Math 502AB - Lecture 18

Dr. Jamshidian

October 30, 2017

1 Lecture - Part 1

1.1 Relationships Between Various Modes of Convergence

1. $X_n \xrightarrow{\text{D}} X \not\Rightarrow X_n \xrightarrow{\text{P}} X$

   Example:

   Let $X \sim unif(-1, 1)$. This implies that $-X \sim unif(-1, 1)$.

   $$X_n = \begin{cases} x & \text{if } n \text{ is odd} \\ -x & \text{if } n \text{ is even} \end{cases}$$

   Clearly, we have:

   $$X_n \xrightarrow{\text{D}} X$$

   However, consider:

   $$|X_n - X| = \begin{cases} 0 & \text{if } n \text{ is odd} \\ 2|X| & \text{if } n \text{ is even} \end{cases}$$

   Let $\epsilon = 1$. Then:

   $$P(|X_n - X| < 1) = \begin{cases} 1 & \text{if } n \text{ is odd} \\ P(2|X| < 1) = P(|X| < \frac{1}{2}) = \frac{1}{2} & \text{if } n \text{ is even} \end{cases}$$

   This shows the above: that convergence in *distribution* **does not** imply convergence in *probability.*

2. **Theorem:** If $X_n \xrightarrow{\text{P}} X$, then $X_n \xrightarrow{\text{D}} X$. This essentially says that convergence in *distribution* is the weakest kind of convergence we ahve talked about.

3. **Theorem:** $X_n \xrightarrow{\text{P}} a \iff X_n \xrightarrow{\text{D}} a$

   Proof:

   ($\Rightarrow$): This follows from the above theorem.

   ($\Leftarrow$): Assume that $X_n \xrightarrow{\text{D}} a$. This means that:

   $$F_{X_n}(x) \to F_X(x) = \begin{cases} 1 & x \geq a \\ 0 & x < a \end{cases}$$

   Let $\epsilon > 0$. We then look at the limit:

   $$\begin{aligned} \lim_{n\to\infty} P\left[|X_n - a| < \epsilon\right] &= \lim_{n\to\infty} P\left[a - \epsilon \leq X_n \leq a + \epsilon\right] \\ &= \lim_{n\to\infty} F_{X_n}(a+\epsilon) - F_{X_n}(a-\epsilon) \\ &= 1 \end{aligned}$$

4. **Theorem:** Suppose that $X_n \xrightarrow{\text{D}} X$, and that g is a continuous function on the support of X. Then $g\left(X_n\right) \xrightarrow[n\to\infty]{\text{D}} g(x)$. (The proof of this theorem requires measure theory)

5. **Theorem: (Slutsky's Theorem)** Let $X_n \xrightarrow{\text{D}} X$ and $Y_n \xrightarrow{\text{P}} a$. Then:

   (a) $X_n Y_n \xrightarrow{\text{D}} aX$

   (b) $X_n + Y_n \xrightarrow{\text{D}} X + a$

   Example: Show that:

   $$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \xrightarrow{\text{D}} N(0,1)$$

   If $\lim_{n\to\infty} Var(S_n^2) = 0$, then we have $S_n^2 \xrightarrow{\text{P}} \sigma^2$. This implies that:

   $$\Rightarrow \frac{S_n^2}{\sigma^2} \xrightarrow{\text{P}} 1 \Rightarrow \sqrt{\frac{S_n^2}{\sigma^2}} = \frac{S_n}{\sigma} \xrightarrow{\text{P}} 1$$

   By the **Central Limit Theorem**:

   $$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{\text{D}} N(0,1)$$

   By **Slutsky's theorem**, if we multiply this by $\frac{\sigma}{S_n}$ (which converges to 1 as shown above), we get the Example result.

1.2 The Delta Method

Idea: The general idea of the delta method is to approximate non-linear functions by their Taylor expansion (a polynomial) and deal with them in an easier way. For example, if we had $\log(X)$ and wanted to write the expected value, , we don't have a good formula for it. But if we *expand* it, and we know moments of x, we can approximate it.

