

## 8.1: 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 To find out what would happen if this law were not true, see the article by Robert M. Coates.<sup>1</sup>

### Chebyshev Inequality

To discuss the Law of Large Numbers, we first need an important inequality called the

**(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

$$P(|X - \mu| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2}. \quad (8.1.1)$$

Let  $m(x)$  denote the distribution function of  $X$ . Then the probability that  $X$  differs from  $\mu$  by at least  $\epsilon$  is given by

$$P(|X - \mu| \geq \epsilon) = \sum_{|x - \mu| \geq \epsilon} m(x). \quad (8.1.2)$$

We know that

$$V(X) = \sum_x (x - \mu)^2 m(x), \quad (8.1.3)$$

and this is clearly at least as large as

$$\sum_{|x - \mu| \geq \epsilon} (x - \mu)^2 m(x), \quad (8.1.4)$$

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,

$$P(|X - \mu| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2}. \quad (8.1.5)$$

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

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

$$P(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2 \sigma^2} = \frac{1}{k^2}. \quad (8.1.6)$$

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

$$p_X = \begin{pmatrix} -\epsilon & +\epsilon \\ 1/2 & 1/2 \end{pmatrix}. \quad (8.1.7)$$

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

$$P(|X - \mu| \geq \epsilon) = \frac{V(X)}{\epsilon^2} = 1. \quad (8.1.8)$$

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

## Law of Large Numbers

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

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \rightarrow 0 \quad (8.1.9)$$

as  $n \rightarrow \infty$ . Equivalently,

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) \rightarrow 1 \quad (8.1.10)$$

as  $n \rightarrow \infty$ . Since  $X_1, X_2, \dots, X_n$  are independent and have the same distributions, we can apply Theorem [thm 6.9]. We obtain

$$V(S_n) = n\sigma^2, \quad (8.1.11)$$

and

$$V\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}. \quad (8.1.12)$$

Also we know that

$$E\left(\frac{S_n}{n}\right) = \mu. \quad (8.1.13)$$

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

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}. \quad (8.1.14)$$

Thus, for fixed  $\epsilon$ ,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \rightarrow 0 \quad (8.1.15)$$

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

$$P\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) \rightarrow 1 \quad (8.1.16)$$

as  $n \rightarrow \infty$ .

## 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 [exer 8.1.16].

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

$$P\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) \rightarrow 1 \quad (8.1.17)$$

as  $n \rightarrow \infty$ . 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 [fig 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

Consider  $n$  rolls of a die. Let  $X_j$  be the outcome of the  $j$ th 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  $\epsilon > 0$

$$P\left(\left|\frac{S_n}{n} - \frac{7}{2}\right| \geq \epsilon\right) \rightarrow 0 \quad (8.1.18)$$

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

$$P\left(\left|\frac{S_n}{n} - \frac{7}{2}\right| < \epsilon\right) \rightarrow 1 \quad (8.1.19)$$

as  $n \rightarrow \infty$ .

## 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.

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

$$A_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n} \quad (8.1.20)$$

is the 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,  $\epsilon = .1$ ,

$$P(|A_n - .3| \geq .1) \leq \frac{.21}{n(.1)^2} = \frac{21}{n}. \quad (8.1.21)$$

Thus, if  $n = 100$ ,

$$P(|A_{100} - .3| \geq .1) \leq .21, \quad (8.1.22)$$

or if  $n = 1000$ ,

$$P(|A_{1000} - .3| \geq .1) \leq .021. \quad (8.1.23)$$

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 published posthumously in 1713.<sup>2</sup> 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 [exer 8.1.13]). 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.<sup>3</sup>

In 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 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 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.<sup>4</sup>*

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.<sup>5</sup>*

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 usefulness, coupled with its just as great difficulty, can exceed in weight and value all the remaining chapters of this thesis.<sup>6</sup>*

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 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.<sup>7</sup>*

## Exercises

### Exercise 8.1.1

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)?

### Exercise 8.1.2

Write a program that uses the function  $\text{binomial}(n, p, x)$  to compute the exact probability that you estimated in Exercise [exer 8.1.1]. Compare the two results.

### Exercise 8.1.3

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?

### Exercise 8.1.4

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?

