Chapter 5: Limit Theorems

November 2, 2009

1 The Law of Large Numbers

Definition 1.1 *A sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a random variable X if, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

or equivalently,

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1.$$

The $X_1, X_2, \ldots$ in Definition 1.1 (and the other definitions in this section) are typically not independent and identically distributed random variables, as in a random sample. The distribution of X_n changes as the subscript changes, and the convergence concepts discussed in this section describes different ways in which the distribution of X_n converges to some limiting distribution as the subscript becomes large.

Theorem 1.1 *(Law of large numbers) Let $X_1, X_2, \ldots$ be iid random variable with $EX_i = \mu$ and $VarX_i = \sigma^2$. Define $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, for every $\epsilon > 0$,*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1;$$

that is, $\bar{X}_n$ converges in probability to μ.

PROOF: We have, for every $\epsilon > 0$,

$$P(|\bar{X}_n - \mu| \geq \epsilon) = P((\bar{X}_n - \mu)^2 \geq \epsilon)$$
$$\leq \frac{E(\bar{X}_n - \mu)^2}{\epsilon^2} = \frac{\mathrm{Var}\bar{X}}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Hence, $P(|\bar{X}_n - \mu| < \epsilon) = 1 - P(|\bar{X}_n - \mu| \geq \epsilon) = 1 - \frac{\sigma^2}{n\epsilon^2} \to 1$, as $n \to \infty$. □

The law of large numbers (also known as the weak law of large numbers or WLLN) quite elegantly states that under general conditions, the sample mean approaches the population mean as $n \to \infty$.

A type of convergence that is stronger than convergence in probability is almost sure convergence. This type of convergence is similar to pointwise convergence of a sequence of functions, except that the convergence need not occur on a set with probability 0 (hence the "almost" sure).

EX A (Consistency of S^2) Suppose we have a sequence $X_1, X_2, \ldots$ of iid random variables with $EX_i = \mu$ and $\text{Var}X_i = \sigma^2 < \infty$. If we define

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2,$$

using Chebychev's Inequality, we have

$$P(|S^2 - \sigma^2| \geq \epsilon) \leq \frac{E(S_n^2 - \sigma^2)^2}{\epsilon^2} = \frac{\text{Var}S_n^2}{\epsilon^2},$$

and thus, a sufficient condition that S_n^2 converges in probability to σ^2 is that $\text{Var}S_n^2 \to 0$ as $n \to \infty$.

EX B (Monte Carlo Integration) Suppose that we wish to calculate

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

where the integration can not be done by elementary means. The Monte Carlo method works in the following way. Generate independent uniform random variables on $[0, 1]$—that is, $X_1, \ldots, X_n$—and compute

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

By the law of large numbers, this should be close to $E[f(x)] = \int_0^1 f(x)dx$.

2 Convergence in Distribution and the Central Limit Theorem

Definition 2.1 *A sequence of random variables, $X_1, X_2, \ldots$, converges in distribution to a random variable X if*

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous.

EX A (Maximum of uniforms) If $X_1, X_2, \ldots$ are iid uniform(0,1) and $X_{(n)} = \max_{1 \le i \le n} X_i$, let us examine if $X_{(n)}$ converges in distribution.

As $n \to \infty$, we have for any $\epsilon > 0$,

$$P(|X_n - 1| \ge \epsilon) = P(X_{(n)} \le 1 - \epsilon)$$
$$= P(X_i \le 1 - \epsilon, i = 1, \ldots, n) = (1 - \epsilon)^n,$$

which goes to 0. However, if we take $\epsilon = t/n$, we then have

$$P(X_{(n)} \le 1 - t/n) = (1 - t/n)^n \to e^{-t},$$

which, upon rearranging, yields

$$P(n(1 - X_{(n)}) \le t) \to 1 - e^{-t};$$

that is, the random variable $n(1-X_{(n)})$ converges in distribution to an exponential(1) random variable.