Let $g(x)$ be a function with $n+1$ continuous derivatives. Then for a given real value x_0:

$$\begin{aligned} g(x) = g(x_0) + g'(x_0)(x-x_0) + \frac{g''(x_0)}{2!}(x-x_0)^2 + \dots \\ + \frac{g^n(x_0)}{n!}(x-x_0)^n + \frac{g^{(n+1)}(\eta)}{(n+1)!}(x+x_0)^{n+1} \end{aligned}$$

Where η is a number between x and x_0. **Note:** This is *not* an approximation to $g(x)$, but the answer itself (for some η).

1.2.1 Linear Approximation

Let's begin with a linear approximation of a function $g(x)$, around $x_0 = a$.

$$g(x) = g(a) + g'(a)(x-a) + o(|x-a|)$$

Note: In terms of *convergence in probability*, we write:

$$Y_n = o_p(X_n) \iff \frac{Y_n}{X_n} \xrightarrow[n\to\infty]{P} 0$$

Examples:

1. Suppose we wanted to find $E[g(x)]$. Then, we would have:

   $$E[g(X)] \cong E[g(a) + g'(a)(x-a)]$$

   In particular, let $a = E[X] = \mu$. Then:

   $$\begin{aligned} E[g(X)] &\cong E[g(\mu) + g'(\mu)(X-\mu)] \\ &= g(\mu) = g(E[X]) \end{aligned}$$

   This says that if $g(x)$ is a function that is approximated well with a line near $x = \mu$, you can use this linear approximation. To find the variance, similarly we have:

   $$\begin{aligned} Var[g(X)] &= Var[g(\mu) + g'(\mu)(X-\mu)] \\ &= (g'(\mu))^2\, Var(X) \end{aligned}$$

2. Let $X_1, \ldots, X_n \sim \mathit{Bernoulli}(p)$ (**iid**). Consider the *sample proportion*, $\hat{p} = \frac{\sum X_i}{n}$.

$$E[\hat{p}] = E[X_i] = p$$

To estimate the odds, $\frac{p}{1-p}$, we use $\frac{\hat{p}}{1-\hat{p}}$. The question is, what is our random variable? It is a *non-linear* function of $\hat{p}$, so it is non-trivial to calculate the expectation. How would we go about computing the expected value? Consider:

$$g(x) = \frac{x}{1-x}$$

First, let's find $E[g(x)]$. Consider $x = \hat{p}$. Then:

$$E\left[\frac{\hat{p}}{1-\hat{p}}\right] \cong \frac{p}{1-p}$$

Now, to find the variance:

$$\begin{aligned} g'(x) &= \frac{1}{(1-x)^2} \\ \Rightarrow Var[g(x)] &\cong \left[g'(p)\right]^2 Var\left(\hat{p}\right) \\ &= \left(\frac{1}{(1-p)^2}\right)^2 \cdot \frac{p(1-p)}{n} = \frac{p}{n(1-p)^3} \end{aligned}$$

This estimator is *approximately* unbiased, and, as n gets large, the variance approaches 0.

3. Let $X \sim \mathit{exp}(1)$. Approximate $E[\log x]$ and $Var(\log x)$ using the *delta method*. Well, we know that if we wanted to calculate the exact value of the expected value, we could find the "true" value:

$$E[\log x] = \int_0^\infty (\log x) e^{-x} dx \cong -0.5772$$

We know that

$$E[X] = 1 \quad \text{and} \quad Var(X) = 1$$

If we wanted to use the *linear approximation*, we would have:

$$\begin{aligned} E[g(X)] &\cong g\left(E[X]\right) \\ &\cong \log 1 = 0 \end{aligned}$$

So, let's use the **second order Taylor expansion**. We then have:

$$
\begin{aligned}
g(x) &\cong g(\mu) + g'(\mu)(x-\mu) + \frac{g''(\mu)}{2}(x-\mu)^2 \\
g(x) &= \log x \quad g'(x) = \frac{1}{x} \quad g''(x) = -\frac{1}{x^2} \\
\Rightarrow g(x) &\cong \log \mu + \frac{x-\mu}{\mu} - \frac{1}{2\mu^2}(x-\mu)^2 \\
E[g(x)] &\cong \log \mu - \frac{1}{2\mu^2} Var(x) \\
E[\log x] &\cong \log 1 - \frac{1}{2} = -\frac{1}{2}
\end{aligned}
$$

