

## 6.3: The Law of Large Numbers

### Basic Theory

This section continues the discussion of the sample mean from the last section, but we now consider the more interesting setting where the variables are random. Specifically, suppose that we have a basic random experiment with an underlying probability measure  $\mathbb{P}$ , and that  $X$  is random variable for the experiment. Suppose now that we perform  $n$  independent replications of the basic experiment. This defines a new, compound experiment with a sequence of independent random variables  $\mathbf{X} = (X_1, X_2, \dots, X_n)$ , each with the same distribution as  $X$ . Recall that in statistical terms,  $\mathbf{X}$  is a random sample of size  $n$  from the distribution of  $X$ . All of the relevant statistics discussed in the the previous section, are defined for  $\mathbf{X}$ , but of course now these statistics are random variables with distributions of their own. For the most part, we use the notation established previously, except that for the usual convention of denoting random variables with capital letters. Of course, the deterministic properties and relations established previously apply as well. When we acutally *run* the experiment and observe the values  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  of the random variables, then we are precisely in the setting of the previous section.

Suppose now that the basic variable  $X$  is real valued, and let  $\mu = \mathbb{E}(X)$  denote the expected value of  $X$  and  $\sigma^2 = \text{var}(X)$  the variance of  $X$  (assumed finite). The *sample mean* is

$$M = \frac{1}{n} \sum_{i=1}^n X_i \quad (6.3.1)$$

Often the distribution mean  $\mu$  is unknown and the sample mean  $M$  is used as an estimator of this unknown parameter.

### Moments

The mean and variance of  $M$  are

1.  $\mathbb{E}(M) = \mu$
2.  $\text{var}(M) = \sigma^2/n$

Proof

1. This follows from the linear property of expected value:

$$\mathbb{E}(M) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu \quad (6.3.2)$$

2. This follows from basic properties of variance. Recall in particular that the variance of the sum of independent variables is the sum of the variances.

$$\text{var}(M) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \quad (6.3.3)$$

Part (a) means that the sample mean  $M$  is an *unbiased* estimator of the distribution mean  $\mu$ . Therefore, the variance of  $M$  is the *mean square error*, when  $M$  is used as an estimator of  $\mu$ . Note that the variance of  $M$  is an increasing function of the distribution variance and a decreasing function of the sample size. Both of these make intuitive sense if we think of the sample mean  $M$  as an estimator of the distribution mean  $\mu$ . The fact that the mean square error (variance in this case) decreases to 0 as the sample size  $n$  increases to  $\infty$  means that the sample mean  $M$  is a *consistent* estimator of the distribution mean  $\mu$ .

Recall that  $X_i - M$  is the *deviation* of  $X_i$  from  $M$ , that is, the directed distance from  $M$  to  $X_i$ . The following theorem states that the sample mean is uncorrelated with each deviation, a result that will be crucial for showing the independence of the sample mean and the sample variance when the sampling distribution is normal.

$M$  and  $X_i - M$  are uncorrelated.

Proof

This result follows from simple properties of covariance. Note that  $\text{cov}(M, X_i - M) = \text{cov}(M, X_i) - \text{cov}(M, M)$ . By independence,

$$\text{cov}(M, X_i) = \text{cov}\left(\frac{1}{n} \sum_{j=1}^n X_j, X_i\right) = \frac{1}{n} \sum_{j=1}^n \text{cov}(X_j, X_i) = \frac{1}{n} \text{cov}(X_i, X_i) = \frac{1}{n} \text{var}(X_i) = \frac{\sigma^2}{n} \quad (6.3.4)$$

But by previous theorem,  $\text{cov}(M, M) = \text{var}(M) = \sigma^2/n$ .

## The Weak and Strong Laws of Large Numbers

The *law of large numbers* states that the sample mean converges to the distribution mean as the sample size increases, and is one of the fundamental theorems of probability. There are different versions of the law, depending on the *mode* of convergence.