### Exercise 8.1.5

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

### Exercise 8.1.6

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  $\epsilon > 0$

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}. \quad (8.1.24)$$

### Exercise 8.1.7

Find the maximum possible value for  $p(1-p)$  if  $0 < p < 1$ . Using this result and Exercise [exer 8.1.6], show that the estimate

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{1}{4n\epsilon^2} \quad (8.1.25)$$

is valid for any  $p$ .

### Exercise 8.1.8

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?

### Exercise 8.1.9

In Exercise [sec 6.2].[exer 6.2.16], 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  $n \geq 11$ .

### Exercise 8.1.10

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

$$P(X \geq k+1) \leq \frac{1}{k^2}. \quad (8.1.26)$$

### Exercise 8.1.11

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?

### Exercise 8.1.12

(Chebyshev<sup>8</sup>) Assume that  $X_1, X_2, \dots, X_n$  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  $M_n = m_1 + m_2 + \dots + m_n$ . Assume that  $\sigma_k^2 < R$  for all  $k$ . Prove that, for any  $\epsilon > 0$ ,

$$P\left(\left|\frac{S_n}{n} - \frac{M_n}{n}\right| < \epsilon\right) \rightarrow 1 \quad (8.1.27)$$

as  $n \rightarrow \infty$ .

### Exercise 8.1.13

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."

### Exercise 8.1.14

Prove the following analogue of Chebyshev's Inequality:

$$P(|X - E(X)| \geq \epsilon) \leq \frac{1}{\epsilon} E(|X - E(X)|). \quad (8.1.28)$$

### Exercise 8.1.15

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

$$A_n = \frac{S_n}{n} \rightarrow \frac{1}{2} \quad (8.1.29)$$

as  $n \rightarrow \infty$ . 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

$$P\left(\frac{S_n}{n} \rightarrow \frac{1}{2}\right) = 1. \quad (8.1.30)$$

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

$$E = \left\{ \omega : \frac{S_n}{n} \rightarrow \frac{1}{2} \right\}. \quad (8.1.31)$$

Could we assign the equiprobable measure to this space?

### Exercise 8.1.16

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.) .1in Suppose we have an infinite sequence of mutually independent events  $A_1, A_2, \dots$ . Let  $a_i = P(A_i)$ , and let  $r$  be a positive integer.

- Find an expression of the probability that none of the  $A_i$  with  $i > r$  occur.
- Use the fact that  $x - 1 \leq e^{-x}$  to show that

$$P(\text{No } A_i \text{ with } i > r \text{ occurs}) \leq e^{-\sum_{i=r}^{\infty} a_i} \quad (8.1.32)$$

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

$$P(\text{infinitely many } A_i \text{ occur}) = 1. \quad (8.1.33)$$

.1in Now, let  $X_i$  be a sequence of mutually independent random variables such that for each positive integer  $i \geq 2$ ,

$$P(X_i = i) = \frac{1}{2i \log i}, \quad P(X_i = -i) = \frac{1}{2i \log i}, \quad P(X_i = 0) = 1 - \frac{1}{i \log i}. \quad (8.1.34)$$

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

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

$$P\left(\left|\frac{S_n}{n}\right| \geq \epsilon\right) \rightarrow 0, \quad (8.1.35)$$

as  $n$  tends to infinity. .1in We now show that  $\{X_i\}$  does not satisfy the Strong Law of Large Numbers. Suppose that  $S_n/n \rightarrow 0$ . Then because

$$\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1}, \quad (8.1.36)$$

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

- 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).
- Prove that  $A_i$  occurs for infinitely many  $i$ .
- Prove that

$$P\left(\frac{S_n}{n} \rightarrow 0\right) = 0, \quad (8.1.37)$$

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

### Exercise 8.1.17

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 [exer 8.1.16]. Then, with probability 1,

$$\frac{S_n}{n} \rightarrow p \quad (8.1.38)$$

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

$$f\left(\frac{S_n}{n}\right) \rightarrow f(p). \quad (8.1.39)$$

Finally, we could hope that

$$E\left(f\left(\frac{S_n}{n}\right)\right) \rightarrow E(f(p)) = f(p). \quad (8.1.40)$$

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

---

This page titled [8.1: Discrete Random Variables](#) is shared under a [GNU Free Documentation License 1.3](#) license and was authored, remixed, and/or curated by [Charles M. Grinstead & J. Laurie Snell](#) ([American Mathematical Society](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.