Thus, we have shown that by increasing the order, we have vastly improved the approximation. If we want to find the **variance**, we will need to find the 4^{th} moment. We need to make an expansion here:

$$
g(x) = \log \mu + \frac{1}{\mu}(X-1) - \frac{1}{2\mu^2}(x^2 - 2\mu x + \mu^2)
$$

Then, we have:

$$
\begin{aligned}
Var[g(X)] = & \frac{1}{\mu^2} Var(X) - \frac{1}{4\mu^4} Var(X^2) + \frac{1}{\mu^2} Var(X) \\
& - \frac{1}{2\mu^3} Cov(X, X^2) + \frac{1}{\mu^2} Cov(X, X) + Cov\left(-\frac{1}{2\mu^2} X^2, \frac{1}{\mu} X\right)
\end{aligned}
$$

With a bit of computation, the results follow.

2 Lecture - Part 2

2.1 The General Delta Method Theorem

Theorem:

Let a_n be an increasing sequence of real numbers such that $a_n \to \infty$ and let θ be a fixed value. Furthermore, suppose that $g(x)$ is a continuously differentiable function such that $g'(\theta) \neq 0$. Then, if $Y_n = a_n(X_n - \theta) \xrightarrow[n\to\infty]{\text{D}} Y$ where Y is a random variable, and X_n is a sequence of random variables, then:

$$
a_n \left[g(X_n) - g(\theta) \right] \xrightarrow{\text{D}} g'(\theta) \cdot Y
$$

Proof:

By Taylor's Theorem, we have:

$$
g(X_n) = g(\theta) + g'(w_n)(X_n - \theta)
$$

Where w_n is between X_n and θ

Facts:

1. $X_n \xrightarrow{\text{P}} \theta$

   Proof: We know that $a_n(X_n - \theta) \xrightarrow{\text{D}} Y$, and that $\frac{1}{a_n} \xrightarrow{\text{P}} 0$. By **Slutsky's Theorem**, we have:

$$\frac{1}{a_n} a_n(X_n - \theta) \xrightarrow{\text{D}} 0$$
$$\Rightarrow X_n - \theta \xrightarrow{\text{D}} 0 \Rightarrow X_n \xrightarrow{\text{D}} \theta \Rightarrow X_n \xrightarrow{\text{P}} \theta$$

2. $w_n \xrightarrow{\text{P}} \theta$

   Proof: Since w_n is between θ and x_n, we know that

$$|X_n - \theta| \geq |w_n - \theta|$$

   This implies that

$$P\left(|X_n - \theta| < \epsilon\right) \leq P\left(|w_n - \theta| < \epsilon\right)$$
$$\Rightarrow P(|w_n - \theta < \epsilon|) \to 1$$

3. Since g' is *continuous*, using **fact 2**, we have

$$g'(w_n) \xrightarrow{\text{P}} g'(\theta)$$

Now consider, by Taylor's Theorem:

$$a_n(g(X_n) - g(\theta)) = g'(w_n)a_n(X_n - \theta)$$

What do we know about this? Well $g'(w_n) \xrightarrow{\text{P}} g'(\theta)$, and $a_n(X_n - \theta) \xrightarrow{\text{D}} Y$. Thus we have, by **Slutsky's Theorem**:

$$a_n(g(X_n) - g(\theta)) \xrightarrow{\text{D}} g'(\theta)Y$$

Theorem: (5.5.4)

Let Y_n be a sequence of random variables that satisfy:

$$\sqrt{n}(Y_n - \theta) \xrightarrow{\text{D}} N(0, \sigma^2)$$

for a given function g and a specific value of θ. Suppose that $g'(\theta)$ exists and $g'(\theta) \neq 0$. Then:

$$\sqrt{n}[g(Y_n) - g(\theta)] \xrightarrow{\text{D}} N\left(0, \sigma^2 \left(g'(\theta)^2\right)\right)$$

