that you estimated in Exercise 1. Compare the two results.

3 Write a program to toss a coin 10,000 times. Let S_n be the number of heads in the first n tosses. Have your program print out, after every 1000 tosses, $S_n - n/2$. On the basis of this simulation, is it correct to say that you can expect heads about half of the time when you toss a coin a large number of times?

4 A 1-dollar bet on craps has an expected winning of $-.0141$. What does the Law of Large Numbers say about your winnings if you make a large number of 1-dollar bets at the craps table? Does it assure you that your losses will be small? Does it assure you that if n is very large you will lose?

5 Let X be a random variable with $E(X) = 0$ and $V(X) = 1$. What integer value k will assure us that $P(|X| \geq k) \leq .01$?

6 Let S_n be the number of successes in n Bernoulli trials with probability p for success on each trial. Show, using Chebyshev's Inequality, that for any $\epsilon > 0$

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2} .$$

7 Find the maximum possible value for $p(1-p)$ if $0 < p < 1$. Using this result and Exercise 6, show that the estimate

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{1}{4n\epsilon^2}$$

is valid for any p.

[7] ibid., pp. 65–66.

8 A fair coin is tossed a large number of times. Does the Law of Large Numbers assure us that, if n is large enough, with probability $> .99$ the number of heads that turn up will not deviate from $n/2$ by more than 100?

9 In Exercise 6.2.15, you showed that, for the hat check problem, the number S_n of people who get their own hats back has $E(S_n) = V(S_n) = 1$. Using Chebyshev's Inequality, show that $P(S_n \geq 11) \leq .01$ for any $n \geq 11$.

10 Let X by any random variable which takes on values 0, 1, 2, ..., n and has $E(X) = V(X) = 1$. Show that, for any positive integer k,

$$P(X \geq k+1) \leq \frac{1}{k^2} .$$

11 We have two coins: one is a fair coin and the other is a coin that produces heads with probability 3/4. One of the two coins is picked at random, and this coin is tossed n times. Let S_n be the number of heads that turns up in these n tosses. Does the Law of Large Numbers allow us to predict the proportion of heads that will turn up in the long run? After we have observed a large number of tosses, can we tell which coin was chosen? How many tosses suffice to make us 95 percent sure?

12 (Chebyshev[8]) Assume that $X_1, X_2, \ldots, X_n$ are independent random variables with possibly different distributions and let S_n be their sum. Let $m_k = E(X_k)$, $\sigma_k^2 = V(X_k)$, and $M_n = m_1 + m_2 + \cdots + m_n$. Assume that $\sigma_k^2 < R$ for all k. Prove that, for any $\epsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \frac{M_n}{n}\right| < \epsilon\right) \to 1$$

as $n \to \infty$.

13 A fair coin is tossed repeatedly. Before each toss, you are allowed to decide whether to bet on the outcome. Can you describe a betting system with infinitely many bets which will enable you, in the long run, to win more than half of your bets? (Note that we are disallowing a betting system that says to bet until you are ahead, then quit.) Write a computer program that implements this betting system. As stated above, your program must decide whether to bet on a particular outcome before that outcome is determined. For example, you might select only outcomes that come after there have been three tails in a row. See if you can get more than 50% heads by your "system."

***14** Prove the following analogue of Chebyshev's Inequality:

$$P(|X - E(X)| \geq \epsilon) \leq \frac{1}{\epsilon} E(|X - E(X)|) .$$

[8] P. L. Chebyshev, "On Mean Values," *J. Math. Pure. Appl.*, vol. 12 (1867), pp. 177–184.

***15** We have proved a theorem often called the "Weak Law of Large Numbers." Most people's intuition and our computer simulations suggest that, if we toss a coin a sequence of times, the proportion of heads will really approach 1/2; that is, if S_n is the number of heads in n times, then we will have

$$A_n = \frac{S_n}{n} \to \frac{1}{2}$$

as $n \to \infty$. Of course, we cannot be sure of this since we are not able to toss the coin an infinite number of times, and, if we could, the coin could come up heads every time. However, the "Strong Law of Large Numbers," proved in more advanced courses, states that

$$P\left(\frac{S_n}{n} \to \frac{1}{2}\right) = 1 \ .$$

Describe a sample space Ω that would make it possible for us to talk about the event

$$E = \left\{\omega : \frac{S_n}{n} \to \frac{1}{2}\right\} \ .$$

Could we assign the equiprobable measure to this space? (See Example 2.18.)

