11. Uniform convergence

Lehmann §2.6

In the definition of convergence in distribution, we saw **pointwise convergence** of distribution functions: If $F(x)$ is continuous, then $F_n \xrightarrow{\mathcal{L}} F$ means that for each x, $F_n(x) \to F(x)$. In other words, for every x and $\epsilon > 0$, there exists N such that

$$|F_n(x) - F(x)| < \epsilon \text{ for all } n > N. \tag{13}$$

A very important fact about the above wording is the stipulation that N depends on x. In other words, for a given ϵ, a value of N that makes statement (13) true for some x might not work for some other x.

The idea of **uniform convergence**, on the other hand, is that we can choose N without regard to the value of x. Thus, for a given ϵ, we can select N so that (13) is true for all x.

Definition 11.1 $g_n(x)$ converges uniformly to $g(x)$ if for every $\epsilon > 0$, there exists N such that

$$|g_n(x) - g(x)| < \epsilon \text{ for all } n > N \text{ and for all } x.$$

An equivalent definition of uniform convergence is as follows:

Theorem 11.1 $g_n(x)$ converges uniformly to $g(x)$ if and only if $\sup_x |g_n(x) - g(x)| \to 0$.

Unlike uniform convergence, pointwise convergence merely asserts that $|g_n(x) - g(x)| \to 0$ for all x. It is tempting to believe that $\sup_x |g_n(x) - g(x)| \to 0$ and $|g_n(x) - g(x)| \to 0$ for all x are equivalent statements, but that would be a mistake.

Intuitively, $g_n(x) \to g(x)$ uniformly if it is possible to draw an ϵ-band around the graph of $g(x)$ that contains all of the graphs of $g_n(x)$ for large enough n.

Example 11.1 It is easy to demonstrate that uniform convergence is not the same thing as pointwise convergence by exhibiting examples in which pointwise convergence holds but uniform convergence does not.

- If $g_n(x) = x(1 + 1/n)$ and $g(x) = x$, then obviously $g_n(x) \to g(x)$ for all x (i.e., pointwise convergence holds). However, since $\sup_x |g_n(x) - g(x)| = \infty$ for all n, uniform convergence does not hold.
- If $g_n(x) = x^n$ for all $x \in (0, 1)$, then $g_n(x) \to 0$ for all fixed x as $n \to \infty$, but $\sup_{x \in (0,1)} |g_n(x)| = 1$.

Note that in both of these examples that for small $\epsilon > 0$, an ϵ-band around $g(x) = x$ in the first example and $g(x) = 0$ in the second example fails to capture the graphs of **any** $g_n(x)$.

Thus, it is clear that pointwise convergence does not in general imply uniform convergence. However, the following theorem gives a special case in which it does.

Theorem 11.2 If $F_n(x)$ and $F(x)$ are cdf's and $F(x)$ is continuous, then pointwise convergence of F_n to F implies uniform convergence of F_n to F.

Problems

Problem 11.1 This problem deals with the question of when $Y_n \xrightarrow{\mathcal{L}} Y$ implies E $Y_n \to$ E Y. As you know, this is not in general the case. As stated in the following definition and theorem (which you are not asked to prove), a sufficient condition for E $Y_n \to$ E Y is the **uniform integrability** of the Y_n.

Definition 11.2 The random variables $Y_1, Y_2, \ldots$ are said to be uniformly integrable if

$$\sup_n E\left(|Y_n|\, I\{|Y_n| \geq \alpha\}\right) \to 0 \text{ as } \alpha \to \infty.$$

Theorem 11.3 If $Y_n \xrightarrow{\mathcal{L}} Y$ and the Y_n are uniformly integrable, then E $Y_n \to$ E Y.

Prove that if there exists $\epsilon > 0$ such that $\sup_n \mathrm{E}\,|Y_n|^{1+\epsilon} < \infty$, then the Y_n are uniformly integrable.

Problem 11.2 Prove that if there exists a random variable Z such that E $|Z| = \mu < \infty$ and $P(|Y_n| \geq t) \leq P(|Z| \geq t)$ for all n and for all $t > 0$, then the Y_n are uniformly integrable. This result gives one version of the **dominated convergence theorem**:

Theorem 11.4 If $Y_n \xrightarrow{\mathcal{L}} Y$ and $|Y_n| \leq Z$ for all n for an integrable random variable Z, then E $Y_n \to$ E Y. (A random variable Z is defined to be integrable if E $|Z| < \infty$.)