Suppose again that  $X$  is a real-valued random variable for our basic experiment, with mean  $\mu$  and standard deviation  $\sigma$  (assumed finite). We repeat the basic experiment indefinitely to create a new, compound experiment with an infinite sequence of independent random variables  $(X_1, X_2, \dots)$ , each with the same distribution as  $X$ . In statistical terms, we are *sampling* from the distribution of  $X$ . In probabilistic terms, we have an independent, identically distributed (IID) sequence. For each  $n$ , let  $M_n$  denote the sample mean of the first  $n$  sample variables:

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (6.3.5)$$

From the result above on [variance](#), note that  $\text{var}(M_n) = \mathbb{E}[(M_n - \mu)^2] \rightarrow 0$  as  $n \rightarrow \infty$ . This means that  $M_n \rightarrow \mu$  as  $n \rightarrow \infty$  in mean square. As stated in the next theorem,  $M_n \rightarrow \mu$  as  $n \rightarrow \infty$  in probability as well.

$\mathbb{P}(|M_n - \mu| > \epsilon) \rightarrow 0$  as  $n \rightarrow \infty$  for every  $\epsilon > 0$ .

**Proof**

This follows from Chebyshev's inequality:

$$\mathbb{P}(|M_n - \mu| > \epsilon) \leq \frac{\text{var}(M_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty \quad (6.3.6)$$

Recall that in general, convergence in mean square implies convergence in probability. The convergence of the sample mean to the distribution mean in mean square and in probability are known as *weak laws of large numbers*.

Finally, the *strong law of large numbers* states that the sample mean  $M_n$  converges to the distribution mean  $\mu$  with probability 1. As the name suggests, this is a much stronger result than the weak laws. We will need some additional notation for the proof. First let  $Y_n = \sum_{i=1}^n X_i$  so that  $M_n = Y_n/n$ . Next, recall the definitions of the *positive and negative parts* a real number  $x$ :  $x^+ = \max\{x, 0\}$ ,  $x^- = \max\{-x, 0\}$ . Note that  $x^+ \geq 0$ ,  $x^- \geq 0$ ,  $x = x^+ - x^-$ , and  $|x| = x^+ + x^-$ .

$M_n \rightarrow \mu$  as  $n \rightarrow \infty$  with probability 1.

**Proof**

The proof is in three major steps. The first step is to show that with probability 1,  $M_{n^2} \rightarrow \mu$  as  $n \rightarrow \infty$ . From Chebyshev's inequality,  $\mathbb{P}(|M_{n^2} - \mu| > \epsilon) \leq \sigma^2/n^2\epsilon^2$  for every  $n \in \mathbb{N}_+$  and every  $\epsilon > 0$ . Since  $\sum_{n=1}^{\infty} \sigma^2/n^2\epsilon^2 < \infty$ , it follows from the first Borel-Cantelli lemma that for every  $\epsilon > 0$ ,

$$\mathbb{P}(|M_{n^2} - \mu| > \epsilon \text{ for infinitely many } n \in \mathbb{N}_+) = 0 \quad (6.3.7)$$

Next, from Boole's inequality it follows that

$$\mathbb{P}(\text{For some rational } \epsilon > 0, |M_{n^2} - \mu| > \epsilon \text{ for infinitely many } n \in \mathbb{N}_+) = 0 \quad (6.3.8)$$

This is equivalent to the statement that  $M_{n^2} \rightarrow \mu$  as  $n \rightarrow \infty$  with probability 1.

For our next step, we will show that if the underlying sampling variable is nonnegative, so that  $\mathbb{P}(X \geq 0) = 1$ , then  $M_n \rightarrow \mu$  as  $n \rightarrow \infty$  with probability 1. Note first that with probability 1,  $Y_n$  is increasing in  $n$ . For  $n \in \mathbb{N}_+$ , let  $k_n$  be the unique positive integer such that  $k_n^2 \leq n < (k_n + 1)^2$ . From the increasing property and simple algebra, it follows that with probability 1,

$$\frac{Y_{k_n^2}}{(k_n + 1)^2} \leq \frac{Y_n}{n} \leq \frac{Y_{(k_n+1)^2}}{k_n^2} \quad (6.3.9)$$