***16** In this exercise, we shall construct an example of a sequence of random variables that satisfies the weak law of large numbers, but not the strong law. The distribution of X_i will have to depend on i, because otherwise both laws would be satisfied. (This problem was communicated to us by David Maslen.)

Suppose we have an infinite sequence of mutually independent events $A_1, A_2, \ldots$. Let $a_i = P(A_i)$, and let r be a positive integer.

(a) Find an expression of the probability that none of the A_i with $i > r$ occur.

(b) Use the fact that $x - 1 \le e^{-x}$ to show that

$$P(\text{No } A_i \text{ with } i > r \text{ occurs}) \le e^{-\sum_{i=r}^{\infty} a_i}$$

(c) (The first Borel-Cantelli lemma) Prove that if $\sum_{i=1}^{\infty} a_i$ diverges, then

$$P(\text{infinitely many } A_i \text{ occur}) = 1.$$

Now, let X_i be a sequence of mutually independent random variables such that for each positive integer $i \ge 2$,

$$P(X_i = i) = \frac{1}{2i \log i}, \quad P(X_i = -i) = \frac{1}{2i \log i}, \quad P(X_i = 0) = 1 - \frac{1}{i \log i}.$$

When $i = 1$ we let $X_i = 0$ with probability 1. As usual we let $S_n = X_1 + \cdots + X_n$. Note that the mean of each X_i is 0.

(d) Find the variance of S_n.

(e) Show that the sequence $\langle X_i \rangle$ satisfies the Weak Law of Large Numbers, i.e. prove that for any $\epsilon > 0$

$$P\left(\left|\frac{S_n}{n}\right| \geq \epsilon\right) \to 0 \ ,$$

as n tends to infinity.

We now show that $\{X_i\}$ does not satisfy the Strong Law of Large Numbers. Suppose that $S_n/n \to 0$. Then because

$$\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n}\frac{S_{n-1}}{n-1} \ ,$$

we know that $X_n/n \to 0$. From the definition of limits, we conclude that the inequality $|X_i| \geq \frac{1}{2}i$ can only be true for finitely many i.

(f) Let A_i be the event $|X_i| \geq \frac{1}{2}i$. Find $P(A_i)$. Show that $\sum_{i=1}^{\infty} P(A_i)$ diverges (use the Integral Test).

(g) Prove that A_i occurs for infinitely many i.

(h) Prove that

$$P\left(\frac{S_n}{n} \to 0\right) = 0,$$

and hence that the Strong Law of Large Numbers fails for the sequence $\{X_i\}$.

***17** Let us toss a biased coin that comes up heads with probability p and assume the validity of the Strong Law of Large Numbers as described in Exercise 15. Then, with probability 1,

$$\frac{S_n}{n} \to p$$

as $n \to \infty$. If $f(x)$ is a continuous function on the unit interval, then we also have

$$f\left(\frac{S_n}{n}\right) \to f(p) \ .$$

Finally, we could hope that

$$E\left(f\left(\frac{S_n}{n}\right)\right) \to E(f(p)) = f(p) \ .$$

Show that, if all this is correct, as in fact it is, we would have proven that any continuous function on the unit interval is a limit of polynomial functions. This is a sketch of a probabilistic proof of an important theorem in mathematics called the *Weierstrass approximation theorem.*

8.2 Law of Large Numbers for Continuous Random Variables

In the previous section we discussed in some detail the Law of Large Numbers for discrete probability distributions. This law has a natural analogue for continuous probability distributions, which we consider somewhat more briefly here.

Chebyshev Inequality

Just as in the discrete case, we begin our discussion with the Chebyshev Inequality.

Theorem 8.3 (Chebyshev Inequality) Let X be a continuous random variable with density function $f(x)$. Suppose X has a finite expected value $\mu = E(X)$ and finite variance $\sigma^2 = V(X)$. Then for any positive number $\epsilon > 0$ we have

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \ .$$

□

The proof is completely analogous to the proof in the discrete case, and we omit it.

Note that this theorem says nothing if $\sigma^2 = V(X)$ is infinite.

Example 8.4 Let X be any continuous random variable with $E(X) = \mu$ and $V(X) = \sigma^2$. Then, if $\epsilon = k\sigma = k$ standard deviations for some integer k, then

$$P(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2} \ ,$$

just as in the discrete case. □

Law of Large Numbers

With the Chebyshev Inequality we can now state and prove the Law of Large Numbers for the continuous case.