Moment-generating function are often useful for establishing the convergence of distribution functions. We know from chapter 4 that a distribution function F_n is uniquely determined by its mgf, M_n. The following theorem states that this unique determination holds for limit as well.

Theorem 2.1 *(Continuity Theorem) Let F_n be a sequence of cumulative distribution functions with the corresponding moment-generating function M_n. Let F be a cumulative distribution function with the moment-generating function M. If $M_n(t) \to M(t)$ for all t in an open interval containing zero, then $F_n(x) \to F(x)$ at all continuity points of F.*

EX B We will show that the Poisson distribution can be approximated by the normal distribution for large values of λ.

Let $\lambda_1, \lambda_2, \ldots$ be an increasing sequence with $\lambda_n \to \infty$, and let $\{X_n\}$ be a sequence of Poisson random variables with the corresponding parameters. We know that $E(X_n) = \mathrm{Var}(X_n) = \lambda_n$. Let

$$Z_n = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}},$$

we then have $E(Z_n) = 0$ and $\mathrm{Var}(Z_n) = 1$, and we will show that the mgf of Z_n converges to the mgf of standard normal distribution.

The mgf of X_n is $\psi_{X_n}(t) = e^{\lambda_n(e^t-1)}$. The mgf of Z_n is then

$$\begin{aligned}\psi_{Z_n}(t) &= e^{-t\sqrt{\lambda_n}}\psi_{X_n}(\frac{t}{\sqrt{\lambda_n}}) \\ &= e^{-t\sqrt{\lambda_n}}e^{\lambda_n(e^{t/\sqrt{\lambda_n}}-1)} \\ &= e^{\frac{t^2}{2}+\frac{t^3}{3!\sqrt{\lambda_n}}+\cdots}.\end{aligned}$$

Thus, $\lim_{n\to\infty}\psi_{Z_n}(t) = e^{t^2/2}$, which concludes the proof.

EX C A certain type of particle is emitted at a rate of 900 per hour. What is the probability that more than 950 particles will be emitted in a given hour if the counts form a Poisson process?

Let X be a Poisson random variable with mean 900. By Example B, we have

$$P(X > 950) = P(\frac{X-900}{\sqrt{900}} > \frac{950-900}{\sqrt{900}}) \approx 1 - \Phi(5/3) = 0.04779.$$

For comparison, the exact probability is 0.04712.

Theorem 2.2 *(Central limit theorem) Let $X_1, X_2, \ldots$ be a sequence of iid random variables whose mgfs exist in a neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive h). Let $EX_i = 0$ and $VarX_i = \sigma^2$. Define $S_n = \sum_{i=1}^n X_i$. Then, for any x, $-\infty < x < \infty$,*

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \le x\right) = \Phi(x).$$

PROOF: Let $Z_n = S_n/(\sigma\sqrt{n})$. We will show that the mgf of Z_n tends to the mgf of the standard normal distribution. Since S_n is a sum of independent normal random variables,

$$\psi_{S_n}(t) = [\psi(t)]^n$$

and

$$\psi_{Z_n}(t) = \left[\psi\left(\frac{t}{\sigma\sqrt{n}}\right)\right].$$

$\psi(s)$ has a Taylor expansion about zero:

$$\psi(s) = \psi(0) + s\psi'(0) + \frac{1}{2}s^2\psi''(0) + \epsilon_s,$$

where $\epsilon_s/s^2 \to 0$ as $s \to 0$. Since $\psi'(0) = 0$ and $\psi''(0) = \sigma^2$,

$$\psi\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{1}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \epsilon_n,$$

where $\epsilon_n/(t^2/(n\sigma^2)) \to 0$ as $n \to \infty$. We thus have

$$\psi_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + \epsilon_n\right)^n.$$

It can be shown that if $a_n \to a$, then

$$\lim_{n\to\infty}(1+\frac{a_n}{n})^n = e^a.$$

From this result, it follows that

$$\psi_{Z_n}(t) \to e^{t^2/2}, \quad \text{as } n \to \infty,$$

which concludes the proof. □

The above theorem is one of the simplest versions of the central limit theorem; there are many central limit theorems of various degrees of abdtraction and generality. We have proved the above theorem under the assumption that the moment-generating function exist, which is a rather strong assumption. The next theorem only requires that the first and second moments exist.

