STATS300B – Lecture 5

Julia Palacios
Scribed by Michael Howes

01/18/22

Contents

1 Convergence of random variables

1.1 Skorokhod's theorem

We ended last lecture with the statement and a proof of Skorokhod's theorem.

Theorem 1 (Skorokhod)**.** *Suppose that $X_n \xrightarrow{d} X_0$. Then there exist random variables X_n^* such that $X_n^* \overset{\text{dist}}{=} X_n$ and $X_n^* \xrightarrow{a.s.} X_0^*$.*

The idea behind the proof was to work with the inverse CDFs,

$$F_n^{-1}(t) = \inf\{x \in \mathbb{R} : F_n(x) \geq t\},$$

where $F_n(x) = \mathbb{P}(X_n \leq x)$. We defined $X_n^* = F_n^{-1}(\xi)$ where $\xi \sim U(0,1)$. We showed that if $X_n \xrightarrow{d} X_0$, then $F_n^{-1}(\xi) \xrightarrow{a.s.} F_0^{-1}(\xi)$.

1.2 Marginal convergence

Let X_n and X be random k-vectors. Then

1. $X_n \xrightarrow{p} X$ if and only if $X_{n,j} \xrightarrow{p} X_j$ for all $j = 1, \ldots, k$.
2. $X_n \xrightarrow{a.s.} X$ if and only if $X_{n,j} \xrightarrow{a.s.} X_j$ for all $j = 1, \ldots, k$.
3. $X_n \xrightarrow{L^p} X$ if and only if $X_{n,j} \xrightarrow{L^p} X_j$ for all $j = 1, \ldots, k$.
4. If $X_n \xrightarrow{d} X$, then $X_{n,j} \xrightarrow{d} X_j$ for all $j = 1, \ldots, k$. But (!) it is possible that $X_{n,j} \xrightarrow{d} X_j$ for all $j = 1, \ldots, k$ and $X_n \overset{d}{\not\to} X$. Indeed, X_n need not have any distribution limit (see homework).

So for convergence in distribution, marginal (or element-wise) convergence is not enough to imply joint convergence. The Cramer–Wald device provides one work around. To show that $X_n \xrightarrow{d} X$, it suffices to show that $a^T X_n \xrightarrow{d} a^T X$ for all $a \in \mathbb{R}^k$ and if $X_n \xrightarrow{d} X$, then $a^T X_n \xrightarrow{d} a^T X$. There is a special case when marginal convergence in distribution does imply joint convergence in distribution. This is when all but one of the entries of X are constant. More precisely, one can show, that if X_n, Y_n, X are random vectors and y is a constant, then

$$X_n \xrightarrow{d} X, Y_n \xrightarrow{p} y \Longrightarrow (X_n, Y_n) \xrightarrow{d} (X, y).$$

Note that since y is a constant, the condition $Y_n \xrightarrow{p} y$ is equivalent to $Y_n \xrightarrow{d} y$. This theorem can be combined with the continuous mapping theorem to prove Slutsky's theorem which is a real workhorse of asymptotic statistics.

Theorem 2 (Slutsky's)**.** *Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$, then*

1. $X_n + Y_n \xrightarrow{d} X + c$
2. $Y_n X_n \xrightarrow{d} cX$,
3. $X_n / Y_n \xrightarrow{d} X/c$ *provided* $c \neq 0$.

Proof. As claimed above, we know that $(X_n, Y_n) \xrightarrow{d} (X, c)$. The function $(x, y) \mapsto x + y$, $(x, y) \mapsto xy$ and $(x, y) \mapsto x/y$ are all continuous on their domains. Thus, by the continuous mapping theorem the above results hold. □

Example 1 (One-sided t-test)**.** Suppose $X_1, \dots, X_n$ are i.i.d. with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Suppose we wish to test $H_0 : \mu \leq \mu_0$ against $H_1 : \mu > \mu_0$.