For example, if $Y \sim N(0, \sigma^2)$, then we would have:

$$g'(\theta)Y \sim N(0, \sigma^2(g'(\theta))^2)$$

2.1.1 Multidimensional Generalization

Let $g(\vec{x}) : \mathbb{R}^n \to \mathbb{R}^1$. Then for any given vector $\vec{x_0} \in \mathbb{R}^n$:

$$g(\vec{x}) = g(\vec{x_0}) + \nabla g(\vec{x_0})^T(\vec{x} - \vec{x_0}) + \frac{1}{2}(\vec{x} - \vec{x_0})^T H(\eta)(x - x_0)$$

Where η is a value between $\vec{x}$ and $\vec{x_0}$, and where:

$$\nabla g(\vec{x}) = \begin{pmatrix} \frac{\partial g}{\partial x_1} \\ \vdots \\ \frac{\partial g}{\partial x_n} \end{pmatrix}$$

And:

$$H(x) = \begin{pmatrix} \frac{\partial^2 g}{\partial x_1^2} & \cdots & \frac{\partial^2 g}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 g}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 g}{\partial x_n \partial x_1} \end{pmatrix}$$

Let $\vec{X} = (X_1, ..., X_n)^T$ be a random vector with mean $\vec{\mu} = (\mu_1, ..., \mu_n)^T$. (**Note that:** $E[X_i] = \mu_i$). Then,

$$g(\vec{X}) \cong g(\vec{\mu}) + \nabla g(\mu)^T(x - \mu)$$

We can now have:

$$E[g(\vec{X})] \cong g(\vec{\mu})$$
$$Var[g(\vec{X})] = \nabla g(\vec{\mu})^T Cov(\vec{X}) \nabla g(\vec{\mu}$$
$$\text{Where, } Cov(\vec{X}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \ddots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$$

Example:

The position of an aircraft relative to an observer is estimated by measuring the aircraft distance R from the observer, and an angle Θ which is the angle from the observer to the aircraft, relative to the horizon. We then have the *altitude* of the aircraft $Y = R\sin(\Theta)$.

Suppose that both Θ and R are subject to random errors, and that R and Θ are independent. Provided that $E[R] = 1000$, $E[\Theta] = \frac{\pi}{4}$, $Var(R) = 10$, and $Var(\Theta) = \frac{\pi^2}{100}$, obtain the approximate variance of the altitude of the plane.

First, we note that $g(R, \Theta) : \mathbb{R}^2 \to \mathbb{R}^1$. Thus we have:

$$
\begin{aligned}
g(R,\Theta) &= R\sin\Theta \\
\nabla g(R,\Theta) &= \begin{pmatrix} \sin\Theta \\ R\cos\Theta \end{pmatrix} \\
\vec{\mu} &= \begin{pmatrix} E[R] \\ E[\Theta] \end{pmatrix} = \begin{pmatrix} 1000 \\ \frac{\pi}{4} \end{pmatrix} \\
\Rightarrow \nabla g(\vec{\mu}) &= \begin{pmatrix} \sin\frac{\pi}{4} \\ 1000\cos\frac{\pi}{4} \end{pmatrix} \\
Cov(R,\Theta) &= \begin{pmatrix} Var(R) & Cov(R,\Theta) \\ Cov(R,\Theta) & Var(\Theta) \end{pmatrix} \\
Var(Y) &= \begin{pmatrix} \sin\frac{\pi}{4} & 1000\cos\frac{\pi}{4} \end{pmatrix} \begin{pmatrix} 10 & 0 \\ 0 & \frac{\pi^2}{100} \end{pmatrix} \begin{pmatrix} \sin\frac{\pi}{4} \\ 1000\cos\frac{\pi}{4} \end{pmatrix} \\
&= 10\sin^2\frac{\pi}{4} + 10^4\pi^2\cos^2\frac{\pi}{4}
\end{aligned}
$$