Theorem 8.4 (Law of Large Numbers) Let $X_1, X_2, \ldots, X_n$ be an independent trials process with a continuous density function f, finite expected value μ, and finite variance σ^2. Let $S_n = X_1 + X_2 + \cdots + X_n$ be the sum of the X_i. Then for any real number $\epsilon > 0$ we have

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0 \ ,$$

or equivalently,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) = 1 \ .$$

□

Note that this theorem is not necessarily true if σ^2 is infinite (see Example 8.8).

As in the discrete case, the Law of Large Numbers says that the average value of n independent trials tends to the expected value as $n \to \infty$, in the precise sense that, given $\epsilon > 0$, the probability that the average value and the expected value differ by more than ϵ tends to 0 as $n \to \infty$.

Once again, we suppress the proof, as it is identical to the proof in the discrete case.

Uniform Case

Example 8.5 Suppose we choose at random n numbers from the interval $[0,1]$ with uniform distribution. Then if X_i describes the ith choice, we have

$$\begin{aligned} \mu &= E(X_i) = \int_0^1 x\,dx = \frac{1}{2}\ , \\ \sigma^2 &= V(X_i) = \int_0^1 x^2\,dx - \mu^2 \\ &= \frac{1}{3} - \frac{1}{4} = \frac{1}{12}\ . \end{aligned}$$

Hence,

$$\begin{aligned} E\left(\frac{S_n}{n}\right) &= \frac{1}{2}\ , \\ V\left(\frac{S_n}{n}\right) &= \frac{1}{12n}\ , \end{aligned}$$

and for any $\epsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \epsilon\right) \leq \frac{1}{12n\epsilon^2}\ .$$

This says that if we choose n numbers at random from $[0,1]$, then the chances are better than $1 - 1/(12n\epsilon^2)$ that the difference $|S_n/n - 1/2|$ is less than ϵ. Note that ϵ plays the role of the amount of error we are willing to tolerate: If we choose $\epsilon = 0.1$, say, then the chances that $|S_n/n - 1/2|$ is less than 0.1 are better than $1 - 100/(12n)$. For $n = 100$, this is about .92, but if $n = 1000$, this is better than .99 and if $n = 10{,}000$, this is better than .999.

We can illustrate what the Law of Large Numbers says for this example graphically. The density for $A_n = S_n/n$ is determined by

$$f_{A_n}(x) = n f_{S_n}(nx)\ .$$

We have seen in Section 7.2, that we can compute the density $f_{S_n}(x)$ for the sum of n uniform random variables. In Figure 8.2 we have used this to plot the density for A_n for various values of n. We have shaded in the area for which A_n would lie between .45 and .55. We see that as we increase n, we obtain more and more of the total area inside the shaded region. The Law of Large Numbers tells us that we can obtain as much of the total area as we please inside the shaded region by choosing n large enough (see also Figure 8.1). $\square$

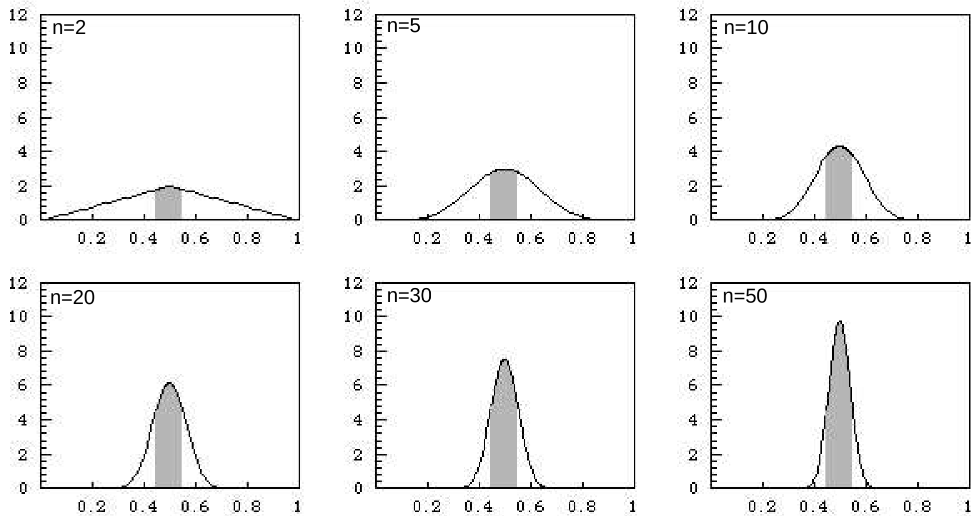

Figure 8.2: Illustration of Law of Large Numbers — uniform case.