You may use the fact (without proof) that for a nonnegative X,

$$\mathrm{E}\,(X) = \int_0^\infty P(X \geq t)\, dt.$$

Hints: Consider the random variables $|Y_n| I\{|Y_n| \geq t\}$ and $|Z| I\{|Z| \geq t\}$. In addition, use the fact that

$$\mathrm{E}\,|Z| = \sum_{i=1}^{\infty} \mathrm{E}\,\left(|Z| I\{i-1 \leq |Z| < i\}\right)$$

to argue that E $(|Z| I\{|Z| < \alpha\}) \to$ E $|Z|$ as $\alpha \to \infty$.

Problem 11.3 Let $X_1, X_2, \ldots$ be iid Poisson random variables with mean $\lambda = 1$. Define $Y_n = \sqrt{n}(\overline{X}_n - 1)$.

(a) Find $\mathrm{E}\,(Y_n^+)$, where $Y_n^+ = Y_n I\{Y_n > 0\}$.

(b) Find, with proof, the limit of $\mathrm{E}\,(Y_n^+)$ and prove Stirling's formula

$$n! \quad \sim \quad \sqrt{2\pi}\, n^{n+1/2} e^{-n}.$$

Hint: Use the result of Problem 11.1.

Problem 11.4 **(a)** Prove that as $x \to \infty$,

$$1 - \Phi(x) \quad \sim \quad \frac{1}{\sqrt{2\pi}} \frac{e^{-x^2/2}}{x},$$

where $\Phi(x)$ denotes the cdf of the standard normal distribution. (The asymptotic equivalence $\sim$ has the obvious definition for a continuous variable $x \to \infty$: $f(x) \sim g(x)$ if $f(x)/g(x) \to 1$.)

(b) Create a table in which you give $1 - \Phi(x)$ exactly and using the approximation based on part (a) for $x \in \{1, 2, 3, 5, 10, 15, 20\}$. Comment on the quality of the approximation.

Hints: For part (a), try writing the left hand side as an integral and then use integration by parts to express it as the right hand side plus another integral. For part (b), you'll need to use the fact that $1 - \Phi(x) = \Phi(-x)$ or the software will simply report 0 for large x.

Problem 11.5 Do Problems 4.6 and 4.7 on p. 125. Of course, $\theta = 0$ in each case. In the graphs for 4.7(ii), use 1000 replications for each n and plot the four cdfs (three empirical and one continuous) on the same graph (and label them) if possible. Use $a = 1$ for simulating from 4.6(ii).

Hint: Study Example 2.4.9 on p. 80.

12. Central Limit Theorem, Part I: iid case

Lehmann §2.4; Ferguson §5

Theorem 12.1 *Central limit theorem, iid case.* If $X_1, X_2, \ldots$ are iid with $\mathrm{E}\,(X_i) = \xi$ and $\mathrm{Var}\,(X_i) = \sigma^2 < \infty$, then

$$\frac{\sqrt{n}(\overline{X}_n - \xi)}{\sigma} \xrightarrow{\mathcal{L}} N(0, 1).$$

There are many examples in which this theorem is useful.

Example 12.1 *Distribution of T statistics:* Suppose $X_1, X_2, \ldots$ are iid with $\mathrm{E}\,(X_i) = \xi$ and $\mathrm{Var}\,(X_i) = \sigma^2 < \infty$. Define

$$s_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2, \tag{14}$$

and let

$$T_n = \frac{\sqrt{n}(\overline{X}_n - \xi)}{s_n}.$$

Letting

$$A_n = \frac{\sqrt{n}(\overline{X}_n - \xi)}{\sigma}$$

and $B_n = \sigma/s_n$, clearly $t_n = A_n B_n$. Therefore, since $A_n \xrightarrow{\mathcal{L}} N(0, 1)$ by the central limit theorem and $B_n \xrightarrow{P} 1$ by the weak law of large numbers, Slutsky's theorem implies that $t_n \xrightarrow{\mathcal{L}} N(0, 1)$. In other words, T statistics are asymptotically normal under the null hypothesis.