From our first step, with probability 1,

$$\frac{Y_{k_n^2}}{(k_n + 1)^2} = \frac{Y_{k_n^2}}{k_n^2} \frac{k_n^2}{(k_n + 1)^2} \rightarrow \mu \text{ as } n \rightarrow \infty \quad (6.3.10)$$

Similarly with probability 1

$$\frac{Y_{(k_n+1)^2}}{k_n^2} = \frac{Y_{(k_n+1)^2}}{(k_n + 1)^2} \frac{(k_n + 1)^2}{k_n^2} \rightarrow \mu \text{ as } n \rightarrow \infty \quad (6.3.11)$$

Finally by the squeeze theorem for limits it follows that with probability 1,  $M_n = Y_n/n \rightarrow \mu$  as  $n \rightarrow \infty$ .

Finally we relax the condition that the underlying sampling variable  $X$  is nonnegative. From step two, it follows that  $\frac{1}{n} \sum_{i=1}^n X_i^+ \rightarrow \mathbb{E}(X^+)$  as  $n \rightarrow \infty$  with probability 1, and  $\frac{1}{n} \sum_{i=1}^n X_i^- \rightarrow \mathbb{E}(X^-)$  as  $n \rightarrow \infty$  with probability 1. Now from algebra and

the linearity of expected value, with probability 1,

$$\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n (X_i^+ - X_i^-) = \frac{1}{n} \sum_{i=1}^n X_i^+ - \frac{1}{n} \sum_{i=1}^n X_i^- \rightarrow \mathbb{E}(X^+) - \mathbb{E}(X^-) = \mathbb{E}(X^+ - X^-) = \mathbb{E}(X) \text{ as } n \rightarrow \infty \quad (6.3.12)$$

The proof of the strong law of large numbers given above requires that the variance of the sampling distribution be finite (note that this is critical in the first step). However, there are better proofs that only require that  $\mathbb{E}(|X|) < \infty$ . An elegant proof showing that  $M_n \rightarrow \mu$  as  $n \rightarrow \infty$  with probability 1 and in mean, using backwards martingales, is given in the chapter on martingales. In the next few paragraphs, we apply the law of large numbers to some of the special statistics studied in the previous section.

### Empirical Probability

Suppose that  $X$  is the outcome random variable for a basic experiment, with sample space  $S$  and probability measure  $\mathbb{P}$ . Now suppose that we repeat the basic experiment indefinitely to form a sequence of independent random variables  $(X_1, X_2, \dots)$  each with the same distribution as  $X$ . That is, we sample from the distribution of  $X$ . For  $A \subseteq S$ , let  $P_n(A)$  denote the empirical probability of  $A$  corresponding to the sample  $(X_1, X_2, \dots, X_n)$ :

$$P_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \in A) \quad (6.3.13)$$

Now of course,  $P_n(A)$  is a random variable for each event  $A$ . In fact, the sum  $\sum_{i=1}^n \mathbf{1}(X_i \in A)$  has the binomial distribution with parameters  $n$  and  $\mathbb{P}(A)$ .

For each event  $A$ ,

1.  $\mathbb{E}[P_n(A)] = \mathbb{P}(A)$
2.  $\text{var}[P_n(A)] = \frac{1}{n} \mathbb{P}(A)[1 - \mathbb{P}(A)]$
3.  $P_n(A) \rightarrow \mathbb{P}(A)$  as  $n \rightarrow \infty$  with probability 1.

Proof

These results follow from the results of this section, since  $P_n(A)$  is the sample mean for the random sample  $\{\mathbf{1}(X_i \in A) : i \in \{1, 2, \dots, n\}\}$  from the distribution of  $\mathbf{1}(X \in A)$ .

This special case of the law of large numbers is central to the very concept of probability: the relative frequency of an event converges to the probability of the event as the experiment is repeated.

