Lecture Notes 12: Stochastic Convergence

Professor: Zhihua Zhang *Scribe:*

12 Stochastic Convergence

Definition 12.1. *A sequence $X_1,\ X_2,\ldots$ of random variables is said to converge in distribution, or converge weakly, or converge in law to a random variable X if*

$$\lim_{n\to\infty} F_n(x) = F(x),$$

for every number $x \in \boldsymbol{R}$ at which F is continuous. Here F_n and F are the cumulative distribution functions of random variables X_n and X, respectively.

We denote convergence in distribution as $X_n \rightsquigarrow X$.

Definition 12.2. *A sequence X_n of random variables converges in probability towards the random variable X if for all $\varepsilon > 0$*

$$\lim_{n\to\infty} \Pr\big(|X_n - X| \geq \varepsilon\big) = 0.$$

Formally, pick any $\varepsilon > 0$ and any $\delta > 0$. Let P_n be the probability that X_n is outside the ball of radius ε centered at X. Then for X_n to converge in probability to X there should exist a number N (which will depend on ε and δ) such that for all $n \geq N$ the probability Pn is less than δ.

Convergence in probability is denoted by adding the letter p over an arrow indicating convergence, or using the "plim" probability limit operator:

$$X_n \xrightarrow{p} X, \quad X_n \xrightarrow{P} X, \quad \operatorname*{plim}_{n\to\infty} X_n = X.$$

For random elements X_n on a separable metric space (S, d), convergence in probability is defined similarly by

$$\forall\varepsilon > 0, \Pr\big(d(X_n, X) \geq \varepsilon\big) \to 0.$$

Definition 12.3. *To say that the sequence X_n converges almost surely or almost everywhere or with probability 1 or strongly towards X means that*

$$\Pr\Big(\lim_{n\to\infty} X_n = X\Big) = 1.$$

Almost sure convergence is often denoted by adding the letters a.s. over an arrow indicating convergence:

$$X_n \xrightarrow{\text{a.s.}} X.$$

Using the probability space $(\Omega, \mathcal{F}, Pr)$ and the concept of the random variable as a function from Ω to R, for generic random elements X_n on a metric space (S, d), convergence almost surely is defined similarly:

$$\Pr\Big(\omega \in \Omega : \, d\big(X_n(\omega), X(\omega)\big) \underset{n\to\infty}{\longrightarrow} 0\Big) = 1$$

Definition 12.4. *Given a real number $r \geq 1$, we say that the sequence X_n converges in the r-th mean (or in the Lr-norm) towards the random variable X, if the r-th absolute moments $E(|X_n|^r)$ and $E(|X|^r)$ of X_n and X exist, and*

$$\lim_{n\to\infty} \mathrm{E}\left(|X_n - X|^r\right) = 0,$$

where the operator E denotes the expected value. Convergence in r-th mean tells us that the expectation of the r-th power of the difference between X_n and X converges to zero.

This type of convergence is often denoted by adding the letter L_r over an arrow indicating convergence:

$$X_n \xrightarrow{L_r} X.$$

The most important cases of convergence in r-th mean are:

1. When X_n converges in r-th mean to X for $r = 1$, we say that X_n converges in mean to X.

2. When X_n converges in r-th mean to X for $r = 2$, we say that X_n converges in mean square to X.

Example 12.1. *Let $X_n \sim N(0, \frac{1}{n})$ and let F be distribution function for a point mass at 0, namely, $P(X = 0) = 1$. Let $Z \sim N(0, 1)$,*

$$\begin{aligned} F_n(x) &= Pr(X_n < x) \\ &= Pr(\sqrt{n}X_n < \sqrt{n}x) \\ &= Pr(Z < \sqrt{n}x) \end{aligned}$$

Thus,

1. *For $x < 0$, $\sqrt{n}x \longrightarrow -\infty$, $\lim_{n\to\infty} F_n(x) = 0$*

2. *For $x > 0$, $\sqrt{n}x \longrightarrow \infty$, $\lim_{n\to\infty} F_n(x) = 1$*

So, for $x \neq 0$, $\lim_{n\to\infty} F_n(x) = F(x)$, $X_n \rightsquigarrow 0$. And since $E(X) = 0$, by Chebyshev's inequality,

$$Pr(|X_n| \geq \varepsilon) \leq \frac{E(X_n^2)}{\varepsilon} = \frac{1}{n\varepsilon} \longrightarrow 0$$

So, $X_n \xrightarrow{p} 0$. Note here we use notation $X_n \rightsquigarrow c$ or $X_n \xrightarrow{p} c$ means X_n convergence to a point mass distribution $P(X = c) = 1$.

Lemma 12.1 (portmanteau lemma)**.** *The following statement are equivalent:*

1. $\lim Pr(X_n \in A) = Pr(X \in A)$ *for all continuity sets* A *of random variable* X.

2. $Eg(X_n) \longrightarrow Eg(X)$ *for all bounded, continuous functions* g.

