# 14.381 Recitation 1

Joonhwan Lee

September 14, 2013

## 1 Modes of Convergence

### 1.1 Almost Sure Convergence

**Definition 1.** Let $\{X_{n\}}$ be a sequence of random variables. Then $X_n \overset{a.s}{\to} X$ if

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

*Remark.* Almost sure convergence means a sequence of functions $X_n(\omega)$ converges point-wise to $X(\omega)$ except for some measure zero set.

**Theorem.** $X_n \overset{a.s}{\to} X$ *if and only if*

$$\forall\epsilon > 0, \quad \lim_{m\to\infty} P(|X_k - X| \le \epsilon \; \forall k \ge m) = 1$$

*Proof.* ($\Rightarrow$) Let $\Omega_0 = \{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$. Suppose $P(\Omega_0) = 1$. Let $\epsilon > 0$ be given. Let $A_m = \cap_{k=m}^{\infty}\{|X_k - X| \le \epsilon\}$. Then $A_m \subset A_{m+1}$ $\forall m$ and $\lim_{m\to\infty} P(A_m) = P(\cup_{m=1}^{\infty} A_m)$ by continuity of probability measure. For each $\omega_0$, there exists $m(\omega_0)$ such that $|X_k(\omega_0) - X(\omega_0)| \le \epsilon$ for all $k \ge m(\omega_0)$. Therefore, $\forall\omega_0 \in \Omega_0$, $\omega_0 \in A_m$ for some $m$ and we can conclude that $\Omega_0 \subset \cup_{m=1}^{\infty} A_m$ and $1 = P(\Omega_0) \le P(\cup_{m=1}^{\infty} A_m) = \lim_{m\to\infty} P(A_m) = 1$.
($\Leftarrow$) Let $A_m(\frac{1}{n})$ be a set defined above with given $\epsilon = \frac{1}{n}$. Suppose that $\lim_{m\to\infty} P(A_m(\frac{1}{n})) = 1$ for all $n$. By continuity, we have $P(A(\frac{1}{n})) = 1$ where $A(\frac{1}{n}) = \cup_{m=1}^{\infty} A_m(\frac{1}{n})$. Let $A = \cap_{n=1}^{\infty} A(\frac{1}{n})$. Then by the continuity, $P(A) = 1$ because $A(\frac{1}{n})$'s are monotone decreasing sequence of sets. Therefore, $\forall\omega_0 \in A$ and $\forall\epsilon > 0$, there exists $M$ such that $|X_m(\omega_0) - X(\omega_0)| \le \epsilon$ for all $m \ge M$. We conclude that $P(\{\omega : \lim_{m\to\infty} X_m(\omega) = X(\omega)\}) = 1$. $\square$

### 1.2 $L^p$-convergence

**Definition 2.** Let $p \in (1, \infty)$ and $\{X_n\}$ be a sequence of random variables. Then,$X_n \overset{L^p}{\to} X$ if

$$E|X_n|^p < \infty,\, E|X|^p \text{ and } \lim_{n\to\infty} E|X_n - X|^p = 0$$

*Remark.* We often use the case when $p = 2$ because $L^2$-space is an inner product space and many interesting results can be derived.

### 1.3 Convergence in probability

**Definition 3.** $X_n \overset{p}{\to} X$ if

$$\lim_{n\to\infty} P(|X_n - X| \le \epsilon) = 1 \quad \forall\epsilon > 0$$

## 1.4 Convergence in distribution

**Definition 4.** Let $\{F_{X_n}\}$ and $F_X$ be distribution functions of random variables $\{X_n\}$ and $X$. $X_n \Rightarrow X$ if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \forall x \in C(F_X)$$

where $C(F_X)$ denotes the set of all points where $F_X$ is continuous.

## 1.5 Relations of modes of convergence

1. $X_n \overset{a.s}{\to} X$ implies $X_n \overset{p}{\to} X$ (Obvious by the Theorem above)

2. $X_n \overset{L^p}{\to} X$ implies $X_n \overset{p}{\to} X$ (Use Chebyshev inequality or its variants)

3. $X_n \overset{p}{\to} X$ implies $X_n \Rightarrow X$
   *(proof)* Let $x \in C(F_X)$ and $\epsilon > 0$ be given. We have,

$$\begin{aligned} F_{X_n}(x) &= P[X_n \le x] = P[\{X_n \le x\} \cap \{|X_n - X| < \epsilon\}] + P[\{X_n \le x\} \cap \{|X_n - X| \ge \epsilon\}] \\ &\le P[X \le x + \epsilon] + P[|X_n - X| \ge \epsilon] \end{aligned}$$

   Hence, $\limsup F_{X_n}(x) \le F_X(x+\epsilon)$ because the latter term converges to 0 by in probability convergence. Similarly,

$$\begin{aligned} 1 - F_{X_n}(x) &= P[X_n \ge x] = P[\{X_n \ge x\} \cap \{|X_n - X| < \epsilon\}] + P[\{X_n \ge x\} \cap \{|X_n - X| \ge \epsilon\}] \\ &\le P[X \ge x - \epsilon] + P[|X_n - X| \ge \epsilon] \end{aligned}$$

   Hence, $\liminf F_{X_n}(x) \ge F_X(x-\epsilon)$ and we conclude that $\lim F_{X_n}(x) = F_X(x)$ by continuity of $F_X$ at $x$.