### The Empirical Distribution Function

Suppose now that  $X$  is a real-valued random variable for a basic experiment. Recall that the distribution function of  $X$  is the function  $F$  given by

$$F(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R} \quad (6.3.14)$$

Now suppose that we repeat the basic experiment indefinitely to form a sequence of independent random variables  $(X_1, X_2, \dots)$ , each with the same distribution as  $X$ . That is, we sample from the distribution of  $X$ . Let  $F_n$  denote the empirical distribution function corresponding to the sample  $(X_1, X_2, \dots, X_n)$ :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq x), \quad x \in \mathbb{R} \quad (6.3.15)$$

Now, of course,  $F_n(x)$  is a random variable for each  $x \in \mathbb{R}$ . In fact, the sum  $\sum_{i=1}^n \mathbf{1}(X_i \leq x)$  has the binomial distribution with parameters  $n$  and  $F(x)$ .

For each  $x \in \mathbb{R}$ ,

1.  $\mathbb{E}[F_n(x)] = F(x)$
2.  $\text{var}[F_n(x)] = \frac{1}{n} F(x)[1 - F(x)]$
3.  $F_n(x) \rightarrow F(x)$  as  $n \rightarrow \infty$  with probability 1.

Proof

These results follow immediately from the results in this section, since  $F_n(x)$  is the sample mean for the random sample  $\{\mathbf{1}(X_i \leq x) : i \in \{1, 2, \dots, n\}\}$  from the distribution of  $\mathbf{1}(X \leq x)$ .

### Empirical Density for a Discrete Variable

Suppose now that  $X$  is a random variable for a basic experiment with a discrete distribution on a countable set  $S$ . Recall that the probability density function of  $X$  is the function  $f$  given by

$$f(x) = \mathbb{P}(X = x), \quad x \in S \quad (6.3.16)$$

Now suppose that we repeat the basic experiment to form a sequence of independent random variables  $(X_1, X_2, \dots)$  each with the same distribution as  $X$ . That is, we sample from the distribution of  $X$ . Let  $f_n$  denote the empirical probability density function corresponding to the sample  $(X_1, X_2, \dots, X_n)$ :

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i = x), \quad x \in S \quad (6.3.17)$$

Now, of course,  $f_n(x)$  is a random variable for each  $x \in S$ . In fact, the sum  $\sum_{i=1}^n \mathbf{1}(X_i = x)$  has the binomial distribution with parameters  $n$  and  $f(x)$ .

For each  $x \in S$ ,

1.  $\mathbb{E}[f_n(x)] = f(x)$
2.  $\text{var}[f_n(x)] = \frac{1}{n} f(x) [1 - f(x)]$
3.  $f_n(x) \rightarrow f(x)$  as  $n \rightarrow \infty$  with probability 1.

Proof

These results follow immediately from the results in this section, since  $f_n(x)$  is the sample mean for the random sample  $\{\mathbf{1}(X_i = x) : i \in \{1, 2, \dots, n\}\}$  from the distribution of  $\mathbf{1}(X = x)$ .

Recall that a countable intersection of events with probability 1 still has probability 1. Thus, in the context of the previous theorem, we actually have

$$\mathbb{P}[f_n(x) \rightarrow f(x) \text{ as } n \rightarrow \infty \text{ for every } x \in S] = 1 \quad (6.3.18)$$

### Empirical Density for a Continuous Variable

Suppose now that  $X$  is a random variable for a basic experiment, with a continuous distribution on  $S \subseteq \mathbb{R}^d$ , and that  $X$  has probability density function  $f$ . Technically,  $f$  is the probability density function with respect to the standard (Lebesgue) measure  $\lambda_d$ . Thus, by definition,

$$\mathbb{P}(X \in A) = \int_A f(x) dx, \quad A \subseteq S \quad (6.3.19)$$

Again we repeat the basic experiment to generate a sequence of independent random variables  $(X_1, X_2, \dots)$  each with the same distribution as  $X$ . That is, we sample from the distribution of  $X$ . Suppose now that  $\mathcal{A} = \{A_j : j \in J\}$  is a partition of  $S$  into a countable number of subsets, each with positive, finite size. Let  $f_n$  denote the empirical probability density function corresponding to the sample  $(X_1, X_2, \dots, X_n)$  and the partition  $\mathcal{A}$ :

$$f_n(x) = \frac{P_n(A_j)}{\lambda_d(A_j)} = \frac{1}{n \lambda_d(A_j)} \sum_{i=1}^n \mathbf{1}(X_i \in A_j); \quad j \in J, x \in A_j \quad (6.3.20)$$

