Lecture 14: August 22

Lecturer: Alessandro Rinaldo *Scribes: Trey McNeely*

14.1 Convergence in Measure/Probability

Let $(\Omega,\ \mathcal{F},\ \mu)$ be a measure space, and let $\{X_n\}$ be a sequence of random variables taking values in a metric space $(\mathcal{X},\ d)$. Let X also be a random variable taking a value in $(\mathcal{X},\ d)$. Recall that a metric $d:\ \mathcal{X}\times\mathcal{X}\mapsto[0,\ \infty)$ satisfies the following properties:

1. $d(x_1,\ x_2)\geq 0$
2. $d(x,\ y)\leq d(x,\ z)+d(y,\ z)$
3. $d(x,\ y)=0$ iff $x=y$

Definition 14.1 (Convergence in Measure) *X_n is said to converge in measure to X when $\forall\epsilon>0$, $\mu\big(\{\omega:\ d(X_n(\omega),\ X(\omega))>\epsilon\}\big)\to 0$ as $n\to\infty$. If μ is a probability measure, this becomes convergence in probability. This is denoted $X_n\xrightarrow{p}X$.*

To conclude that $X_n\xrightarrow{p}X$, we need to know joint distribution of $X_n,\ X$. See the the Bernoulli example, Example 13.3, for intuition.

14.1.1 Extension to Random Vectors

Now let $\{X_n\}$ be a sequence of random vectors in $\mathbb{R}^d$. Let $\{X\}$ also be a random vector in $\mathbb{R}^d$. Then, $X_n\xrightarrow{p}X$ means that $\forall\epsilon>0,\ P(\|X_n-X\|>\epsilon)\to 0$ as $n\to\infty$.

For homework: Let $X_n(J)$ and $X(J)$ indicate the Jth coordinate of X_n, X respectively $(J=1,...,\ k)$. We will show that $X_n\xrightarrow{p}X$ iff $X_n(J)\xrightarrow{p}X(p)\ \forall J$.

14.1.2 Little-oh Notation

Definition 14.2 ($o(\cdot)$ notation) *Let $\{X_n\}$ and $\{Y_n\}$ be sequences in some common probability space. Then, in little-oh notation, $X_n=o(Y_n)$ means that $\frac{X_n}{Y_n}\xrightarrow{p}0$. Then, $\forall\epsilon>0\ \exists n_0=n_0(\epsilon)$ s.t. $\forall n>n_0$, $|\frac{X_n}{Y_n}|<\epsilon$.*

Definition 14.3 ($op(\cdot)$ notation) *Let $\{X_n\}$ be a sequence of random variables/vectors in some probability space, and let $\{r_n\}$ be a sequence of positive numbers. Then, $X_n=op(r_n)$ means that $\frac{X_n}{r_n}\xrightarrow{p}0$. Then, $\forall\epsilon>0$ $\exists n_0=n_0(\epsilon)$ s.t. $\forall n>n_0$, $|\frac{X_n}{r_n}|<\epsilon$ or $\frac{\|X_n\|}{Y_n}<\epsilon$.*

We can use the $o(\cdot)$ and $op(\cdot)$ to cleanly express the Weak Law of Large Numbers.

Theorem 14.4 (WLLN) *Let $\{X_n\}$ be a sequence of random variables s.t. $\mathbb{E}[X_n] = \mu \; \forall n$, $V[X_n] = \sigma^2 < \infty$, and $cov(x_n,\ x_{n'}) = 0 \; \forall n \neq n'$. If $\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 \to 0$, then $\frac{1}{n}\sum_{i=1}^n X_n \xrightarrow{p} 0$.*

Proof: Let $S_n = \sum_{i=1}^n X_n$. By Chebyshev's inequality, $\forall \epsilon > 0$,

$$P\big(|\frac{S_n}{n} - \mu|\big) \leq \frac{V[\frac{S_n}{n}]}{\epsilon^2} = \frac{1}{n^2}\frac{1}{\epsilon}\sum_{i=1}^n \sigma_i^2 \to 0$$

■

Then, we can say that $\frac{S_n}{n} - \mu = op(1) \iff \frac{S_n}{n} = \mu + op(1)$. We can view $op(1)$ as the random fluctuations about the mean, which converge to 0.