4. $X_n \overset{a.s}{\to} X$ does not necessarily imply $X_n \overset{L^p}{\to} X$.
   *(counterexample)* Let $X_n$ be random variables defined on $(\Omega, \mathcal{F}, P)$ where $\Omega = (0,1)$, $\mathcal{F}$ is Borel sets on $(0,1)$ and $P$ is the Lebesgue measure. Let $X_n = n^{\frac{1}{p}} 1_{\{0<\omega<\frac{1}{n}\}}(\omega)$. Then $\forall \omega \in (0,1)$ there exists $N$ such that $X_n(\omega) = 0$ for all $n > N$. Thus, $X_n$ converges to 0 everywhere and obviously $X_n \overset{a.s.}{\to} 0$. However, $E|X_n|^p = \int_0^{1/n} n d\omega = 1$ for all $n$ and we can see that $X_n$ does not converge in $L^p$ to 0.

5. $X_n \overset{L^p}{\to} X$ does not necessarily imply $X_n \overset{a.s}{\to} X$.
   *(counterexample)* Let $Y_{k,j} = 1_{\{\frac{j-1}{k}<\omega<\frac{j}{k}\}}$ where $k \ge 1$, $1 \le j \le k$. Let $X_n$ be the lexicographic ordering of $Y_{k,j}$. That is, $X_1 = Y_{1,1}$, $X_2 = Y_{2,1}$, $X_3 = Y_{2,2}$, $X_4 = Y_{3,1}$ and so on. Let $k_n$ be the corresponding value of $k$ for given $X_n$. Then it is easy to see that $E|X_n|^p = \frac{1}{k_n}$.Thus, $X_n \overset{L^p}{\to} 0$. However, given any $\omega \in (0,1)$, $X_n$ does not converge to 0 because $X_n(\omega) = 1$ infinitely often.

6. Two counterexamples above directly imply that the converses of both (1) and (2) are not true.

7. $X_n \Rightarrow X$ does not necessarily imply $X_n \overset{p}{\to} X$
   (counterexample) Let our probability space have sample space $\Omega = (-\frac{1}{2}, \frac{1}{2})$ equipped with Lebesgue measure. Let $X(\omega) = \omega$ and $X_n = -X$. It is easy to show that both $X$ and $-X$ has uniform distribution on $(-\frac{1}{2}, \frac{1}{2})$ and therefore $X_n \Rightarrow X$. However, given $\epsilon = \frac{1}{2}$, $P(|X_n - X| < \epsilon) = P(\{\omega : 2|\omega| > \frac{1}{2}\}) = \frac{1}{2}$ for all $n$. We conclude that $X_n$ does not converge in probability to $X$.

# 2 Limit Theorems and Delta Method

## 2.1 Law of Large Numbers

**Theorem.** *Let $\{X_n\}$ be a sequence of independent and identically distributed random variables. Suppose $E|X_1| < \infty$. Then we have*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s} \mu$$

*where $\mu = E[X_1]$.*

*Remark.* This is known as Strong law of large numbers. From the first section, we can see that it directly implies the weak law of large numbers. i.e.

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$$

We can considerably relax the i.i.d. condition and there are many versions of both SLLN and WLLN. However, some kind of independence structure is essential for any law of large numbers.

## 2.2 Central Limit Theorem

**Theorem.** *Let $\{X_n\}$ be a sequence of independent and identically distributed random variables. Suppose $E|X_1|^2 < \infty$. Let $\overline{X}_n = \frac{1}{n}\sum_i X_i$, $\mu = E[X_1]$ and $\sigma^2 = E[X_1^2] - E[X_1]^2$. Then we have*

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

*where $N(0,1)$ denotes a random variable whose distribution function is*

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx$$

## 2.3 Delta Method

**Lemma.** *(Taylor Expansion) Let $g(x)$ be a $r$-times continuously differentiable function in the neighborhood of$a$. Then we have*

$$g(x) = \sum_{n=0}^{r} \frac{g^{(n)}(a)(x-a)^n}{n!} + R_r(x-a)$$

*where $R_r(h) = o(h^r)$ or in other words $\lim_{h\to 0} \frac{R_r(h)}{h^r} = 0$.*

**Definition.** For a sequence of random variable $\{X_n\}$ and a positive sequence $a_n$, $X_n = O_p(a_n)$ if for any $\epsilon > 0$, there exists $C$ and $N$ such that $P(|\frac{X_n}{a_n}| > C) < \epsilon$ for all $n > N$.

