Lecture 19 – Introduction to large sample theory

Last Modified: November 3, 2026

Outline:

19.1 Why we need large sample theory 1
19.2 Modes of convergence 2
19.3 Operations which preserve convergence 5

19.1 Why we need large sample theory

So far, we have only studied statistics whose distributions we could work out (e.g. the distribution of an MLE or the distribution of a test statistic under the null hypothesis). What happens when we don't know this distribution? Large sample theory gives us another way forward: we study what happens as $n \to \infty$, and use the limiting behavior to approximate what happens for large n.

Example 1 (Unknown distribution of sample mean)**.** Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim} f$, where X_i is the change in blood pressure for patient i after taking a medication. Let $\mu = \mathbb{E}[X_1]$. We want to test

$$H_0 : \mu = 0 \qquad \text{versus} \qquad H_1 : \mu \neq 0.$$

A natural test rejects when $|\bar{X}_n| > c_\alpha$: a large positive or negative average change is evidence against H_0. To choose a cutoff giving a level-α test, we would like to solve

$$\mathbb{P}_{\mu=0}(|\bar{X}_n| > c_\alpha) = \alpha.$$

But if we do not know the density f, we do not know the distribution of $\bar{X}_n$, even under H_0. How can we evaluate this probability and choose the cutoff? Large sample theory will let us approximate the null distribution of a standardized sample mean and choose a cutoff with Type I error probability approximately α for large n.

Example 2 (No closed form for MLE)**.** Suppose $X_1, \dots, X_n$ are i.i.d. Cauchy random variables with

$$f_\theta(x) = \frac{1}{\pi\{1 + (x - \theta)^2\}}, \qquad x \in \mathbb{R}.$$

We would like to understand the bias and variance of the MLE $\hat{\theta}_{\text{MLE}}$. However, even computing the estimator is not straightforward. The log-likelihood is

$$\ell_n(\theta) = -n \log \pi - \sum_{i=1}^{n} \log\{1 + (X_i - \theta)^2\},$$

so the MLE must satisfy

$$\ell_n'(\hat{\theta}_{\text{MLE}}) = \sum_{i=1}^{n} \frac{2(X_i - \hat{\theta}_{\text{MLE}})}{1 + (X_i - \hat{\theta}_{\text{MLE}})^2} = 0.$$

There is no simple formula for $\hat{\theta}_{\text{MLE}}$ based on this equation, which makes computing its bias and variance seems challenging. Asymptotic theory provides a way to approximate these quantities for large n.

The large n case is called the *asymptotic* setting. We need to consider *sequences* of random variables, often denoted by

$$Y_1, Y_2, Y_3, \ldots \qquad \text{or} \qquad \{Y_n\}_{n=1}^{\infty} \qquad \text{or simply} \qquad \{Y_n\}.$$

For example, Y_n might represent the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ when the sample size is n. We first formalize what it means for random variables to converge.

19.2 Modes of convergence

Definition 1. A sequence of numbers $a_1, a_2, a_3, \ldots$ **converges** to a if the following holds:

$$\forall \epsilon > 0, \quad \exists N_\epsilon < \infty, \text{ such that } \quad \forall n \geq N_\epsilon, \quad |a_n - a| \leq \epsilon$$

Intuitively, the definition says that eventually, a_n is as close to a as you want it to be: if you set any error tolerance, then for large enough n, you are guaranteed to have a_n within that tolerance of a. Typically the smaller your error tolerance, the larger n you will need to reach it.

Convergence is denoted by

$$\lim_{n\to\infty} a_n = a, \qquad \text{or} \qquad a_n \xrightarrow{n\to\infty} a \qquad \text{or simply} \qquad a_n \to a$$

Example 3. Let's prove that $1/n \to 0$. First consider it intuitively. Set your error tolerance to be 0.01. Is $1/n$ eventually within 0.01 of 0? Yes, once $n > 100$. Is $1/n$ eventually within 0.001 of 0? Also yes, once $n > 1000$.

Now let's do it formally: pick any $\epsilon > 0$ and observe that for all $n \geq 1/\epsilon$, we have $|1/n - 0| = 1/n \leq \epsilon$. Hence $1/n$ converges to 0.

Remark 1. On exams or in homework, I will not ask you to formally prove convergence of a sequence of deterministic numbers.

