CHAPTER 5

Limit Theorems

5.1 Introduction

This chapter is principally concerned with the limiting behavior of the sum of independent random variables as the number of summands becomes large. The results presented here are both intrinsically interesting and useful in statistics, since many commonly computed statistical quantities, such as averages, can be represented as sums.

5.2 The Law of Large Numbers

It is commonly believed that if a fair coin is tossed many times and the proportion of heads is calculated, that proportion will be close to $\frac{1}{2}$. John Kerrich, a South African mathematician, tested this belief empirically while detained as a prisoner during World War II. He tossed a coin 10,000 times and observed 5067 heads. The law of large numbers is a mathematical formulation of this belief. The successive tosses of the coin are modeled as independent random trials. The random variable X_i takes on the value 0 or 1 according to whether the ith trial results in a tail or a head, and the proportion of heads in n trials is

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The law of large numbers states that $\overline{X}_n$ approaches $\frac{1}{2}$ in a sense that is specified by the following theorem.

THEOREM A *Law of Large Numbers*

Let $X_1, X_2, \ldots, X_i \ldots$ be a sequence of independent random variables with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Let $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Then, for any $\varepsilon > 0$,

$$P(|\overline{X}_n - \mu| > \varepsilon) \to 0 \qquad \text{as } n \to \infty$$

Proof

We first find $E(\overline{X}_n)$ and $\text{Var}(\overline{X}_n)$:

$$E(\overline{X}_n) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu$$

Since the X_i are independent,

$$\text{Var}(\overline{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{\sigma^2}{n}$$

The desired result now follows immediately from Chebyshev's inequality, which states that

$$P(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{\text{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0, \qquad \text{as } n \to \infty \qquad \blacksquare$$

In the case of a fair coin toss, the X_i are Bernoulli random variables with $p = 1/2$, $E(X_i) = 1/2$ and $\text{Var}(X_i) = 1/4$. If tossed 10,000 times

$$\text{Var}(\overline{X}_{10,000}) = 2.5 \times 10^{-5}$$

and the standard deviation of the average is the square root of the variance, 0.005. The proportion observed by Kerrich, 0.5067, is thus a little more than one standard deviation away from its expected value of 0.5, consistent with Chebyshev's inequality. (Recall from Section 4.2 that Chebyshev's inequality can be written in the form $P(|\overline{X}_n - \mu| > k\sigma) \leq 1/k^2$.)

If a sequence of random variables, $\{Z_n\}$, is such that $P(|Z_n - \alpha| > \varepsilon)$ approaches zero as n approaches infinity, for any $\varepsilon > 0$ and where α is some scalar, then Z_n is said to **converge in probability** to α. There is another mode of convergence, called *strong convergence* or *almost sure convergence,* which asserts more. Z_n is said to **converge almost surely to** α if for every $\varepsilon > 0$, $|Z_n - \alpha| > \varepsilon$ only a finite number of times with probability 1; that is, beyond some point in the sequence, the difference is always less than ε, but where that point is random. The version of the law of large numbers stated and proved earlier asserts that $\overline{X}_n$ converges to μ in probability. This version is usually called the *weak law of large numbers*. Under the same assumptions, a strong law of large numbers, which asserts that $\overline{X}_n$ converges almost surely to μ, can also be proved, but we will not do so.

We now consider some examples that illustrate the utility of the law of large numbers.

EXAMPLE A *Monte Carlo Integration*

Suppose that we wish to calculate

$$I(f) = \int_0^1 f(x)\, dx$$

where the integration cannot be done by elementary means or evaluated using tables of integrals. The most common approach is to use a numerical method in which the integral is approximated by a sum; various schemes and computer packages exist for doing this. Another method, called the **Monte Carlo method,** works in the following way. Generate independent uniform random variables on [0, 1]—that is, $X_1, X_2, \ldots, X_n$—and compute

$$\hat{I}(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$$

By the law of large numbers, this should be close to $E[f(X)]$, which is simply

$$E[f(X)] = \int_0^1 f(x)\, dx = I(f)$$

This simple scheme can be easily modified in order to change the range of integration and in other ways. Compared to the standard numerical methods, it is not especially efficient in one dimension, but becomes increasingly efficient as the dimensionality of the integral grows.

As a concrete example, let us consider the evaluation of

$$I(f) = \frac{1}{\sqrt{2\pi}}\int_0^1 e^{-x^2/2}\, dx$$

The integral is that of the standard normal density, which cannot be evaluated in closed form. From the table of the normal distribution (Table 2 in Appendix B), an accurate numerical approximation is $I(f) = .3413$. If 1000 points, $X_1, \ldots, X_{1000}$, uniformly distributed over the interval $0 \leq x \leq 1$, are generated using a pseudorandom number generator, the integral is then approximated by

$$\hat{I}(f) = \frac{1}{1000}\left(\frac{1}{\sqrt{2\pi}}\right)\sum_{i=1}^{1000} e^{-X_i^2/2}$$

which produced for one realization of the X_i the value .3417. ■

EXAMPLE B *Repeated Measurements*

Suppose that repeated independent unbiased measurements, $X_1, \ldots, X_n$, of a quantity are made. If n is large, the law of large numbers says that $\overline{X}$ will be close to the true value, μ, of the quantity, but how close $\overline{X}$ is depends not only on n but on the variance of the measurement error, σ^2, as can be seen in the proof of Theorem A.