| | |
|---|---|
| **Stat210B: Theoretical Statistics** | **Lecture Date: January 16, 2007** |
| Lecture 1 | |
| *Lecturer: Michael I. Jordan* | *Scribe: Karl Rohe* |

**Reading:** Chapter two of van der Vaart's book *Asymptotic Statistics*.

# 1 Convergence

There are four types of convergence that we will discuss.

**Definition 1.** **Weak convergence**, also known as convergence in distribution or law, is denoted

$$X_n \xrightarrow{d} X$$

A sequence of random variables $X_n$ converges in law to random variable $X$ if $P(X_n \leq x) \to P(X \leq x)$ for all $x$ at which $P(X \leq x)$ is continuous.

**Definition 2.** $X_n$ is said to **converge in probability** to $X$ if for all $\epsilon > 0$, $P(\ d(X_n, X) > \epsilon) \to 0$. This is denoted $X_n \xrightarrow{P} X$.

**Definition 3.** $X_n$ is said to **converge in r^th^ mean** to $X$ if $E(\ d(X_n, X)^r) \to 0$. This is denoted $X_n \xrightarrow{r} X$.

**Definition 4.** $X_n$ is said to **converge almost surely** to $X$ if $P(\ \lim_n d(X_n, X) = 0) = 1$. This is denoted $X_n \xrightarrow{\text{a.s.}} X$.

**Theorem 5.**

- *A.s. convergence implies convergence in probability.*
- *Convergence in $r^{th}$ mean also implies convergence in probability.*
- *Convergence in probability implies convergence in law.*
- $X_n \xrightarrow{d} c$ *implies* $X_n \xrightarrow{P} c$*. Where $c$ is a constant.*

**Theorem 6.** ***The Continuous Mapping Theorem***

*Let $g$ be continuous on a set $C$ where $P(X \in C) = 1$. Then,*

1. $X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$
2. $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$
3. $X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$

**Example 7.** Let $X_n \xrightarrow{d} X$, where $X \sim N(0, 1)$. Define the function $g(x) = x^2$. The CMT says $g(X_n) \xrightarrow{d} g(X)$. But, $X^2 \sim \chi^2_1$. So, $g(X_n) \xrightarrow{d} \chi^2_1$.

**Example 8.** Let $X_n = \frac{1}{n}$ and $g(x) = \mathbf{1}_{x>0}$. Then $X_n \xrightarrow{d} 0$ and $g(X_n) \xrightarrow{d} 1$. But, $g(0) \neq 1$.

**Theorem 9.** ***Slutsky's Theorems***

1. $X_n \xrightarrow{d} X$ and $X_n - Y_n \xrightarrow{P} 0$ together imply $Y_n \xrightarrow{d} X$.
2. $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$ together imply

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X \\ c \end{pmatrix}$$

3. $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$ together imply $X_n + Y_n \xrightarrow{d} X + c$.
4. $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$ together imply $X_n Y_n \xrightarrow{d} Xc$.
5. $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{P} c$ together imply $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$ when $c \neq 0$.

**Example 10.** Let $X_n$ be iid with mean $\mu$ and variance $\sigma^2$. From the Weak Law of Large Numbers we know the sample mean $\bar{X}_n \xrightarrow{P} \mu$. Similarly, $\frac{1}{n}\sum_i X_i^2 \xrightarrow{P} E(X^2)$. By Slutsky's Theorem we know $S_n^2 = \frac{1}{n}\sum_i X_i^2 - \bar{X}^2 \xrightarrow{d} \sigma^2$. Together with the CMT, this implies $S_n \xrightarrow{P} \sigma$. From the CLT $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$. Together these facts imply

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

Where this last equality is due to Slutsky. So, the t-statistic is asymptotically normal.

**Definition 11.** $X_n = o_p(R_n)$, pronounced "$X_n$ is little oh-pee-$R_n$," means $X_n = Y_n R_n$, where $Y_n \xrightarrow{P} 0$.

**Definition 12.** $X_n = O_p(R_n)$, pronounced "$X_n$ is big oh-pee-$R_n$," means $X_n = Y_n R_n$, where $Y_n = O_p(1)$. $O_p(1)$ denotes a sequence $Z_n$ which for any $\epsilon > 0$ there exists an $M$ such that $P(|Z_n| > M) < \epsilon$.

**Lemma 13.** *Let $R : \mathbb{R}^k \to \mathbb{R}$ and $R(0) = 0$. Let $X_n = o_p(1)$. Then, as $h \to 0$, for all $p > 0$*

1. $R(h) = o(\|h\|^p)$ *implies* $R(X_n) = o_p(\|X_n\|^p)$.
2. $R(h) = O(\|h\|^p)$ *implies* $R(X_n) = O_p(\|X_n\|^p)$.

To prove this, apply the CMT to $\frac{R(h)}{\|h\|^p}$.

- Any random variable is **tight**. I.e. for all $\epsilon > 0$, there exists and $M$ such that $P(\|X\| > M) < \epsilon$.
- $\{X_\alpha : \alpha \in A\}$ is called **Uniformly Tight** (UT) if for all $\epsilon > 0$, there exists and $M$ such that $\sup_\alpha P(\|X_\alpha\| > M) < \epsilon$.

**Theorem 14.** ***Prohorov's theorem (cf. Heine-Borel)***

1. *If* $X_n \xrightarrow{d} X$, *then* $X_n$ *is UT.*
2. *If* $\{X_n\}$ *is UT, then there exits a subsequence* $\{X_{nj}\}$ *with* $X_{nj} \xrightarrow{d} X$ *as* $j \to \infty$ *for some* $X$.

As we move on in the course we will wish to describe weak convergence for things other than random variables. At this point, the our previous definition will not make sense. We can then use this following theorem as a definition.

**Theorem 15.** ***Portmanteau***

$$X_n \xrightarrow{d} X \Longleftrightarrow Ef(X_n) \rightarrow Ef(X) \text{ for all bounded continuous } f.$$

In this theorem, "bounded and continuous" can be replaced with

- "continuous and vanishes outside of compacta"
- "bounded and measurable, such that $P(X \in C(g)) = 1$" where $C(g)$ is the set of $g$'s continuity points.
- "bounded Lipshitz"
- "$f(X) = e^{itX}$." This is the next theorem.

**Theorem 16.** ***Continuity theorem***

$$X_n \xrightarrow{d} X \Longleftrightarrow E\exp(it^T X_n) \rightarrow E\exp(it^T X)$$

**Example 17.** To demonstrate why $f$ must be bounded, observe what happens if $g(x) = x$ and

$$X_n = \begin{cases} n & \text{w.p. } 1/n \\ 0 & \text{otherwise} \end{cases}$$

$X_n \xrightarrow{d} 0$, $Eg(X_n) = 1 \neq Eg(0) = 0$.

**Example 18.** To demonstrate why $f$ must be continuous, observe what happens if $X_n = 1/n$ and

$$g(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$$

**Theorem 19.** ***(Scheffè)*** *For random variables, $X_n \geq 0$, if $X_n \xrightarrow{a.s.} X$ and $EX_n \rightarrow EX < \infty$, then $E|X_n - X| \rightarrow 0$. For densities, if $f_n(x) \rightarrow g(x)$ for all $x$, then $\int |f_n(x) - g(x)|dx \rightarrow 0$.*