Theorem 2.3 *(Stronger form of the central limit theorem) Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $EX_i = \mu$ and $0 < VarX_i = \sigma^2 < \infty$. Define $\bar{X}_n = (\frac{1}{n})\sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,*

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution.

EX D Because the uniform distribution on $[0,1]$ has mean $1/2$ and variance $1/12$, the sum of 12 uniform random variables, minus 6, has mean 0 and variance 1. The distribution of this sum is quite close to normal; in fact, before better algorithms were developed, it was commonly used in computers for generating normal random variables from uniform ones.

Figure 5.1 shows a histogram of 1000 such sums with a superimposed normal density function.

EX E The sum of n independent random variables with parameter $\lambda = 1$ follows a gamma distribution with $\lambda = 1$ and $\alpha = n$. The exponential density is quite skewed; therefore, a good approximation of a standardized gamma by a standardized normal would not be expected for small n. Figure **??** shows that the approximation to normal improves as n increases.

EX F Since a binomial random variable is the sum of independent bernoulli random variables, its is distribution can be approximated by a normal distribution. The approximation is best when the binomial distribution is symmetric—that is, when $p = 1/2$. A frequently used

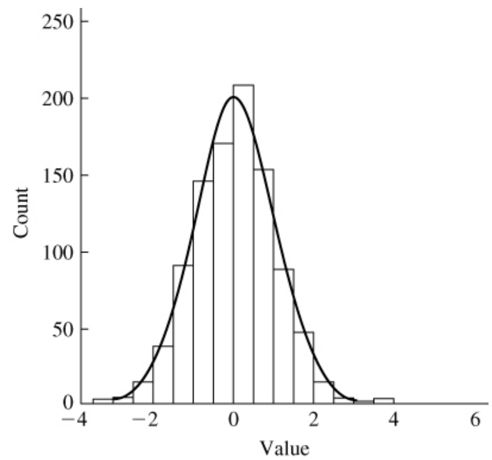

Figure 1: A histogram of 1000 values, each of which is the sum of 12 uniform $[-0.5, 0.5]$ random variables.

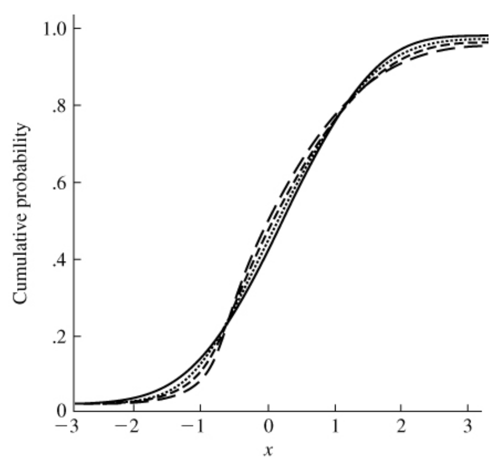

Figure 2: The standard normal cdf (solid line) and the cdf's of standardized gamma distributions with $\alpha = 5$ (log dashed), $\alpha = 10$ (short dashed), and $\alpha = 30$ (dots).

rule of thumb is that the approximation is reasonable when $np > 5$ and $n(1-p) > 5$. The approximation is especially useful for large values of n.

Suppose that a coin is tossed 100 times and lands heads up 60 times. Should we be supprised and doubt that the coin is fair?

$$\begin{aligned} P(X \geq 60) &= P\left(\frac{X-50}{\sqrt{25}} \geq \frac{100-50}{\sqrt{25}}\right) \\ &\approx 1 - \Phi(2) \\ &= 0.0228 \end{aligned}$$

The probability is rather small, so the fairness of the coin is called into question.