5.5 Convergence Concepts

This section treats the somewhat fanciful idea of allowing the sample size to approach infinity and investigates the behavior of certain sample quantities as this happens. We are mainly concerned with three types of convergence, and we treat them in varying amounts of detail. In particular, we want to look at the behavior of $\bar{X}_n$, the mean of n observations, as $n \to \infty$.

5.5.1 Convergence in Probability

Definition 5.5.1 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 5.5.1 (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 5.5.2 (Weak law of large numbers)

Let $X_1, X_2, \ldots$ be iid random variable with $EX_i = \mu$ and $\mathrm{Var}X_i = \sigma^2 < \infty$. 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 weak law of large numbers (WLLN) quite elegantly states that under general conditions, the sample mean approaches the population mean as $n \to \infty$.

Example (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 $n \to \infty$.

Theorem 5.5.4

Suppose that $X_1, X_2, \ldots$ converges in probability to a random variable X and that h is a continuous function. Then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.

Proof: If h is continuous, given $\epsilon > 0$ there exists $\delta > 0$ such that $|h(x_n) - h(x)| < \epsilon$ for $|x_n - x| < \delta$. Since $X_1, X_2, \ldots$ converges in probability to the random variable X, then

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

Thus,

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

□

Example (Consistency of S)

If S_n^2 is a consistent estimator of σ^2, then by Theorem 5.5.4, the sample standard deviation $S_n = \sqrt{S_n^2}$ is a consistent estimator of σ.

5.5.2 Almost sure convergence

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).

Example (Almost sure convergence)

Let the sample space S be the closed interval $[0, 1]$ with the uniform probability distribution. Define random variables $X_n(s) = s + s^n$ and $X(s) = s$. For every $s \in [0, 1)$, $s^n \to 0$ as $n \to$ and $X_n(s) \to s = X(s)$. However, $X_n(1) = 2$ for every n so $X_n(1)$ does not converge to $1 = X(1)$. But since the convergence occurs on the set $[0, 1)$ and $P([0, 1)) = 1$, X_n converges to X almost surely.

Example (Convergence in probability, not almost surely)

Let the sample space be $[0, 1]$ with the uniform probability distribution. Define the sequence $X_1, X_2, \ldots$ as follows:

$$X_1(s) = s + I_{[0,1]}(s), \quad X_2(s) = s + I_{[0,\frac{1}{2}]}(s), \quad X_3(s) = s + I_{[\frac{1}{2},1]}(s),$$
$$X_4(s) = s + I_{[0,\frac{1}{3}]}(s), \quad X_5(s) = s + I_{[\frac{1}{3},\frac{2}{3}]}(s), \quad X_6(s) = s + I_{[\frac{2}{3},1]}(s),$$
$$\ldots$$

Let $X(s) = s$. As $n \to \infty$, $P(|X_n - X| \geq \epsilon)$ is equal to the probability of an interval of s values whose length is going to 0. However, X_n does not converge to X almost surely. Indeed, there is no value of $s \in S$ for which $X_n(s) \to s = X(s)$. For every s, the value $X_n(s)$ alternates between the values s and $s + 1$ infinitely often. For example, if $s = \frac{3}{8}$, $X_1(s) = 11/8, X_2(s) = 11/8, X_3(s) = 3/8, X_4(s) = 3/8, X_5(s) = 11/8, X_6(s) = 3/8$, etc. No pointwise convergence occurs for this sequence.

Theorem 5.5.9 (Strong law of large numbers)

Let $X_1, X_2, \ldots$ be iid random variable with $EX_i = \mu$ and $\mathrm{Var}X_i = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Then, for every $\epsilon > 0$,

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

that is, $\bar{X}_n$ converges almost surely to μ.

For both the weak and strong law of large numbers we had the assumption of a finite variance. In fact, both the weak and strong laws hold without this assumption. The only moment condition needed is that $E|X_i| < \infty$.

5.5.3 Convergence in Distribution

Definition 5.5.10

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.

Example (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 $\epsilon > 0$,

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

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.

Note that although we talk of a sequence of random variables converging in distribution, it is really the cdfs that converge, not the random variables. In this very fundamental way convergence in distribution is quite different from convergence in probability or convergence almost surely.

Theorem 5.5.12

If the sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a random variable X, the sequence also converges in distribution to X.

Theorem 5.5.13

The sequence of random variables, $X_1, X_2, \ldots$, converges in probability to a constant μ if and only if the sequence also converges in distribution to μ. That is, the statement

$$P(|X_n - \mu| > \epsilon) \to 0 \quad \text{for every } \epsilon > 0$$

is equivalent to

$$P(X_n \leq x) \to \begin{cases} 0 & \text{if } x < \mu \\ 1 & \text{if } x > \mu. \end{cases}$$

Theorem 5.5.14 (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 $EX_i = \mu$ and $\mathrm{Var}X_i = \sigma^2 > 0$. (Both μ and σ^2 are finite since the mgf exists.) 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.

Theorem 5.5.15 (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 $0 < \mathrm{Var}X_i = \sigma^2 < \infty$. 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.

The proof is almost identical to that of Theorem 5.5.14, except that characteristic functions are used instead of mgfs.

Example (Normal approximation to the negative binomial)

Suppose $X_1, \ldots, X_n$ are a random sample from a negative binomial(r, p) distribution. Recall that

$$EX = \frac{r(1-p)}{p}, \quad \text{Var}X = \frac{r(1-p)}{p^2}$$

and the central limit theorem tells us that

$$\frac{\sqrt{n}(\bar{X} - r(1-p)/p)}{\sqrt{r(1-p)/p^2}}$$

is approximately $N(0, 1)$. The approximate probability calculation are much easier than the exact calculations. For example, if $r = 10$, $p = \frac{1}{2}$, and $n = 30$, an exact calculation would be

$$\begin{aligned} P(\bar{X} \leq 11) &= P(\sum_{i=1}^{30} X_i \leq 330) \\ &= \sum_{x=0}^{330} \binom{300 + x - 1}{x} (\frac{1}{2})^{300+x} = 0.8916 \end{aligned}$$

Note $\sum X$ is negative binomial(nr, p). The CLT gives us the approximation

$$P(\bar{X} \leq 11) = P(\frac{\sqrt{30}(\bar{X} - 10)}{\sqrt{20}} \leq \frac{\sqrt{30}(11 - 10)}{\sqrt{20}}) \approx P(Z \leq 1.2247) = .8888.$$

Theorem 5.5.17 (Slutsky's theorem)

If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then

(a) $Y_n X_n \to aX$ in distribution.

(b) $X_n + Y_n \to X + a$ in distribution.

Example (Normal approximation with estimated variance)

Suppose that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to N(0, 1),$$

but the value σ is unknown. We know $S_n \to \sigma$ in probability. By Exercise 5.32, $\sigma/S_n \to 1$ in probability. Hence, Slutsky's theorem tells us

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sigma}{S_n}\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to N(0, 1).$$