This notation says nothing about rates. In fact, asymptotics are hidden by the notation. For example:

$$X_n = op(1) \;\&\; Y_n = op(1) \implies X_n + Y_n = op(1)$$
$$X_n = op(r_n) \;\&\; Y_n = op(r_n) \implies X_n + Y_n = op(r_n)$$

Similarly, $X_n = op(r_n) \implies KX_n = op(r_n) \; \forall K \in \mathbb{R}$. We can also extend $op(\cdot)$ notation to include $X_n = op(Y_n)$, which signifies $\frac{X_n}{Y_n} = op(1)$. Again, this tells us nothing about the asymptotics.

14.1.3 Taylor Expansions

Let a function f have d derivatives at the point θ_0. Also suppose that $X_n = \theta_0 + op(1)$ (i.e. $X_n \xrightarrow{p} \theta_0$). Then there exists a sequence $\{Y_n\}$ s.t. $Y_n = op(1)$ s.t.

$$f(X_n) = f(\theta_0) + (X_n - \theta_0)f'(\theta_0) + ... + \frac{(X_n - \theta_0)^d}{d!}\big[f^{(d)}(\theta_0) + Y_n\big]$$

In particular,

$$\frac{(X_n - \theta_0)^d}{d!}\big[f^{(d)}(\theta_0) + Y_n\big] = \frac{(X_n - \theta_0)^d}{d!}f^{(d)}(\theta_0) + op\big((X_n - \theta_0)^d\big)$$

where $op\big((X_n - \theta_0)^d\big) = op(1)$; recall that $X_n op(1) = op(X_n)$.

Proof: By the Taylor theorem, $\forall \epsilon > 0$, $\exists \delta = \delta(\epsilon)$ s.t. $\|X_n(\omega) - \theta_0| < \delta \implies {}_n(\omega) < \epsilon \; \forall \omega \in A$ s.t. $P(A^c) = 0$. Some details are omitted. ■

Proof: In more detail, $\{|X_n - \theta_0|\delta\} \implies \{|Y_n| < \epsilon\}$. Since $X_n \xrightarrow{p} \theta_0$, $\exists n_0 = n_0(\delta,\ \epsilon)$ s.t. $\forall n > n_0$, $P(|X_n - \theta_0| > \delta) < \epsilon$, so $\forall n > n_0$

$$P(|Y_n| < \epsilon) \geq P(|X_n - \theta_0| < \delta) > 1 - \epsilon$$

Then, $Y_n = op(1)$ ■

We can take this into multiple dimensions. Let $f : \mathbb{R}^d \mapsto \mathbb{R}$ be twice-differentiable at θ_0, and let $\{X_n\}$ be a sequence of random vectors in $\mathbb{R}^d$ such that $X_n \xrightarrow{p} \theta_0$. Then, express

$$f(X_n) = f(\theta_0) + \nabla f^T(\theta_0)(X_n - \theta_0) + \frac{1}{2}(X_n - \theta_0)^T\nabla^2 f(\theta_0)(X_n - \theta_0) + Y_n$$

In the above, $Y_n = op(1)$. In fact, $Y_n = op(\|X_n - \theta_0\|^2)$. See once again that $op(\cdot)$ notation hides rates from us.

14.1.4 Weak Law of Large Numbers

We can now prove a version of the weak law of large numbers with weaker requirements; we drop the finite variance requirement.

Theorem 14.5 (WLLN) *Let $\{X_n\}$ be independent and identically distributed random variables s.t. $\mu = \mathbb{E}[X_1] < \infty$. Let $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then $\frac{S_n}{n} \xrightarrow{p} \mu$.*

Proof: We use the truncation technique. Let $t \in (0,\ \infty)$. Let

$$X_{t,\ k} = X_k \mathbb{I}_{\{|X_k|<t\}}$$

$$Y_{t,\ k} = X_k \mathbb{I}_{\{|X_k|>t\}}$$

Then, $X_k = X_{t,\ k} + Y_{t,\ k}$, so $\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_{t,\ k} + \frac{1}{n}\sum_{i=1}^{n} Y_{t,\ k}$.

Next, let $U_{t,\ k} = \frac{1}{n}\sum_{i=1}^{n} X_{t,\ k}$ and $V_{t,\ k} = \frac{1}{n}\sum_{i=1}^{n} Y_{t,\ k}$.