Normal Case

Example 8.6 Suppose we choose n real numbers at random, using a normal distribution with mean 0 and variance 1. Then

$$\begin{aligned} \mu &= E(X_i) = 0 \ , \\ \sigma^2 &= V(X_i) = 1 \ . \end{aligned}$$

Hence,

$$\begin{aligned} E\left(\frac{S_n}{n}\right) &= 0 \ , \\ V\left(\frac{S_n}{n}\right) &= \frac{1}{n} \ , \end{aligned}$$

and, for any $\epsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - 0\right| \geq \epsilon\right) \leq \frac{1}{n\epsilon^2} \ .$$

In this case it is possible to compare the Chebyshev estimate for $P(|S_n/n - \mu| \geq \epsilon)$ in the Law of Large Numbers with exact values, since we know the density function for S_n/n exactly (see Example 7.9). The comparison is shown in Table 8.1, for $\epsilon = .1$. The data in this table was produced by the program **LawContinuous**. We see here that the Chebyshev estimates are in general *not* very accurate. □

n	$P(\|S_n/n\| \geq .1)$	Chebyshev
100	.31731	1.00000
200	.15730	.50000
300	.08326	.33333
400	.04550	.25000
500	.02535	.20000
600	.01431	.16667
700	.00815	.14286
800	.00468	.12500
900	.00270	.11111
1000	.00157	.10000

Table 8.1: Chebyshev estimates.

Monte Carlo Method

Here is a somewhat more interesting example.

Example 8.7 Let $g(x)$ be a continuous function defined for $x \in [0,1]$ with values in $[0,1]$. In Section 2.1, we showed how to estimate the area of the region under the graph of $g(x)$ by the Monte Carlo method, that is, by choosing a large number of random values for x and y with uniform distribution and seeing what fraction of the points $P(x,y)$ fell inside the region under the graph (see Example 2.2).

Here is a better way to estimate the same area (see Figure 8.3). Let us choose a large number of independent values X_n at random from $[0,1]$ with uniform density, set $Y_n = g(X_n)$, and find the average value of the Y_n. Then this average is our estimate for the area. To see this, note that if the density function for X_n is uniform,

$$\begin{aligned}\mu &= E(Y_n) = \int_0^1 g(x)f(x)\,dx \\ &= \int_0^1 g(x)\,dx \\ &= \text{average value of } g(x)\ ,\end{aligned}$$

while the variance is

$$\sigma^2 = E((Y_n-\mu)^2) = \int_0^1 (g(x)-\mu)^2\,dx < 1\ ,$$

since for all x in $[0,1]$, $g(x)$ is in $[0,1]$, hence μ is in $[0,1]$, and so $|g(x)-\mu| \leq 1$. Now let $A_n = (1/n)(Y_1 + Y_2 + \cdots + Y_n)$. Then by Chebyshev's Inequality, we have

$$P(|A_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} < \frac{1}{n\epsilon^2}\ .$$

This says that to get within ϵ of the true value for $\mu = \int_0^1 g(x)\,dx$ with probability at least p, we should choose n so that $1/n\epsilon^2 \leq 1-p$ (i.e., so that $n \geq 1/\epsilon^2(1-p)$). Note that this method tells us how large to take n to get a desired accuracy. $\square$

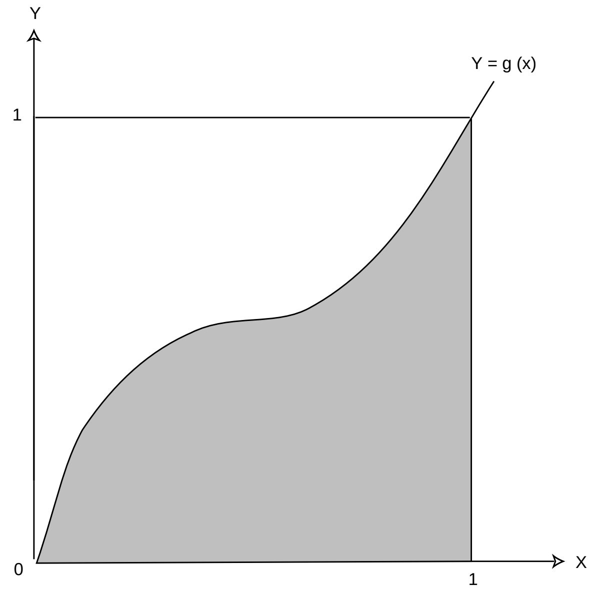

Figure 8.3: Area problem.