If X_i was normally distributed, we know that the uniformly most powerful unbiased test rejects when $T_n \geq t_{n-1,\alpha}$ where

$$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n},$$

and

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

When X_i are normally distributed and the null holds, we know that T_n has the students t distribution with $n - 1$ degrees of freedom. We would like to know the asymptotic distribution of T_n in the non-normal cases. In particular,

1. What is the asymptotic distribution of T_n when $\mu = \mu_0$?
2. What is the asymptotic distribution of T_n when $\mu > \mu_0$?

With Slutsky's theorem we can answer both of these,

1. We have seen before that the weak law of large numbers implies that $S_n \xrightarrow{p} \sigma$ and that the central limit theorem implies that $\sqrt{n}(\bar{X}_n - \mu_0) \xrightarrow{d} \mathsf{N}(0, \sigma^2)$. Thus, Slutsky's theorem implies that $T_n \xrightarrow{d} \mathsf{N}(0, 1)$.
2. If $\mu > \mu_0$, then we can write T_n as,

$$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu) + \sqrt{n}(\mu - \mu_0)}{S_n} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} + \frac{\sqrt{n}(\mu - \mu_0)}{S_n}.$$

   We know that $\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} \mathsf{N}(0, 1)$ but $\frac{\sqrt{n}(\mu - \mu_0)}{S_n} \xrightarrow{p} +\infty$. Thus, when $\mu > \mu_0$, $T_n \xrightarrow{d} +\infty$. Thus, for $\mu > \mu_0$,

$$\mathbb{P}_{\mu,\sigma^2}(T_n \geq t_{n-1,\alpha}) \overset{n \to \infty}{\to} 1.$$

   Therefor, under any alternative, the test has asymptotic power 1.

Example 2 (Testing variance)**.** Suppose $X_1, X_2, \ldots$ are i.i.d. with $\mathbb{E}[X_1] < \infty$, $\mathrm{Var}(X_1) = \sigma^2$ and $\mathbb{E}[(X_1-\mu)^4] = \mu_4 < \infty$. Let $Y_i = (X_i-\mu)^2$ then $\mathbb{E}[Y_i] = \sigma^2$ and $\mathrm{Var}(Y_i) = \mathbb{E}[(X_i-\mu)^4] - \mathbb{E}[(X_i-\mu)^2]^2 = \mu_4 - \sigma^4$. Thus,

$$T_n = \frac{\sqrt{n}(\bar{Y}_n - \sigma^2)}{\sqrt{\mu_4 - \sigma^4}} \to \mathsf{N}(0,1).$$

Example 3 (Pearson's chi-squared)**.** Suppose $X_1, \ldots, X_n$ are i.i.d. with distribution $\mathsf{Multinomial}_k(1,p)$. That is each X_i is a vector in $\{0,1\}^k$ with exactly one entry equal to one and,

$$\mathbb{P}(X_{i,j} = 1) = p_j,$$

for every j. Let $N = \sum_{i=1}^n X_i \sim \mathsf{Multinomial}_k(n,p)$ and let $\widehat{p} = \frac{N}{n}$. Let H_0 be the hypothesis $p = p_0$ and let H_1 be $p \neq p_0$. Pearson's chi-squared test statistic of H_0 against H_1 is,

$$Q_n = \sum_{i=1}^k \frac{(N_j - np_{0,j})^2}{np_{0,j}}.$$

We will show that under H_0, $Q \xrightarrow{d} \chi^2_{k-1}$ as $n \to \infty$. It suffices to write $Q_n = W_n^T W_n$ where $W_n \xrightarrow{d} \mathsf{N}(0, I_{k-1})$. With this in mind, define,

$$Z_n = \begin{bmatrix} \frac{N_1 - np_{0,1}}{\sqrt{np_{0,1}}} \\ \vdots \\ \frac{N_k - np_{0,k}}{\sqrt{np_{0,k}}} \end{bmatrix} \in \mathbb{R}^k.$$