Definition 2 (Convergence in probability)**.** Let $\{Y_n\}$ be a sequence of random variables. We say that Y_n **converges in probability** to a random variable Y if

$$\forall \delta > 0, \quad \lim_{n\to\infty} \mathbb{P}(|Y_n - Y| > \delta) = 0$$

Convergence in probability is denoted by $Y_n \xrightarrow{p} Y$. The definition essentially says that it becomes increasingly rare for Y_n to be some distance δ from Y. This is true even if δ is very small! So for example $\mathbb{P}(|Y_n - Y| > 10^{-3})$ still converges to zero.

Example 4. Suppose $Y_n \sim \mathsf{Bernoulli}\big(p/n\big)$ for $n = 1, 2, 3, \ldots$. Let's prove that $Y_n \xrightarrow{p} 0$. To do this, we need to take any $\delta > 0$ and show $\mathbb{P}(|Y_n - 0| > \delta)$ converges to zero. Let's consider two cases. If $\delta \geq 1$, then $\mathbb{P}(|Y_n| > \delta) = \mathbb{P}(Y_n < -\delta) + \mathbb{P}(Y_n > \delta) = 0$, since Y_n can't be negative and Y_n can't be greater than 1. Therefore $\mathbb{P}(|Y_n| > \delta)$ converges to zero. Next suppose $0 < \delta < 1$. Then

$$\mathbb{P}(|Y_n - 0| > \delta) = \mathbb{P}(Y_n > \delta) = \mathbb{P}(Y_n = 1) = \frac{p}{n} \to 0.$$

So in both cases, we have shown the necessary convergence to zero.

Definition 3 (Convergence in distribution)**.** Let $\{Y_n\}$ be a sequence of random variables, each with distribution function F_n, and let Y be a random variable with distribution function F. We say Y_n **converges in distribution** to Y if

$$\lim_{n\to\infty} F_n(y) = F(y)$$

for every y such that F is continuous at y. Convergence in distribution is denoted by $Y_n \xrightarrow{d} Y$.

Example 5. Let $Y_n \sim \mathsf{Unif}(0, 1/n)$. We will show that $Y_n \xrightarrow{d} 0$. The cdf of Y_n is

$$F_n(t) = \begin{cases} 0, & t \le 0, \\ nt, & 0 < t < 1/n, \\ 1, & t \ge 1/n. \end{cases}$$

The proposed limit is the constant random variable $Y = 0$, whose cdf is $F(t) = 0$ for $t < 0$ and $F(t) = 1$ for $t \ge 0$.

Fix $t < 0$. Then $F_n(t) = 0 = F(t)$ for every n.

Next, fix any $t > 0$. We keep this same value of t throughout and let n increase. We are studying the sequence of numbers

$$F_1(t),\ F_2(t),\ F_3(t),\ \ldots.$$

On the plot below, these are the heights at which the successive cdf curves meet the vertical line at our chosen t. We are not moving along a single curve: we stay at the same horizontal position and look at a different curve for each n.

When $n < 1/t$, we have $t < 1/n$, so the vertical line meets the rising part of the cdf and $F_n(t) = nt < 1$. When $n > 1/t$, we have $1/n < t$: the cdf has already reached 1 before the vertical line, so $F_n(t) = 1$. For example, at $t = 0.3$, the sequence is

$$F_1(0.3) = 0.3, \quad F_2(0.3) = 0.6, \quad F_3(0.3) = 0.9, \quad F_4(0.3) = 1, \quad F_5(0.3) = 1, \quad \ldots.$$

Here $1/t = 10/3$, so every integer $n \ge 4$ is past the threshold. More generally, once $t > 0$ is fixed, $1/t$ is just a fixed, finite number. When taking the limit as $n \to \infty$, we can always take n large enough that $n > 1/t$. All subsequent terms are then exactly 1; the finitely many earlier terms do not affect the limit. Thus $F_n(t) \to 1 = F(t)$ for every fixed $t > 0$, however small t is. Together with the case $t < 0$, this proves convergence at every continuity point of F, and hence $Y_n \xrightarrow{d} 0$.

At $t = 0$, however, $F_n(0) = 0$ for every n, whereas $F(0) = 1$. This is why the definition excludes discontinuity points of the limiting cdf: all the probability mass approaches zero from above, but $\mathbb{P}(Y_n \le 0)$ remains zero.

