36-752 Advanced Probability Overview **Spring 2018**

Lecture 18: April 5

Lecturer: Alessandro Rinaldo *Scribes: Wanshan Li*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

18.1 Continuous Mapping Theorem

Let $\{X_n\}_{n=1}^{\infty}$ and X be random variables taking values in metric space $(\mathcal{X}, d)$. Recall that last time we defined that $X_n \overset{D}{\to} X$ when $\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)]$ for all bounded continuous function g. This relationship can actually be generalized as follows:

Theorem 18.1. *$X_n \overset{D}{\to} X$ if and only if*

$$\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)]$$

for all bounded g that are continuous a.e.$[\mu_X]$, or $\mu_X(\{x : g \text{ not continuous at } x\}) = 0$.

Remark For the direct proof of this theorem, you can see Theorem 3.9.1 on Durrett's book, or the section on weak convergence of Billingsley's book. You can also prove it by using Skorokhod's Representation Theorem given below:

Theorem 18.2 (Skorokhod's Representation Theorem)**.** *Suppose $X_n \overset{D}{\to} X$, all taking values in metric space $(\mathcal{X}, d)$, and the probability measure μ of X is separable. Then $\exists$ $\{Y_n\}$ and Y, taking values in $(\mathcal{X}, d)$, and defined on some probability space $(\Omega, \mathcal{F}, p)$ s.t. $X_n \overset{d}{=} Y_n$, $X \overset{d}{=} Y$, and*

$$Y_n \overset{a.s.}{\to} Y,$$

where a.s. is w.r.t. p.

In a narrow sense, the so-called continuous mapping theorem concerns the convergence in distribution of random variables, as we will discuss first. This theorem contains three parts. Roughly speaking, the main part of it says that if $X_n \overset{D}{\to} X$ and f is a *a.e.*$[\mu_X]$ continuous function, then $f(X_n) \overset{D}{\to} f(X)$.

Theorem 18.3 (**Continuous Mapping Theorem, I**)**.** *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables and X another random variable, all taking values in the same metric space $\mathcal{X}$. Let $\mathcal{Y}$ be a metric space and $f : \mathcal{X} \to \mathcal{Y}$ a measurable function. Define*

$$C_f = \{x : g \text{ is continuous at } x\}.$$

Suppose that $X_n \overset{D}{\to} X$ and $\mathbb{P}(X \in C_f) = 1$, then $f(X_n) \overset{D}{\to} f(X)$.

Proof. Denote the probability measure of X_n and X as μ_n and μ. Last time we saw that $\forall B \subset \mathcal{Y}$ closed,

$$\overline{f^{-1}(B)} \subset f^{-1}(B) \cup C_f^c,$$

where $\overline{f^{-1}(B)}$ is the closure of $\{x \in \mathcal{X} : f(x) \in B\}$. Therefore,

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}(f(x_n) \in B) &= \limsup_{n\to\infty} \mu_n(f^{-1}(B)) \\
&\le \limsup_{n\to\infty} \mu_n(\overline{f^{-1}(B)}) \\
&\overset{\text{Portmanteau}}{\le} \mu(\overline{f^{-1}(B)}) \\
&\le \mu(\overline{f^{-1}(B)}) + \mu(C_f^c) \\
&= \mu(\overline{f^{-1}(B)}) = \mathbb{P}(f(X) \in B).
\end{aligned}$$

This implies $f(X_n) \xrightarrow{D} f(X)$ by Portamanteau Theorem. ☐

Example 18.4. *Suppose $X_1, X_2, \cdots$ are i.i.d. samples from a (well behaved) distribution with mean and variance μ, σ^2. By CLT we have $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{D} N(0,1)$. Now by CMT we have $\left[\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)\right]^2 \xrightarrow{D} \chi_1^2$.*

Theorem 18.5 (Continuous Mapping Theorem, II). *Let $\{X_n\}_{n=1}^\infty$, X, $\{Y_n\}_{n=1}^\infty$ be random variables taking values in a metric space with metric d. Suppose that $X_n \xrightarrow{D} X$ and $d(X_n, Y_n) \xrightarrow{P} 0$, then $Y_n \xrightarrow{D} X$.*

Theorem 18.6 (Continuous Mapping Theorem, III). *Let $\{X_n\}_{n=1}^\infty$ and $\{Y_n\}_{n=1}^\infty$ be random variables. Suppose that $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{P} c$, then $(X_n, Y_n) \xrightarrow{D} (X, c)$. Furthermore, if $X_n \perp\!\!\!\perp Y_n$ and $Y_n \xrightarrow{D} Y$, then $(X_n, Y_n)' \xrightarrow{D} (X, Y)'$, and $X \perp\!\!\!\perp Y$.*

The CMT can also be generalized to cover the convergence in probability, as the following theorem does.