Example 12.2 *Distribution of sample variance:* Suppose that $X_1, X_2, \ldots$ are iid with $\mathrm{E}\,(X_i) = \xi$, $\mathrm{Var}\,(X_i) = \sigma^2$, and $\mathrm{Var}\,\{(X_i - \xi)^2\} = \tau^2 < \infty$. Define s_n^2 as in equation (14). Letting $Y_i = X_i - \xi$, it is not hard to verify that

$$s_n^2 = \frac{1}{n} \sum_{i=1}^{n} Y_i^2 - \overline{Y}_n^2.$$

By the CLT, we know that

$$\sqrt{n} \left(\frac{1}{n} Y_i^2 - \sigma^2 \right) \xrightarrow{\mathcal{L}} N(0, \tau^2)$$

because $\mathrm{E}\,(Y_i^2) = \mathrm{Var}\,(X_i) = \sigma^2$. Furthermore, the fact that $\sqrt{n}\overline{Y}_n \xrightarrow{\mathcal{L}} N(0, \sigma^2)$ means that $\sqrt{n}\overline{Y}_n^2 \xrightarrow{P} 0$. Therefore, since

$$\sqrt{n}(s_n^2 - \sigma^2) = \sqrt{n} \left(\frac{1}{n} Y_i^2 - \sigma^2 \right) + \sqrt{n}\overline{Y}_n^2,$$

Slutsky's theorem implies that $\sqrt{n}(s_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} N(0, \tau^2)$.

Notice that s_n^2 is defined in equation (14) to have the denominator n, whereas the sample variance is often defined with $n-1$ as the denominator instead. However, it is possible through an additional use of Slutsky's theorem to prove that both versions of s_n^2 have the same limiting behavior.

Example 12.3 The assumptions made in the CLT are crucial. For example, if $X_1, X_2, \ldots$ are iid standard Cauchy random variables, then the stipuation that $\text{Var}(X_i) < \infty$ is violated, and in fact $\overline{X}_n$ is itself standard Cauchy.

It is possible to strengthen the CLT in a couple of directions. We consider one of them, the Berry-Esseen theorem, here. Another one, called Edgeworth expansions, will not be covered, though it is an interesting topic in its own right and you will find it in both Lehmann's book and Ferguson's book if you are interested.

As Ferguson puts it in Section 5, the convergence in the CLT is not uniform in the underlying distribution in the sense that for any n, there are distributions satisfying the hypotheses of the CLT but for which the distribution of $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ approximates the standard normal distribution arbitrarily poorly. However, consider the following theorem:

Theorem 12.2 *Berry-Esseen theorem:* If $X_1, \ldots, X_n$ are iid with mean μ, variance $\sigma^2 > 0$, and absolute third moment $\rho = \text{E}\,|X - \mu|^3 < \infty$, then

$$|G_n(x) - \Phi(x)| < \frac{.7975\rho}{\sigma^3\sqrt{n}} \quad \text{for all } x \text{ and } n, \tag{15}$$

where $G_n(x)$ is the cdf of $\sqrt{n}(\overline{X}_n - \mu)/\sigma$ and $\Phi(x)$ is the standard normal cdf.

Note that the Berry-Esseen theorem is not a limit theorem; it is true **for all** n. However, as $n \to \infty$, the right hand side goes to zero, which means that the CLT is implied for the case of a finite absolute third moment. Interestingly, it is not known whether the constant .7975 is the best possible constant—it is only known that the constant must be greater than .4097.

We won't use the Berry-Esseen theorem extensively in this class, but we now give an example of the sort of result it may be used to prove.

Example 12.4 Suppose that $X_n \sim \text{binomial}(n, p_n)$. We would like to prove a central-limit-theorem-like result for these X_n. View X_n as the sum $\sum_{i=1}^n Y_{ni}$ of iid Bernoulli(p_n) random variables. Unfortunately, the CLT does not apply because the Y_{ni} are not iid. However, note that $\text{E}\,|Y_{ni} - p_n|^3$ must be less than or equal to one. In this case, $\sigma = \sqrt{p_n(1-p_n)}$, so the Berry-Esseen theorem implies that

$$\frac{\sqrt{n}(X_n/n - p_n)}{\sqrt{p_n(1-p_n)}} = \frac{X_n - np_n}{\sqrt{np_n(1-p_n)}} \xrightarrow{\mathcal{L}} N(0,1)$$

as long as $p_n(1-p_n)$ does not converge to zero. (The convergence is even possible if $p_n(1-p_n) \to 0$; see Example 2.4.8 on page 79 of Lehmann's book.)

Example 12.5 *Distribution of sample median:* Suppose $X_1, \ldots, X_n$ are iid such that $P(X_i \le x) = F(x - \theta)$ for some cdf $F(x)$ with $F(0) = 1/2$ and $F'(0)$ exists and is positive. Let $f(0) = F'(0)$. Note that θ is the population median. Suppose n is odd, $n = 2m - 1$, and let $\tilde{X}_n$ denote the sample median. Then $\tilde{X}_n = X_{(m)}$, the mth order statistic.