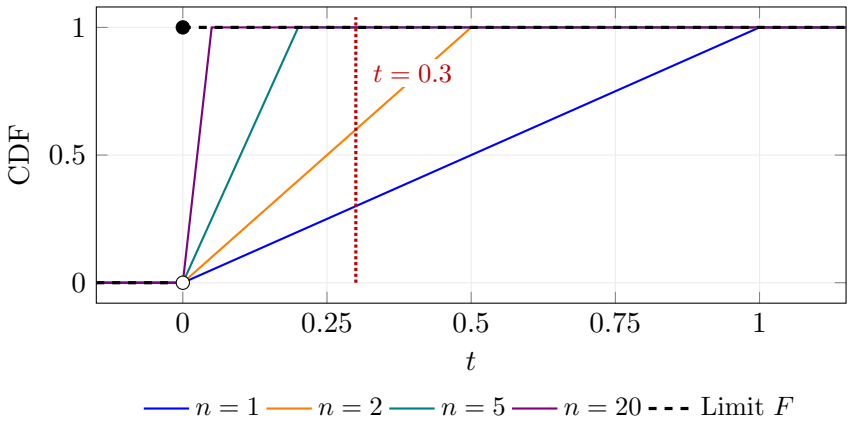


Figure 19.1: CDFs of $\mathsf{Unif}(0, 1/n)$ for increasing n. The vertical dotted line marks the fixed value $t = 0.3$. Its intersections with the plotted curves give $F_1(0.3) = 0.3$, $F_2(0.3) = 0.6$, and $F_5(0.3) = F_{20}(0.3) = 1$. The dashed curve is the limiting cdf. The open circle at $(0, 0)$ and filled circle at $(0, 1)$ indicate that $F(0) = 1$.

Example 6. Suppose $X_i \overset{\text{iid}}{\sim} \mathsf{Unif}(0, 1)$, and let $Y_n = X_{(n)} = \max\{X_1, \ldots, X_n\}$. Show that $Y_n \xrightarrow{d} 1$.
For $0 \le t \le 1$, the distribution function of Y_n is

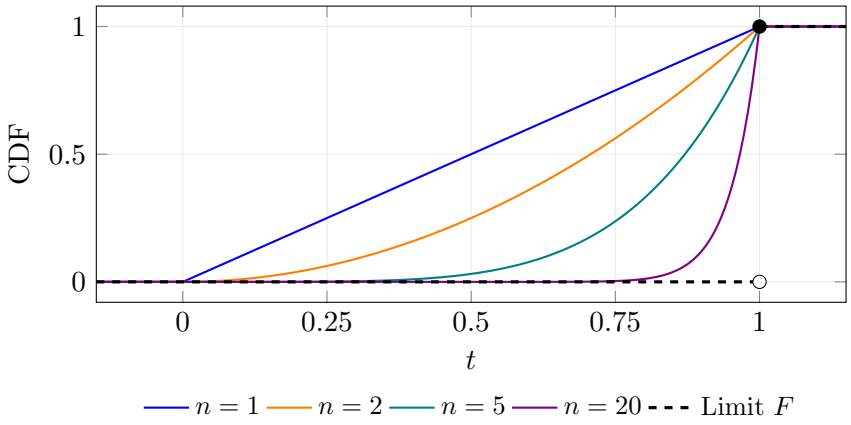
$$F_n(t) = \mathbb{P}(Y_n \le t) = \prod_{i=1}^{n} \mathbb{P}(X_i \le t) = t^n$$

Thus the sequence of distribution functions converges to a step function at 1,

$$\lim_{n\to\infty} F_n(t) = \begin{cases} 0, & t < 1, \\ 1, & t \ge 1. \end{cases}$$

This is exactly the distribution function of the constant "random variable" equal to 1. Hence

$$Y_n = X_{(n)} \xrightarrow{d} 1.$$



Here the limiting cdf jumps at $t = 1$. Unlike the preceding example, convergence also holds at the jump: $F_n(1) = 1 = F(1)$ for every n.

Theorem 1. Convergence in probability implies convergence in distribution, i.e.,

$$Y_n \xrightarrow{p} Y \quad \Longrightarrow \quad Y_n \xrightarrow{d} Y$$

but the reverse implication does not hold in general. In the special case where $Y = c$ is a constant, the reverse implication *does* hold, so we conclude that convergence in probability is *equivalent* to convergence in distribution.

