CHAPTER 8: Law of Large Numbers

Law of Large Numbers for Discrete Random Variables

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.

ChebyShev Inequality

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

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

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

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

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

Chebyshevs 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 Chebyshevs 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 $\mathrm{E}(X) = 0$, $\mathrm{V}(X) = \epsilon^2$, and

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

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

Alternatively, Chebyshev's theorem says that the proportion of any distribution that lies within k standard deviations of the mean is at least: $1-\frac{1}{k^2}$, where k is any positive number greater than 1. This theorem applies to **all distributions**.

For example, Chebyshev's theorem says that within 2 standard deviations of the mean, you will find at least

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4},$$

or at least 75% of data will lie within 2 standard deviations from the mean.

Let $X_1, X_2, \ldots, X_n$ be a Bernoulli trials process with probability .3 for success and .7 for failure. Then, $\mathrm{E}(X_j) = .3$ and $\mathrm{V}(X_j) = .3 \cdot .7 = .21$. If

$$A_n = \frac{S_n}{n} = \frac{X_1, X_2, \ldots, X_n}{n}$$

is the average of the X_j, then $\mathrm{E}(A_n) = .3$ and $\mathrm{V}(A_n) = \frac{\mathrm{V}(S_n)}{n^2} = \frac{.21}{n}$.

Let $\epsilon = .1$. Then, Chebyshev's Inequality states that

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

If $n = 100$, this can be rewritten as

$$\mathbb{P}(.2 < A_{100} < .4) \geq \mathbb{P}_{A_{100}}$$

where $\mathbb{P}_{A_{100}}$ is the probability that the mean of A_n lies between .2 and .4, which is .1 away from $\mathrm{E}(A_n) = .3$. By Chebyshev, we can calculate this by finding how many standard deviations .1 is. The standard deviation, $\sigma = \sqrt{\frac{.21}{100}} = .0458$. Thus, $k = \frac{.1}{.0458} = 2.182$. Thus,

$$\mathbb{P}_{A_{100}} = 1 - \frac{1}{2.182^2} = .79 = 79\%.$$

Law of Large Numbers

Theorem 8.2: Law of Large Numbers Let $X_1, X_2, \ldots, X_n$ be an independent trial process, with finite expected value $\mu = \mathrm{E}(X_j)$, and finite variance $\sigma = \mathrm{V}(X_j)$. Then for any $\epsilon > 0$,

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

as $n \to \infty$. Equivalently,

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

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 $n \to \infty$. □

The **Law of Large Numbers** states that as the number of observations increases, the actual or observed probability approaches the theoretical or expected probabilities.

Weak Law of Large Numbers vs. **Strong Law of Large Numbers**
The **WLLN** says that for a large sample, there is a very high probability that the mean of the sample will be very close to the expected value, due to *weak convergence*. However, the probability of convergence is still very high.

The **SLLN** says that for a large sample, the mean of the sample will almost surely be the expected value, due to *strong convergence*. The proof of the **SLLN** is much more complex and much harder to prove.

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 $n \to \infty$. □

There are exercises for **LLN** on page 312.

Law of Large Numbers for Discrete Random Variables

This law has a natural analogue for continuous probability distributions, which we consider somewhat more briefly here.

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 $\mu\mathrm{E}(X)$ and finite variance $\mathrm{V}(X) = \sigma^2$. Then, for any positive number $\epsilon > 0$ we have

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

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

Law of Large Numbers

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

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

or equivalently,

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

There are exercises for **LLN** on page 320.

CHAPTER 9: Central Limit Theorem

Central Limit Theorem for Bernoulli Trials

The second fundamental theorem of probability is the *Central Limit Theorem*. This theorem says that if Sn is the sum of n mutually independent random variables, then the distribution function of Sn is well-approximated by a certain type of continuous function known as a normal density function, which is given by the formula

$$f_{\mu,\sigma} = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}.$$

In this section, we will deal only with the case that $\mu = 0$ and $\sigma = 1$. We will call this particular normal density function the standard normal density, and we will denote it by $\phi(x)$:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Definition 9.1 The *standardized sum* of S_n is given by

$$S_n^* = \frac{S_n - np}{\sqrt{npq}},$$

where S_n^* has expected value 0 and variance 1.

Approximating Binomial Distributions We can approximate a binomial distribution function as follows

$$\begin{aligned} b(n,p,j) &\sim \frac{\phi(x)}{\sqrt{npq}} \\ &= \frac{1}{\sqrt{npq}} \phi\left(\frac{j-np}{\sqrt{npq}}\right). \end{aligned}$$

An example of a situation where this can be used is when there are 100 flips of a fair coin and we want to know the probability of seeing exactly 55 heads.