We may compute the cdf of $\sqrt{n}(X_{(m)} - \theta)$ directly:

$$P\left\{\sqrt{n}(X_{(m)} - \theta) \le x\right\} = P\left\{X_{(m)} \le \frac{x}{\sqrt{n}} + \theta\right\} = P(Y_n \le m - 1),$$

where $Y_n \sim \text{binomial}\{n, 1 - F(x/\sqrt{n})\}$. Let $p_n = 1 - F(x/\sqrt{n})$. Then

$$P(Y_n \le m-1) = P\left\{\frac{Y - np_n}{\sqrt{np_n(1-p_n)}} \le \frac{\sqrt{n}(1 - 2p_n) - (1/\sqrt{n})}{2\sqrt{p_n(1-p_n)}}\right\}.$$

Because $p_n \to 1/2$ in this case, Example 12.4 implies that

$$P(Y_n \leq m-1) - \Phi\left\{\frac{\sqrt{n}(1-2p_n) - (1/\sqrt{n})}{2\sqrt{p_n(1-p_n)}}\right\} \to 0. \tag{16}$$

The argument of Φ in expression (16) is clearly asymptotically equivalent to $\sqrt{n}(1-2p_n)$, which equals

$$2x\frac{F(x/\sqrt{n}) - F(0)}{x/\sqrt{n}}. \tag{17}$$

Since expression (17) converges to $2xf(0)$ by the definition of derivative, we conclude that $P(Y_n \leq m-1) \to \Phi\{2xf(0)\}$. Therefore, we finally obtain the asymptotic distribution of the sample median:

$$\sqrt{n}(\tilde{X}_n - \theta) \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{4f^2(0)}\right).$$

Problems

Problem 12.1 Using the result of Example 12.2, prove that $\sqrt{n}(v_n - \sigma^2) \xrightarrow{\mathcal{L}} N(0, \tau^2)$, where

$$v_n = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$$

is the unbiased version of the sample variance and everything else is defined as in Example 12.2.

13. Taylor's theorem and the delta method

Lehmann §2.5; Ferguson §7

We begin with Taylor's theorem:

Theorem 13.1 *Taylor's theorem:*

- If $f(x)$ has r derivatives at a, then as $\Delta \to 0$,

$$\begin{aligned} f(a+\Delta) &= f(a) + \Delta f'(a) + \cdots + \frac{\Delta^r}{r!} f^{(r)}(a) + o(\Delta^r) \\ &= f(a) + \Delta f'(a) + \cdots + \frac{\Delta^r}{r!} \left\{ f^{(r)}(a) + o(1) \right\}. \end{aligned}$$

- If $f(x)$ has $r+1$ derivatives in $[a, a+\Delta]$, then there exists $\xi \in [a, a+\Delta]$ such that

$$f(a+\Delta) = f(a) + \Delta f'(a) + \cdots + \frac{\Delta^r}{r!} f^{(r)}(a) + \frac{\Delta^{r+1}}{(r+1)!} f^{(r+1)}(\xi).$$

Based on Taylor's theorem, we have the delta method:

Theorem 13.2 *Delta method:* If $g'(a)$ exists and $n^b(X_n - a) \xrightarrow{\mathcal{L}} X$ for $b > 0$, then $n^b \{g(X_n) - g(a)\} \xrightarrow{\mathcal{L}} g'(a)X$.

The proof of the delta method uses Taylor's theorem: Since $X_n - a \xrightarrow{P} 0$,

$$n^b \{g(X_n) - g(a)\} = n^b (X_n - a) \{g'(a) + o_P(1)\},$$

and thus Slutsky's theorem together with the fact that $n^b(X_n - a) \xrightarrow{\mathcal{L}} X$ proves the result.

Corollary 13.1 As a special case, we have Theorem 2.5.2 on page 86 of Lehmann, which states that if $g'(\xi)$ exists and $\sqrt{n}(\overline{X}_n - \xi) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$, then

$$\sqrt{n} \left\{ g(\overline{X}_n) - g(\xi) \right\} \xrightarrow{\mathcal{L}} N \left\{ 0, \sigma^2 g'(\xi)^2 \right\}.$$

Example 13.1 *Asymptotic distribution of $\overline{X}_n^2$* Suppose $X_1, X_2, \ldots$ are iid with mean ξ and finite variance σ^2. Then by the central limit theorem,

$$\sqrt{n}(\overline{X}_n - \xi) \xrightarrow{\mathcal{L}} N(0, \sigma^2).$$

Therefore, the delta method gives

$$\sqrt{n}(\overline{X}_n^2 - \xi^2) \xrightarrow{\mathcal{L}} N(0, 4\xi^2\sigma^2). \tag{18}$$