Of course now,  $f_n(x)$  is a random variable for each  $x \in S$ . If the partition is sufficiently fine (so that  $\lambda_d(A_j)$  is small for each  $j$ ), and if the sample size  $n$  is sufficiently large, then by the law of large numbers,

$$f_n(x) \approx f(x), \quad x \in S \quad (6.3.21)$$

## Exercises

### Simulation Exercises

In the dice experiment, recall that the dice scores form a random sample from the specified die distribution. Select the average random variable, which is the sample mean of the sample of dice scores. For each die distribution, start with 1 die and increase the sample size  $n$ . Note how the distribution of the sample mean begins to resemble a point mass distribution. Note also that the mean of the sample mean stays the same, but the standard deviation of the sample mean decreases. For selected values of  $n$  and selected die distributions, run the simulation 1000 times and compare the relative frequency function of the sample mean to the true probability density function, and compare the empirical moments of the sample mean to the true moments.

Several apps in this project are simulations of random experiments with events of interest. When you run the experiment, you are performing independent replications of the experiment. In most cases, the app displays the relative frequency of the event and its complement, both graphically in blue, and numerically in a table. When you run the experiment, the relative frequencies are shown graphically in red and also numerically.

In the simulation of Buffon's coin experiment, the event of interest is that the coin crosses a crack. For various values of the parameter (the radius of the coin), run the experiment 1000 times and compare the relative frequency of the event to the true probability.

In the simulation of Bertrand's experiment, the event of interest is that a "random chord" on a circle will be longer than the length of a side of the inscribed equilateral triangle. For each of the various models, run the experiment 1000 times and compare the relative frequency of the event to the true probability.

Many of the apps in this project are simulations of experiments which result in discrete variables. When you run the simulation, you are performing independent replications of the experiment. In most cases, the app displays the true probability density function numerically in a table and visually as a blue bar graph. When you run the simulation, the relative frequency function is also shown numerically in the table and visually as a red bar graph.

In the simulation of the binomial coin experiment, select the number of heads. For selected values of the parameters, run the simulation 1000 times and compare the sample mean to the distribution mean, and compare the empirical density function to the probability density function.

In the simulation of the matching experiment, the random variable is the number of matches. For selected values of the parameter, run the simulation 1000 times and compare the sample mean and the distribution mean, and compare the empirical density function to the probability density function.

In the poker experiment, the random variable is the type of hand. Run the simulation 1000 times and compare the empirical density function to the true probability density function.

Many of the apps in this project are simulations of experiments which result in variables with continuous distributions. When you run the simulation, you are performing independent replications of the experiment. In most cases, the app displays the true probability density function visually as a blue graph. When you run the simulation, an empirical density function, based on a partition, is also shown visually as a red bar graph.

In the simulation of the gamma experiment, the random variable represents a random arrival time. For selected values of the parameters, run the experiment 1000 times and compare the sample mean to the distribution mean, and compare the empirical density function to the probability density function.

In the special distribution simulator, select the normal distribution. For various values of the parameters (the mean and standard deviation), run the experiment 1000 times and compare the sample mean to the distribution mean, and compare the empirical density function to the probability density function.