3. $Eg(X_n) \longrightarrow Eg(X)$ *for all bounded, Lipschitz functions* g. $(\|g(x)-g(y)\| \leq L\|x-y\|)$

The portmanteau lemma provides several equivalent definitions of convergence in distribution. Although these definitions are less intuitive, they are used to prove a number of statistical theorems.

Theorem 12.1 (Continuous mapping)**.** *Let* X_n*,* X *be random elements defined on a metric space* S*. Suppose a function* $g : S \longrightarrow S$ *(where* S *is another metric space) has the set of discontinuity points* D_g *such that* $Pr[X \in D_g] = 0$*. Then*

$$\begin{array}{rcl} X_n \rightsquigarrow X & \Rightarrow & g(X_n) \rightsquigarrow g(X); \\ X_n \xrightarrow{p} X & \Rightarrow & g(X_n) \xrightarrow{p} g(X); \\ X_n \xrightarrow{as} X & \Rightarrow & g(X_n) \xrightarrow{as} g(X). \end{array}$$

Theorem 12.2. *Let* X_n *and* Y_n *be random variables, then*

1. $X_n \xrightarrow{as} X \quad \Rightarrow \quad X_n \xrightarrow{p} X$

2. $X_n \xrightarrow{L_r} X \quad \Rightarrow \quad X_n \xrightarrow{p} X$

3. $X_n \xrightarrow{p} X \quad \Rightarrow \quad X_n \rightsquigarrow X$

4. $X_n \xrightarrow{p} c \quad \Leftrightarrow \quad X \rightsquigarrow c$

5. $X_n \rightsquigarrow X,\ d(X_n, Y_n) \xrightarrow{p} 0 \quad \Rightarrow \quad Y_n \rightsquigarrow X$

6. $X_n \rightsquigarrow X,\ Y_n \xrightarrow{p} c \quad \Rightarrow \quad (X_n, Y_n) \rightsquigarrow (X, c)$

7. $X_n \xrightarrow{p} X,\ Y_n \xrightarrow{p} Y \quad \Rightarrow \quad (X_n, Y_n) \xrightarrow{p} (X, Y),\ X_n + Y_n \xrightarrow{p} X + Y$

Here we give some counter examples to illustrate theorem 12.2.

Example 12.2. *Consider the probability space* $([0,1], B_{[0,1]}, p)$*.* $B_{[0,1]}$ *is* σ*-field of Borel set and* p *is Lebesgue measure. Let* $X = 0$*, for any positive integer* n*, there exist integers* m *and* k *such that* $n = 2^m - 2 + k$ *and* $0 \leq k \leq 2^{m+1}$*. Define*

$$X_n(w) = \begin{cases} 1 & \frac{k}{2^m} \leq w \leq \frac{k+1}{2^m} \\ 0 & otherwise \end{cases}$$

So, we have

$$\begin{array}{rcl} P(|X_n - X| \geq \varepsilon) & \leq & P(\{w : \dfrac{k}{2^m} \leq w \leq \dfrac{k+1}{2^m}\}) \\ & = & \dfrac{1}{2^m} \\ & \to & 0 \end{array}$$

That is, $X_n \xrightarrow{p} X$.

On the other hand, for any fixed $w \in [0,1]$ *and* m*, there exist* $k, 1 \le k \le 2^m$ *such that* $\frac{k-1}{2^m} \le w \le \frac{k}{2^m}$*. So we have a sequence of* $X_{n_m}(w) = 1$*. So* $X_n(w) \xrightarrow{as} X$ *does **not** hold.*

Example 12.3. *Let* $X = 0$*, define*

$$X_n(w) = \begin{cases} 0 & \frac{1}{n} < w \le 1 \\ e^n & 0 \le w \le \frac{1}{n} \end{cases}$$

Then for any $\varepsilon \in (0,1)$*,*

$$P(|X_n - X| \ge \varepsilon) \;=\; P(|X_n| \ne 0) = \frac{1}{n} \to 0$$

That is, $X_n \xrightarrow{p} X$.

But for $p > 0$*,* $E|X_n - X|^p = E|X|^p = \frac{e^{np}}{n} \to \infty$*. So* $X_n \xrightarrow{p} X$ *does **not** imply* $X_n \xrightarrow{L_p} X$.

Example 12.4. *Define*

$$X(w) = \begin{cases} 1 & 0 \le w \le \frac{1}{2} \\ 0 & \frac{1}{2} < w \le 1 \end{cases}$$

and

$$X_n(w) = \begin{cases} 0 & 0 \le w \le \frac{1}{2} \\ 1 & \frac{1}{2} < w \le 1 \end{cases}$$

So for any t*,*

$$P(X < t) = P(X_n \le t) = \begin{cases} 0 & t \ge 1 \\ \frac{1}{2} & 0 \le t < 1 \\ 0 & t \le 0 \end{cases}$$

That is $X_n \rightsquigarrow X$*. But* $|X_n - X| = 1, d(|X_n - X| > \varepsilon) = 1$*. So* $X_n \rightsquigarrow X$ *does **not** imply* $X_n \xrightarrow{p} X$.