However, this is not necessarily the end of the story. If $\xi = 0$, then the normal limit in (18) is degenerate—that is, expression (18) merely states that $\sqrt{n}(\overline{X}_n^2)$ converges in probability to the constant 0. This is not what we mean by the asymptotic distribution! Thus, we must treat the case $\xi = 0$ separately, noting in that case that $\sqrt{n}\overline{X}_n \xrightarrow{\mathcal{L}} N(0, \sigma^2)$ by the central limit theorem, which implies that

$$n\overline{X}_n \xrightarrow{\mathcal{L}} \sigma^2 \chi^2_1.$$

Example 13.2 *Estimating binomial variance:* Suppose $X_n \sim \text{binomial}(n, p)$. Because X_n/n is the MLE for p, the MLE for $p(1-p)$ is $\delta_n = X_n(n - X_n)/n^2$. The central limit theorem tells us that $\sqrt{n}(X_n/n - p) \xrightarrow{\mathcal{L}} N(0, pq)$, so the delta method gives

$$\sqrt{n}\left\{\delta_n - p(1-p)\right\} \xrightarrow{\mathcal{L}} N\left\{0, p(1-p)(1-2p)^2\right\}.$$

Note that in the case $p = 1/2$, this does not give the asymptotic distribution of δ_n.

We have seen in a couple of cases that if $g'(a) = 0$, then the delta method gives something other than the asymptotic distribution we seek. However, if we carry the Taylor expansion out one more term in this case, we obtain the following theorem:

Theorem 13.3 If $g'(a) = 0$ and $g''(a)$ exists, then $n^b(X_n - a) \xrightarrow{\mathcal{L}} X$ implies that

$$n^{2b}\left\{g(X_n) - g(a)\right\} \xrightarrow{\mathcal{L}} \frac{1}{2} g''(a) X^2.$$

We can carry this idea even further if we wish and prove an analagous result for the case in which $g'(a) = 0, g''(a) = 0, \ldots, g^{(r-1)}(a) = 0$ and $g^{(r)}(a)$ exists.

We end this topic with a discussion of **variance stabilizing transformations**. Often, if $\mathrm{E}(X_i) = \xi$ is the parameter of interest, the central limit theorem gives

$$\sqrt{n}(\overline{X}_n - \xi) \xrightarrow{\mathcal{L}} N\{0, \sigma^2(\xi)\}.$$

In other words, the variance of the limiting distribution is a function of ξ. This is a problem if we wish to do inference for ξ, because ideally the limiting distribution should not depend on the unknown ξ. The delta method gives a possible solution: Since

$$\sqrt{n}\left\{g(\overline{X}_n) - g(\xi)\right\} \xrightarrow{\mathcal{L}} N\left\{0, \sigma^2(\xi) g'(\xi)^2\right\},$$

we may search for a transformation $g(x)$ such that $g'(\xi)\sigma(\xi)$ is a constant. Such a transformation is called a variance stabilizing transformation.

Problems

Problem 13.1 **(a)** Do problem 5.6 on page 127. For the sake of simplicity, assume $\sigma^2 = 1$. Note that the case $\theta = 0$ should be treated separately.

(b) Do problem 5.7 on page 127.

Problem 13.2 **(a)** Do Problem 5.4 on page 127. Note that this result states that $f(x) = \sin^{-1}(\sqrt{x})$ is a variance-stabilizing transformation in the binomial case.

(b) Do Problem 5.9 (i) and (ii) on page 127.

(c) Evaluate the two confidence intervals in part (b) numerically for all combinations of $n \in \{10, 100, 1000\}$ and $p \in \{.1, .5\}$ as follows: For 1000 realizations of $X \sim \text{bin}(n, p)$, construct both 95% confidence intervals and keep track of how many times (out of 1000) that the confidence intervals contain p. Report the observed proportion of successes for each (n, p) combination. Comment on the quality of the two methods of producing confidence intervals.

Problem 13.3 **(a)** Suppose that $X_1, X_2, \ldots$ are iid Normal $(0, \sigma^2)$ random variables. Using the result of Example 12.2, find a variance-stabilizing transformation for

$$s_n^2 \quad = \quad \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

(b) Give an approximate test at $\alpha = .05$ for $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_a : \sigma^2 \neq \sigma_0^2$ based on part (a).

(c) For $n = 25$, estimate the true level of the test in part (b) by simulating 5000 samples of size $n = 25$ from the null distribution with $\sigma_0^2 = 1$. Report the proportion of cases in which you reject the null hypothesis according to your test (ideally, this proportion will be about .05).