Theorem 18.7 (CMT for convergence in probability). *If $X_n \xrightarrow{P} X$ and f is continuous a.s.$[\mu_X]$, then $f(X_n) \xrightarrow{P} f(X)$.*

Remark Also notice the trivial fact that if $X_n \xrightarrow{a.s.} X$ then $f(X_n) \xrightarrow{a.s.} f(X)$. Therefore the CMT holds for all these three modes of convergence.

Proof. Fix an arbitrary $\varepsilon > 0$, we want to show that

$$\mathbb{P}(d_{\mathcal{Y}}(f(X_n), f(X)) > \varepsilon) \to 0, \text{ as } n \to \infty.$$

Fix $\delta > 0$ and let B_δ be the subset of $\mathcal{X}$ consisting of all x's such that $\exists y \in \mathcal{X}$ with $d_{\mathcal{X}}(x, y) < \delta$ and $d_{\mathcal{Y}}(f(x), f(y)) > \varepsilon$. Then

$$\{X \notin B_\delta\} \cap \{d_{\mathcal{Y}}(f(X_n), f(X)) > \varepsilon\} \subset \{d_{\mathcal{X}}(X_n, X) \ge \delta\}.$$

By the fact that $A^c \cap B \subset C \Rightarrow B \subset A \cup C$, we have

$$\mathbb{P}(d_{\mathcal{Y}}(f(X_n), f(X)) > \varepsilon) \le \mathbb{P}(X \in B_\delta) + \mathbb{P}(d_{\mathcal{X}}(X_n, X) \ge \delta) \triangleq T_1 + T_2.$$

Now $T_2 \to 0$ as $n \to \infty$ because $X_n \xrightarrow{P} X$. As for T_1, notice that

$$\mathbb{P}(X \in B_\delta) = \mathbb{P}(X \in B_\delta \cap C_f) \downarrow 0 \text{ as } \delta \downarrow 0.$$

Thus

$$\mathbb{P}(d_{\mathcal{Y}}(f(X_n), f(X)) > \varepsilon) \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

For n large enough and δ small enough. ☐

Remark If f (defined on $\mathcal{X}$) is $a.s.[\mu_X]$ continuous and g (defined on $\mathcal{Y}$) is continuous, then $g \circ f$ is $a.s.[\mu_X]$ continuous. Therefore by theorem 18.1, we can get the CMT immediately.

Example 18.8. *Suppose $X_1, X_2, \cdots$ are i.i.d. samples with mean and variance μ and σ^2. We want to estimate σ^2 by an estimator $\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2$. Consider $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. By WLLN we have*

$$\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 \xrightarrow{P} \sigma^2, \bar{X}_n \xrightarrow{P} \mu.$$

Using this, together with the fact that

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 - (\bar{X}_n - \mu)^2,$$

and $(\bar{X}_n - \mu)^2 \xrightarrow{P} 0$ by CMT, we have

$$(\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2, (\bar{X}_n - \mu)^2)' \xrightarrow{P} (\sigma^2, 0)'.$$

Now let $f(x, y) = x - y$, which is continuous, by CMT we have

$$\hat{\sigma}_n^2 = f(\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2, (\bar{X}_n - \mu)^2) \xrightarrow{P} \sigma^2.$$

A direct but useful corollary of continuous mapping theorem is the Slusky's Theorem.

Theorem 18.9 (Slusky's Theorem). *If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} c$, then*

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

Example 18.10. *By Central Limit Theorem we have $\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu) \xrightarrow{D} N(0, 1)$. By Law of Large Numbers we have $\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2$. Then by continuous mapping theorem $\hat{\sigma}_n \overset{P}{\sigma}$ and by Slusky's Theorem,*

$$\frac{\sqrt{n}}{\hat{\sigma}_n}(\bar{X} - \mu) = \frac{\sigma}{\hat{\sigma}_n}\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu) \xrightarrow{D} N(0, 1).$$

18.2 Tightness

Definition 18.11. *A sequence of probability measures $\{\mu_n\}_{n=1}^\infty$ is said to be tight if $\forall \varepsilon > 0$, $\exists C$ compact s.t.*

$$\mu_n(C) > 1 - \varepsilon, \ \forall n.$$

Equivalently, if $X_n \sim \mu_n$, then $\{X_n\}_{n=1}^\infty$ is tight or bounded by probability if $\forall \varepsilon > 0$, $\exists M$ s.t.

$$\sup_n \mathbb{P}(\|X_n\| > M) < \varepsilon.$$

In that case, we denote $X_n = O_P(1)$.