The Law of Large Numbers requires that the variance σ^2 of the original underlying density be finite: $\sigma^2 < \infty$. In cases where this fails to hold, the Law of Large Numbers may fail, too. An example follows.

Cauchy Case

Example 8.8 Suppose we choose n numbers from $(-\infty, +\infty)$ with a Cauchy density with parameter $a = 1$. We know that for the Cauchy density the expected value and variance are undefined (see Example 6.28). In this case, the density function for

$$A_n = \frac{S_n}{n}$$

is given by (see Example 7.6)

$$f_{A_n}(x) = \frac{1}{\pi(1 + x^2)} \ ,$$

that is, *the density function for A_n is the same for all n.* In this case, as n increases, the density function does not change at all, and the Law of Large Numbers does not hold. □

Exercises

1 Let X be a continuous random variable with mean $\mu = 10$ and variance $\sigma^2 = 100/3$. Using Chebyshev's Inequality, find an upper bound for the following probabilities.

(a) $P(|X - 10| \geq 2)$.

(b) $P(|X - 10| \geq 5)$.

(c) $P(|X - 10| \geq 9)$.

(d) $P(|X - 10| \geq 20)$.

2 Let X be a continuous random variable with values unformly distributed over the interval $[0, 20]$.

(a) Find the mean and variance of X.

(b) Calculate $P(|X - 10| \geq 2)$, $P(|X - 10| \geq 5)$, $P(|X - 10| \geq 9)$, and $P(|X - 10| \geq 20)$ exactly. How do your answers compare with those of Exercise 1? How good is Chebyshev's Inequality in this case?

3 Let X be the random variable of Exercise 2.

(a) Calculate the function $f(x) = P(|X - 10| \geq x)$.

(b) Now graph the function $f(x)$, and on the same axes, graph the Chebyshev function $g(x) = 100/(3x^2)$. Show that $f(x) \leq g(x)$ for all $x > 0$, but that $g(x)$ is not a very good approximation for $f(x)$.

4 Let X be a continuous random variable with values exponentially distributed over $[0, \infty)$ with parameter $\lambda = 0.1$.

(a) Find the mean and variance of X.

(b) Using Chebyshev's Inequality, find an upper bound for the following probabilities: $P(|X - 10| \geq 2)$, $P(|X - 10| \geq 5)$, $P(|X - 10| \geq 9)$, and $P(|X - 10| \geq 20)$.

(c) Calculate these probabilities exactly, and compare with the bounds in (b).

5 Let X be a continuous random variable with values normally distributed over $(-\infty, +\infty)$ with mean $\mu = 0$ and variance $\sigma^2 = 1$.

(a) Using Chebyshev's Inequality, find upper bounds for the following probabilities: $P(|X| \geq 1)$, $P(|X| \geq 2)$, and $P(|X| \geq 3)$.

(b) The area under the normal curve between -1 and 1 is .6827, between -2 and 2 is .9545, and between -3 and 3 it is .9973 (see the table in Appendix A). Compare your bounds in (a) with these exact values. How good is Chebyshev's Inequality in this case?

6 If X is normally distributed, with mean μ and variance σ^2, find an upper bound for the following probabilities, using Chebyshev's Inequality.

(a) $P(|X - \mu| \geq \sigma)$.

(b) $P(|X - \mu| \geq 2\sigma)$.

(c) $P(|X - \mu| \geq 3\sigma)$.

(d) $P(|X - \mu| \geq 4\sigma)$.

Now find the exact value using the program **NormalArea** or the normal table in Appendix A, and compare.

7 If X is a random variable with mean $\mu \neq 0$ and variance σ^2, define the *relative deviation* D of X from its mean by

$$D = \left| \frac{X - \mu}{\mu} \right| .$$

(a) Show that $P(D \geq a) \leq \sigma^2/(\mu^2 a^2)$.

(b) If X is the random variable of Exercise 1, find an upper bound for $P(D \geq .2)$, $P(D \geq .5)$, $P(D \geq .9)$, and $P(D \geq 2)$.

8 Let X be a continuous random variable and define the *standardized version* X^* of X by:

$$X^* = \frac{X - \mu}{\sigma} .$$

(a) Show that $P(|X^*| \geq a) \leq 1/a^2$.

(b) If X is the random variable of Exercise 1, find bounds for $P(|X^*| \geq 2)$, $P(|X^*| \geq 5)$, and $P(|X^*| \geq 9)$.