### Probability Exercises

Suppose that  $X$  has probability density function  $f(x) = 12x^2(1-x)$  for  $0 \leq x \leq 1$ . The distribution of  $X$  is a member of the beta family. Compute each of the following

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$
3.  $\mathbb{P}(X \leq \frac{1}{2})$

Answer

1.  $\frac{3}{5}$
2.  $\frac{1}{25}$
3.  $\frac{5}{16}$

Suppose now that  $(X_1, X_2, \dots, X_9)$  is a random sample of size 9 from the distribution in the previous problem. Find the expected value and variance of each of the following random variables:

1. The sample mean  $M$
2. The empirical probability  $P([0, \frac{1}{2}])$

Answer

1.  $\frac{3}{5}, \frac{1}{225}$
2.  $\frac{5}{16}, \frac{55}{2304}$

Suppose that  $X$  has probability density function  $f(x) = \frac{3}{x^4}$  for  $1 \leq x < \infty$ . The distribution of  $X$  is a member of the Pareto family. Compute each of the following

1.  $\mathbb{E}(X)$
2.  $\text{var}(X)$

3.  $\mathbb{P}(2 \leq X \leq 3)$

Answer

1.  $\frac{3}{2}$
2.  $\frac{3}{4}$
3.  $\frac{19}{216}$

Suppose now that  $(X_1, X_2, \dots, X_{16})$  is a random sample of size 16 from the distribution in the previous problem. Find the expected value and variance of each of the following random variables:

1. The sample mean  $M$
2. The empirical probability  $P([2, 3])$

Answer

1.  $\frac{3}{2}, \frac{3}{64}$
2.  $\frac{19}{216}, \frac{3743}{746496}$

Recall that for an *ace-six flat die*, faces 1 and 6 have probability  $\frac{1}{4}$  each, while faces 2, 3, 4, and 5 have probability  $\frac{1}{8}$  each. Let  $X$  denote the score when an ace-six flat die is thrown. Compute each of the following:

1. The probability density function  $f(x)$  for  $x \in \{1, 2, 3, 4, 5, 6\}$
2. The distribution function  $F(x)$  for  $x \in \{1, 2, 3, 4, 5, 6\}$
3.  $\mathbb{E}(X)$
4.  $\text{var}(X)$

Answer

1.  $f(x) = \frac{1}{4}, x \in \{1, 6\}; f(x) = \frac{1}{8}, x \in \{2, 3, 4, 5\}$
2.  $F(1) = \frac{1}{4}, F(2) = \frac{3}{8}, F(3) = \frac{1}{2}, F(4) = \frac{5}{8}, F(5) = \frac{3}{4}, F(6) = 1$
3.  $\mathbb{E}(X) = \frac{7}{2}$
4.  $\text{var}(X) = \frac{15}{4}$

Suppose now that an ace-six flat die is thrown  $n$  times. Find the expected value and variance of each of the following random variables:

1. The empirical probability density function  $f_n(x)$  for  $x \in \{1, 2, 3, 4, 5, 6\}$
2. The empirical distribution function  $F_n(x)$  for  $x \in \{1, 2, 3, 4, 5, 6\}$
3. The average score  $M$

Answer

1.  $\mathbb{E}[f_n(x)] = \frac{1}{4}, x \in \{1, 6\}; \mathbb{E}[f_n(x)] = \frac{1}{8}, x \in \{2, 3, 4, 5\}$
2.  $\text{var}[f_n(x)] = \frac{3}{16n}, x \in \{1, 6\}; \text{var}[f_n(x)] = \frac{7}{64n}, x \in \{2, 3, 4, 5\}$
3.  $\mathbb{E}[F_n(1)] = \frac{1}{4}, \mathbb{E}[F_n(2)] = \frac{3}{8}, \mathbb{E}[F_n(3)] = \frac{1}{2}, \mathbb{E}[F_n(4)] = \frac{5}{8}, \mathbb{E}[F_n(5)] = \frac{3}{4}, \mathbb{E}[F_n(6)] = 1$
4.  $\text{var}[F_n(1)] = \frac{3}{16n}, \text{var}[F_n(2)] = \frac{15}{64n}, \text{var}[F_n(3)] = \frac{1}{4n}, \text{var}[F_n(4)] = \frac{15}{64n}, \text{var}[F_n(5)] = \frac{3}{16n}, \text{var}[F_n(6)] = 0$
5.  $\mathbb{E}(M) = \frac{7}{2}, \text{var}(M) = \frac{15}{4n}$

This page titled [6.3: The Law of Large Numbers](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.