Example 18.12. *Cases where bounded in probability fails:*

1. $\mathbb{P}(X_n = c_n) = 1$, *and* $c_n \to \infty$.
2. $X_n \sim \text{Uniform}([-n, n])$. *(Note that then* $\mathbb{P}(X_n \in [-M, M]^c) = 1 - \frac{M}{n}$*).*

Theorem 18.13. *Let* $\{X_n\}_{n=1}^{\infty}$ *be a sequence of random variables taking values in* $\mathbb{R}^d$.

(i) If $X_n \overset{D}{\to} X$ *then* $\{X_n\}$ *is tight.*

(ii) ***Helly-Bray Selection Theorem.*** *If* $\{X_n\}$ *is tight, then* $\exists$ $\{n_k\}$ *s.t.* $X_{n_k} \overset{D}{\to} X$. *Further, if every convergent (in distribution) sub-sequence converges to the same* X, *then* $X_n \overset{D}{\to} X$.

Proof of (i). X is a random variable, so $\forall \varepsilon > 0$, let $M = M(\varepsilon) > 0$ s.t. $\mathbb{P}(\|X\| > M) < \varepsilon$. Then by Portmanteau Thm, for all $n \geq n_0(\varepsilon, M)$,

$$\mathbb{P}(\|X_n\| \geq M) \leq \mathbb{P}(\|X\| \geq M) + \varepsilon \leq 2\varepsilon.$$

On the other hand, we can find $M_1 > M$ s.t. for all $n < n_0$,

$$\mathbb{P}(\|X_n\| \geq M_1) \leq 2\varepsilon.$$

Now for $n \geq n_0$, we have

$$\mathbb{P}(\|X_n\| \geq M_1) \leq \mathbb{P}(\|X_n\| \geq M) \leq \mathbb{P}(\|X\| \geq M) + \varepsilon \leq 2\varepsilon.$$

Thus $\sup_n \mathbb{P}(\|X_n\| \geq M_1) \leq 2\varepsilon$. □

Theorem 18.14 (Polya Theorem). *Suppose* $\{X_n\}_{n=1}^{\infty}$ *and* X *take values in* $\mathbb{R}^d$, *and* $X_{n_k} \overset{D}{\to} X$. *If the c.d.f.* F *of* X *is continuous, or equivalently* $\mu_X \ll \lambda_d$, *where* λ_d *is the Lebesgue measure on* $\mathbb{R}^d$, *then (*F_n *is the c.d.f. of* X_n*)*

$$\sup_{X \in \mathbb{R}^d} |F_n(x) - F(x)| \to 0.$$

Theorem 18.15 (Glivenko-Cantelli Theorem). *Suppose* $X_1, X_2, \cdots ,, X_i \in \mathbb{R}$ *are i.i.d. samples from a distribution with c.d.f.* F. *For each* n, *let* $\hat{F}_n$ *be the empirical c.d.f.*

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_n \leq x\}.$$

We have

$$\|\hat{F}_n - F\|_\infty = \sup_{X \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \overset{a.s.}{\to} 0.$$

Remark

- Notice that $\mathbb{E}[\hat{F}_n(x)] = F(x)$ for all x. Therefore by SLLN we have $|\hat{F}_n(x) - F(x)| \overset{a.s.}{\to} 0$ for each x. However the Glivenko-Cantelli Theorem is much stronger than this because it asserts the uniform convergence.
- We often use another (even stronger) theorem instead, named after Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz, who in 1956 proved this inequality:

Theorem 18.16 (Dvoretzky-Kiefer-Wolfowitz). *Under the same condition, for any* $\varepsilon > 0$ *and all* n, *we have*

$$\mathbb{P}(\|\hat{F}_n - F\|_\infty \geq \varepsilon) \leq 2\exp\{-2n\varepsilon^2\}.$$