Now, we have $\mathbb{E}[|V_{k,\ t}|] \leq \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[|Y_{k,\ t}|] = \mathbb{E}[X_k \mathbb{I}_{\{|X_1|<t\}}]$

By the Dominated Convergence Theorem, $\mathbb{E}[X_k \mathbb{I}_{\{|X_1|<t\}}] \to 0$ as $t \to \infty$.

We fix $\epsilon \in (0,\ 1)$ and $\delta \in (0,\ 1)$. We choose t sufficiently large to bound the previous expectation; choose $t = t(\epsilon,\ \delta)$ s.t. $\mathbb{E}[X_k \mathbb{I}_{\{|X_1|<t\}}] \leq \frac{2\delta}{6}$. Let $\mu_t = \mathbb{E}[X_{1,\ t}]$. Then,

$$|\mu_t - \mu| \leq |\mathbb{E}[Y_{1,\ t}]| \leq \mathbb{E}[|Y_{1,\ t}|] < \frac{\epsilon\delta}{6} < \frac{\epsilon}{3}$$

We now let $B_n = \{|U_{n,\ t} - \mu_t| > \frac{\epsilon}{3}\}$ and $C_n = \{|V_{n,\ t}| > \frac{\epsilon}{3}\}$.

To bound $P(B_n)$, we can use the weaker WLLN proved earlier; $\mathbb{E}[X_{k,\ t}^2] < t^2\ \forall k$. There exists $n_0 = n_0(\epsilon,\ \delta)$ s.t. $P(B_n) < \frac{\delta}{2}$.

To bound $P(C_n)$, we use Markov's inequality.

$$P(C_n) \leq \frac{3\mathbb{E}[|V_{n,\ t}|]}{\epsilon} \leq \frac{3\mathbb{E}[|V_{1,\ t}|]}{\epsilon} \leq \frac{\delta}{2}$$

Now, $B_n^c \cap C_n^c$ is a good set that has probability at least $1-\delta$ by the Union Bound: $(B_n^c \cap C_n^c) = (B_n \cup C_n)^c$. We can set $|U_{n,\ t} - \mu_t| \leq \frac{\epsilon}{3}$ and $|V_{n,\ t}| \leq \frac{\epsilon}{3}$.

Now, we can decompose $|\frac{S_n}{n} - \mu|$ as

$$|\frac{S_n}{n} - \mu| \leq |U_{n,\ t} - \mu_t| + |\mu_t - \mu| + |V_{n,\ t}| \leq \epsilon$$

Each of these terms has been bounded by $\frac{\epsilon}{3}$, so $P(|\frac{S_n}{n} - \mu| > \epsilon) \leq P(B_n \cup C_n) \leq \delta$. ■

14.2 Almost Sure Convergence

For a sequence to converge almost surely, it can only violate $|X_n - X| < \epsilon$ finitely many times. In other words, $P(|X_n - X| > \epsilon \text{infinitely often}) = 0$.

Definition 14.6 (Almost Sure Convergence) *Let $(\Omega,\ \mathcal{F},\ P)$ be a probability space, let $\{X_n\}$ be a sequence of random variables in that space, and let X be another random variable. Then, X_n is said to converge almost surely to X (denoted $X_n \overset{a.s.}{\to} X$) if any of the following equivalent conditions hold:*

- $P(|X_n - X| > \epsilon \text{infinitely often}) = 0$
- $P\big(\{\omega :\ \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\big) = 1$
- $P(\limsup_n A_{n,\ \omega}) = 0$ *where* $A_{n,\ \omega} = \{|X_n' - X| > \epsilon\}$

We note that this includes no guarantees about uniform convergence. For some $\omega_1 \neq \omega_2$, $X_n(\omega_1) \to X(\omega_1)$ at a different rate than $X_n(\omega_2) \to X(\omega_2)$.

14.3 Next Time: L_p Convergence

Let $p \geq 1$. Then, $X_n \overset{L_p}{\to} X$ if

$$\|X_n - X\|_p = (\mathbb{E}[|X_n - X|^p])^{\frac{1}{p}} \to 0 \text{ as } n \to \infty$$

For the special case where $p = 2$, we refer to this as convergence in quadratic mean. $X_n \overset{L_2}{\to} \mu$ iff $V[X_n] \to 0$ and $\mathbb{E}[X_n] \to \mu$.