Note that $\mathbb{E}_{H_0}[Z] = 0$ and, for $i \neq j$,

$$\begin{aligned} \mathrm{Cov}_{H_0}(Z_{n,i} Z_{n,j}) &= \frac{1}{n\sqrt{p_{0,i}p_{0,j}}} \mathrm{Cov}_{H_0}(N_{n,i} N_{n,j}) \\ &= \frac{1}{n\sqrt{p_{0,i}p_{0,j}}} np_{0,i}p_{0,j} \\ &= \sqrt{p_{0,i}p_{0,j}}. \end{aligned}$$

And

$$\mathrm{Var}_{H_0}(Z_{n,j}) = \frac{1}{np_{0,j}} \mathrm{Var}_{H_0}(N_{n,j}) = \frac{np_{0,j}(1-p_{0,j})}{np_{0,j}} = 1 - p_{0,j}.$$

Thus, $\mathbb{E}[Z_n] = 0$ and $\mathrm{Var}(Z_n) = \Sigma$ where $\Sigma = I - \sqrt{p_0}\sqrt{p_0}^T$. Furthermore, by the multivariate CLT, $Z_n \xrightarrow{d} \mathsf{N}_K(0, \Sigma)$. Now let Γ be an orthogonal matrix with first row equal to $\sqrt{p_0}$. It follows that,

$$\begin{aligned} Z_n^T Z_n &= (\Gamma Z_n)^T (\Gamma Z_n) \\ &= V_n^T V_n. \end{aligned}$$

By Slutsky's we have $V_n \xrightarrow{d} V \sim \Gamma \mathsf{N}_k(0, \Sigma) = \mathsf{N}_k(0, \Gamma^T \Sigma \Gamma)$. Furthermore,

$$\Gamma^T \Sigma \Gamma = \Gamma^T \Gamma - \Gamma^T \sqrt{p_0}\sqrt{p_0}^T \Gamma = I - e_1 e_1^T = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

Thus $V^T V \sim \chi^2_{k-1}$ and $Q_n \xrightarrow{d} \chi^2_{k-1}$.

2 Delta method

The central limit theorem tells us that $\sqrt{n}(\bar{X}_n - \mu)$ converges in distribution to $\mathsf{N}(0, \sigma^2)$. Often we aren't just interested in the mean. The *delta method* allows us to study the asymptotic distribution of functions of the mean.

Theorem 3 (Delta method 1)**.** *Suppose that $X_1, X_2, \dots$ are i.i.d. with mean μ and variance $\sigma^2 < \infty$. Suppose that f is differentiable at μ, then*

$$\sqrt{n}(f(\bar{X}_n) - f(\mu)) \xrightarrow{d} \mathsf{N}(0, [f'(\mu)]^2\sigma^2).$$

Proof. We can prove this result by combining Slutsky's theorem with a Taylor's approximation. We have

$$f(\bar{X}_n) = f(\mu) + f'(\mu)(\bar{X}_n - \mu) + o(\bar{X}_n - \mu).$$

By the central limit theorem $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathsf{N}(0, \sigma^2)$ and so $o(\bar{X}_n - \mu) = o_p(\sqrt{n})$. Thus, rearranging the above we get,

$$\sqrt{n}(f(\bar{X}_n) - f(\mu)) = f'(\mu)\sqrt{n}(\bar{X}_n - \mu) + o_p(1).$$

Slutsky's theorem thus implies that $\sqrt{n}(f(\bar{X}_n) - f(\mu)) \xrightarrow{d} f'(\mu)\mathsf{N}(0, \sigma^2) = \mathsf{N}(0, [f'(\mu)]^2\sigma^2)$. □

The delta method can be generalized to situations other than that of the central limit theorem. There are also higher dimensional versions like the following.