9 (a) Suppose a number X is chosen at random from $[0, 20]$ with uniform probability. Find a lower bound for the probability that X lies between 8 and 12, using Chebyshev's Inequality.

(b) Now suppose 20 real numbers are chosen independently from $[0, 20]$ with uniform probability. Find a lower bound for the probability that their average lies between 8 and 12.

(c) Now suppose 100 real numbers are chosen independently from $[0, 20]$. Find a lower bound for the probability that their average lies between 8 and 12.

10 A student's score on a particular calculus final is a random variable with values of $[0, 100]$, mean 70, and variance 25.

(a) Find a lower bound for the probability that the student's score will fall between 65 and 75.

(b) If 100 students take the final, find a lower bound for the probability that the class average will fall between 65 and 75.

11 The Pilsdorff beer company runs a fleet of trucks along the 100 mile road from Hangtown to Dry Gulch, and maintains a garage halfway in between. Each of the trucks is apt to break down at a point X miles from Hangtown, where X is a random variable uniformly distributed over $[0, 100]$.

(a) Find a lower bound for the probability $P(|X - 50| \leq 10)$.

(b) Suppose that in one bad week, 20 trucks break down. Find a lower bound for the probability $P(|A_{20} - 50| \leq 10)$, where A_{20} is the average of the distances from Hangtown at the time of breakdown.

12 A share of common stock in the Pilsdorff beer company has a price Y_n on the nth business day of the year. Finn observes that the price change $X_n = Y_{n+1} - Y_n$ appears to be a random variable with mean $\mu = 0$ and variance $\sigma^2 = 1/4$. If $Y_1 = 30$, find a lower bound for the following probabilities, under the assumption that the X_n's are mutually independent.

(a) $P(25 \leq Y_2 \leq 35)$.

(b) $P(25 \leq Y_{11} \leq 35)$.

(c) $P(25 \leq Y_{101} \leq 35)$.

13 Suppose one hundred numbers $X_1, X_2, \ldots, X_{100}$ are chosen independently at random from $[0, 20]$. Let $S = X_1 + X_2 + \cdots + X_{100}$ be the sum, $A = S/100$ the average, and $S^* = (S - 1000)/(10/\sqrt{3})$ the standardized sum. Find lower bounds for the probabilities

(a) $P(|S - 1000| \leq 100)$.

(b) $P(|A - 10| \leq 1)$.

(c) $P(|S^*| \leq \sqrt{3})$.

14 Let X be a continuous random variable normally distributed on $(-\infty, +\infty)$ with mean 0 and variance 1. Using the normal table provided in Appendix A, or the program **NormalArea**, find values for the function $f(x) = P(|X| \geq x)$ as x increases from 0 to 4.0 in steps of .25. Note that for $x \geq 0$ the table gives $NA(0, x) = P(0 \leq X \leq x)$ and thus $P(|X| \geq x) = 2(.5 - NA(0, x)$. Plot by hand the graph of $f(x)$ using these values, and the graph of the Chebyshev function $g(x) = 1/x^2$, and compare (see Exercise 3).

15 Repeat Exercise 14, but this time with mean 10 and variance 3. Note that the table in Appendix A presents values for a standard normal variable. Find the standardized version X^* for X, find values for $f^*(x) = P(|X^*| \geq x)$ as in Exercise 14, and then rescale these values for $f(x) = P(|X - 10| \geq x)$. Graph and compare this function with the Chebyshev function $g(x) = 3/x^2$.

16 Let $Z = X/Y$ where X and Y have normal densities with mean 0 and standard deviation 1. Then it can be shown that Z has a Cauchy density.

(a) Write a program to illustrate this result by plotting a bar graph of 1000 samples obtained by forming the ratio of two standard normal outcomes. Compare your bar graph with the graph of the Cauchy density. Depending upon which computer language you use, you may or may not need to tell the computer how to simulate a normal random variable. A method for doing this was described in Section 5.2.

(b) We have seen that the Law of Large Numbers does not apply to the Cauchy density (see Example 8.8). Simulate a large number of experiments with Cauchy density and compute the average of your results. Do these averages seem to be approaching a limit? If so can you explain why this might be?

17 Show that, if $X \geq 0$, then $P(X \geq a) \leq E(X)/a$.

18 (Lamperti[9]) Let X be a non-negative random variable. What is the best upper bound you can give for $P(X \geq a)$ if you know

(a) $E(X) = 20$.

(b) $E(X) = 20$ and $V(X) = 25$.

(c) $E(X) = 20$, $V(X) = 25$, and X is symmetric about its mean.

[9] Private communication.