Proof that $Y_n \xrightarrow{d} c$ *implies* $Y_n \xrightarrow{p} c$. Let F_n be the cdf of Y_n, and fix any $\delta > 0$. We need to show $\mathbb{P}(|Y_n - c| > \delta)$ converges to zero. But

$$\begin{aligned} \mathbb{P}(|Y_n - c| > \delta) &\le \mathbb{P}(Y_n \le c - \delta) + \mathbb{P}(Y_n > c + \delta) \\ &= F_n(c - \delta) + 1 - F_n(c + \delta) \to 0 + 1 - 1 = 0, \end{aligned} \tag{19.1}$$

since $c - \delta$ and $c + \delta$ are continuity points of the cdf of the constant c. □

Example 7 (Convergence need not preserve expectations)**.** $Y_n \xrightarrow{p} Y$ does not guarantee that $\mathbb{E}[Y_n] \to \mathbb{E}[Y]$ (and therefore $Y_n \xrightarrow{d} Y$, which is weaker than convergence in probability, also does not guarantee convergence of means). To see why consider the following example:

$$Y_n = \begin{cases} Y, & \text{if } X_n = 1, \\ n^2, & \text{if } X_n = 0, \end{cases} \tag{19.2}$$

where $X_n \sim \text{Bernoulli}(1 - 1/n)$, and X_n is independent of Y. Assume $\mathbb{E}[|Y|] < \infty$.

To prove convergence in probability, fix any $\delta > 0$. Since $Y_n = Y$ whenever $X_n = 1$, the event $\{|Y_n - Y| > \delta\}$ can occur only when $X_n = 0$. Hence

$$\mathbb{P}(|Y_n - Y| > \delta) \leq \mathbb{P}(X_n = 0) = \frac{1}{n} \to 0.$$

Thus $Y_n \xrightarrow{p} Y$.

On the other hand, using independence of X_n and Y,

$$\mathbb{E}[Y_n] = \left(1 - \frac{1}{n}\right)\mathbb{E}[Y] + \frac{1}{n}\, n^2 = \left(1 - \frac{1}{n}\right)\mathbb{E}[Y] + n \to +\infty.$$

Therefore $\mathbb{E}[Y_n]$ does not converge to the finite value $\mathbb{E}[Y]$. The exceptional value n^2 occurs with probability tending to zero, but its contribution to the expectation is n, which grows without bound.

19.3 Operations which preserve convergence

Theorem 2 (Continuous Mapping Theorem)**.** Let $\{Y_n\}$ be a sequence of random variables such that $Y_n \xrightarrow{p} Y$, and let g be a continuous function. Then

$$g(Y_n) \xrightarrow{p} g(Y).$$

Theorem 3 (Convergence of sums and products)**.** If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$ then $X_n + Y_n \xrightarrow{p} X + Y$ and $X_n Y_n \xrightarrow{p} XY$.

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ then $X_n + Y_n \xrightarrow{d} X + c$ and $X_n Y_n \xrightarrow{d} Xc$.

The second statement is known as Slutsky's Theorem.

Remark 2. In general, $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ does *not* imply $X_n + Y_n \xrightarrow{d} X + Y$ or $X_n Y_n \xrightarrow{d} XY$. In fact, a statement like this does not even make sense, without additional information about the joint distribution of X_n and Y_n, and the joint distribution of X and Y.

For example, take a sequence of X_n's such that $X_n \sim \mathcal{N}(0, 1)$, and let $Y_n = -X_n$. By symmetry of the standard normal, $Y_n \sim \mathcal{N}(0, 1)$ as well. Therefore, we can take any two random variables X, Y which both have the $\mathcal{N}(0, 1)$ distribution, and conclude that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$.

So does $X_n + Y_n = 0$ converge in distribution to $X + Y$? It depends on the relationship between X and Y! If we take $Y = -X$, then $X + Y = 0$, and 0 indeed converges to 0. If we take $Y = X$, then $X + Y = 2X \sim \mathcal{N}(0, 4)$, and 0 does not converge to $\mathcal{N}(0, 4)$.

It is common to write the limit of a convergence in distribution using the law itself, rather than a random variable. So in this example, we would write

$$X_n \xrightarrow{d} \mathcal{N}(0, 1), \qquad Y_n \xrightarrow{d} \mathcal{N}(0, 1).$$

This makes it clear that we don't have enough information to know what happens when we add the two $\mathcal{N}(0, 1)$'s.