Theorem 4 (Delta mathod 2 (higher dimensional))**.** *Suppose that $X_1, X_2, \dots$ are random k-vectors such that,*

$$a_n(X_n - c) \xrightarrow{d} Y.$$

If f is a real-valued function that is differentiable at c, then

$$a_n(f(X_n) - f(c)) \xrightarrow{d} \nabla f(c)^T Y.$$

The proof is via a Taylor's approximation like the previous version.

Example 4. Suppose $X_1, X_2, \dots$ are i.i.d. random vectors with $\mathbb{E}[X_1] = \theta \neq 0$ and $\mathrm{Cov}(X_1) = \Sigma$. Define $\phi(h) = \frac{1}{2}\|h\|_2^2$. By the multivariate central limit theorem, $\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{d} \mathsf{N}_k(0, \Sigma)$. Thus,

$$\sqrt{n}(\phi(\bar{X}_n) - \phi(\theta)) \xrightarrow{d} \theta^T \mathsf{N}(0, \Sigma) = \mathsf{N}(0, \theta^T\Sigma\theta),$$

since $\nabla h(\theta) = \theta$.

Example 5. Let $S_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n^2$. We know that $S_n^2 \xrightarrow{p} \sigma^2 = \mathrm{Var}(X_1)$, but can we say something about the limiting distribution of S_n^2? Note that $S_n^2 = \phi(\bar{X}_n, \bar{X^2}_n)$ where $\phi(x, y) = y - x^2$. Note that $\nabla\phi(x, y) = (-2x, 1)^T$. Also, assume X_1 has a finite fourth moment,

$$\sqrt{n}\left(\begin{bmatrix}\bar{X}_n \\ \bar{X^2}n\end{bmatrix} - \begin{bmatrix}\mu \\ \mu^2 + \sigma^2\end{bmatrix}\right) \xrightarrow{d} \mathsf{N}_2(0, \Sigma).$$

Where

$$\Sigma = \begin{bmatrix}\mathrm{Var}(X) & \mathrm{Cov}(X, X^2) \\ \mathrm{Cov}(X, X^2) & \mathrm{Var}(X^2)\end{bmatrix}$$

Note that $\nabla\phi(\mu, \mu^2 + \sigma^2) = (-2\mu, 1)^T$. Thus,

$$\begin{aligned}
\nabla\phi(\mu, \mu^2 + \sigma^2)^T \Sigma \nabla\phi(\mu, \mu^2 + \sigma^2) &= [-2\mu, 1]\begin{bmatrix}-2\mu\sigma^2 + \mathrm{Cov}(X, X^2) \\ -2\mu\,\mathrm{Cov}(X, X^2) + \mathrm{Var}(X^2)\end{bmatrix} \\
&= 4\mu^2\sigma^2 - 4\mu\,\mathrm{Cov}(X, X^2) + \mathrm{Var}(X^2) \\
&=: \gamma.
\end{aligned}$$

By the delta-method $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} \mathsf{N}(0, \gamma)$.

If the first derivate of our function is zero, then we can use the higher order delta method to get a better approximation.

Theorem 5 (Delta method 3 (higher order)). *Suppose that X_n are random k-vectors such that*

$$r_n(X_n - \theta) \xrightarrow{d} X,$$

where r_n is a deterministic function with $r_n \to +\infty$. Let ϕ be a real-valued function that is twice differentiable at θ with $\phi'(\theta) = 0$. Then,

$$r_n^2(\phi(X_n) - \phi(\theta)) \xrightarrow{d} \frac{1}{2} X^T \nabla^2 \phi(\theta) T.$$

Note that since $r_n \to +\infty$, $r_n^2 > r_n$ for sufficiently large n. Thus, the rate of convergence of $\phi(X_n)$ to $\phi(\theta)$ is faster. This is because we have to multiply by large numbers in order to have $\phi(X_n) - \phi(X)$ converge to a non-trivial distribution. We will prove the higher order delta method at the start of the next class.