Example 12.5. *Let* $g(t) = 1 - \mathbf{1}_{\{0\}}(t)$*,* $X = 0$*,* $X_n = \frac{1}{n}$*. Then* $X_n \xrightarrow{p} X$*. But* $g(X_n) = 0, g(X) = 1$*. So,* $g(X_n) \xrightarrow{p} g(X)$ *does **not** hold.*

Lemma 12.2 (Slutsky's theorem)**.** *Let* X_n*,* Y_n *be sequences of random variables. If* X_n *converges in distribution to a random element* X *and* Y_n *converges in probability to a constant* c*, then*

1. $X_n + Y_n \rightsquigarrow X + c$;
2. $X_n Y_n \rightsquigarrow cX$;
3. $Y_n^{-1} X_n \rightsquigarrow c^{-1} X$, $c \ne 0$

Proposition 12.1. *For all* $t, x \ge 0$*,*

$$\lim_{\lambda \to \infty} e^{-\lambda t} \sum_{k \le \lambda x} \frac{(\lambda t)^k}{k!} = \mathbf{1}_{[0,x]}(t)$$

Proof. Assume $X \sim Possion(\lambda t)$, so the proposition is equal to

$$\lim_{\lambda \to \infty} P(X \leq \lambda x) = \mathbf{1}_{[0,x]}(t)$$

We have $E(X) = \lambda t$, $Var(X) = \lambda t$. So for any $t \leq x$,

$$\begin{aligned} P(X \leq \lambda x) &= P(X - \lambda t \leq \lambda(x-t)) \\ &= 1 - P(X - \lambda t > \lambda(x-t)) \\ &\geq 1 - P(|X - \lambda t| > \lambda(x-t)) \\ &\geq 1 - \frac{\lambda t}{\lambda^2 (t-x)^2} \\ &\to 1 \end{aligned}$$

for $t > x$,

$$\begin{aligned} P(X \leq \lambda x) &= P(\lambda t - X > \lambda(t-x)) \\ &\leq P(|\lambda t - X| > \lambda(t-x)) \\ &\leq \frac{\lambda t}{\lambda^2 (t-x)^2} \\ &\to 0 \end{aligned}$$

So,

$$\lim_{\lambda \to \infty} P(X \leq \lambda x) = \mathbf{1}_{[0,x]}(t)$$

□

Theorem 12.3. *Let $X_1, X_2, \cdots$ be iid samples. Let $\mu = E(X_n)$ and $\sigma^2 = Var(X_n)$. Define the sample average*

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

then $\overline{X}_n \xrightarrow{p} \mu$.

Proof.

$$P(|\overline{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0$$

□

Theorem 12.4 (Central limit theorems)**.** *Let $X_1, X_2, \cdots, X_n$ be iid with mean μ and variance Σ. Let $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, then*

$$Z_n = \Sigma^{-\frac{1}{2}} (\overline{X} - \mu) \rightsquigarrow N(0, I)$$

Example 12.6 (t-statistic)**.** *Let $X_1, X_2, \cdots, X_n$ be iid with mean 0 and finite variance. The t-statistic $\frac{\sqrt{n} X_n}{S_n}$, where $S_n = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$ is asymptotically standard normal.*

Proof.
$$S_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2\right) \xrightarrow{p} E(X_i^2) - E(X_i)^2 = Var(X_i)$$

So, $S_n \xrightarrow{p} \sqrt{Var(X_i)}$, hence, $S_n \rightsquigarrow \sqrt{Var(X_i)}$.
And $\sqrt{n}\overline{X}_n \rightsquigarrow N(0, Var(X_i))$, according to Slutslay's lemma, $\frac{\sqrt{n}\overline{X}_n}{S_n} \rightsquigarrow N(0,1)$ $\square$

Theorem 12.5 (The delta method)**.** *Suppose that X_n is a sequence of random variables such that $\sqrt{n}(X_n - \mu) \rightsquigarrow N(0, \Sigma)$.. Let $g : R^k \to R$ and let $\bigtriangledown g(y)$ denotes the gradient and $\bigtriangledown u$ denotes $\bigtriangledown g(y)$ evaluated at u. Assume that the elements of $\bigtriangledown u$ are nonzero. Then*

$$\sqrt{n}(g(X_n) - g(\mu)) \rightsquigarrow N(0, \bigtriangledown u^T \Sigma \bigtriangledown u)$$

Theorem 12.6. *Let ϕ be a map defined on a subset of R^k and differentiable at θ. Let T_n be random vectors taking their values in the domain of ϕ. If $r_n(T_n - \theta) \rightsquigarrow T$ for numbers $r_n \to \infty$, then $r_n(\phi(T_n) - \phi(\theta)) \rightsquigarrow \phi'_\theta(T)$. Moreover, the difference between $r_n(\phi(T_n) - \phi(\theta))$ and $\phi'_\theta(r_n(T_n - \theta))$ convergence to 0 with probability 1.*