$X = o_p(a_n)$ if $\frac{X}{a_n} \xrightarrow{p} 0$.

**Lemma.** $O_p(a_n)o_p(b_n) = o_p(a_n b_n)$

*Proof.* Let $X_n = O_p(a_n)$ and $Y_n = o_p(b_n)$. Let $\epsilon, \eta > 0$ be given. Choose $C$ such that $P(|\frac{X_n}{a_n}| > C) < \frac{\eta}{2}$. Choose $N$ such that $P(|\frac{Y_n}{b_n}| < \frac{\epsilon}{C}) < \frac{\eta}{2}$ for all $n > N$. Then we have,

$$P(|\frac{X_n Y_n}{a_n b_n}| < \epsilon) \leq P(C|\frac{Y_n}{b_n}| < \epsilon) + P(|\frac{X_n}{a_n}| > C) \leq \eta \quad \forall n > N$$

and we conclude that $\frac{X_n Y_n}{a_n b_n} \xrightarrow{p} 0$. □

**Lemma.** *If $X_n \Rightarrow X$, then $X_n = O_p(1)$.*

*Proof.* Let $\epsilon > 0$ be given. Since $X$ has a distribution function $F_X$ and $\lim_{x\to\infty} F_X(x) = 1$, there exists $C$ such that $P(|X| > C) < \frac{\epsilon}{2}$ where $F_X$ is continuous at $C$. By the convergence, we can choose $N$ such that $|F_{X_n}(C) - F_X(C)| < \frac{\epsilon}{2}$ for all $n > N$. Thus we conclude that $P(|X_n| > C) < \epsilon$ for all $n > N$. $\square$

**Theorem.** *(Delta Method) Let $\sqrt{n}(Y_n - \mu) \Rightarrow N(0, \sigma^2)$ and $g(y)$ be a continuously differentiable in the neighborhood of $\mu$ with $g'(\mu) \neq 0$. Then we have,*

$$\sqrt{n}(g(Y_n) - g(\mu)) \Rightarrow N(0, g'(\mu)^2\sigma^2)$$

*Proof.* Using Taylor expansion, we can get

$$\sqrt{n}(g(Y_n) - g(\mu)) = g'(\mu)\sqrt{n}(Y_n - \mu) + \sqrt{n}R_1(Y_n - \mu)$$

Thus, it is sufficient to show that $\sqrt{n}R_1(Y_n - \mu) = o_p(1)$. Let $m(h) = \frac{R_1(h)}{h}$. By Taylor theorem, we have $\lim_{h\to 0} m(h) = 0$. Let $\epsilon > 0$ be given. Choose $\delta > 0$ such that

$$|m(h)| < \epsilon \quad \forall h < \delta$$

We have,

$$\lim_{n\to\infty} P(|m(Y_n - \mu)| < \epsilon) \geq \lim_{n\to\infty} P(|Y_n - \mu| < \delta) = 1$$

by WLLN. Thus, $m(Y_n - \mu) = o_p(1)$ and by the lemmas above, we can get the desired result. $\square$

*Remark.* For $o_p$ and $O_p$ notations, equality (=) means logical implication from left to right. So one can say $o(1) = o_p(1)$ for example. You can derive other elementary results using $o_p$ and $O_p$ notations easily.

# 3 From Class

## 3.1 Transformations

**Theorem.** *(1-to-1 transformation) Let $X$ be a continuous random variable with distribution function $F_X(x)$ defined on $\mathcal{X}$ and let $f_X(x) = F'(x)$. Let $g(x)$ be a continuously differentiable strictly increasing function. Let $Y = g(X)$ and $\mathcal{Y} = g(\mathcal{X})$ . Then we have*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\frac{dg^{-1}(y)}{dy} & \forall y \in \mathcal{Y} \\ 0 & \textit{otherwise} \end{cases}$$

*Proof.* First note that $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$. Thus we have,

$$\begin{aligned} \frac{F_Y(y + \Delta_y) - F_Y(y)}{\Delta_y} &= \frac{F_X(g^{-1}(y + \Delta_y)) - F_X(g^{-1}(y))}{\Delta_y} \\ &= \frac{F_X(g^{-1}(y + \Delta_y)) - F_X(g^{-1}(y))}{\Delta_x} \cdot \frac{\Delta_x}{\Delta_y} \end{aligned}$$

where $\Delta_x = g^{-1}(y + \Delta_y) - g^{-1}(y)$. Taking $\Delta_y \to 0$, we get the desired result by differentiability of $g$. $\square$

*Remark.* It is obvious that you just need to change the sign for a strictly decreasing function.

**Theorem.** *(K to 1 transformation) Suppose there exists a partition, $A_0, A_1, \ldots, A_K$ of $\mathcal{X}$ and $f_X(x)$ is continuous on each $A_i$. Let $g(x)$ be a function such that each $g_i(x) \equiv g(x)1_{\{x \in A_i\}}(x)$'s are strictly monotone on $A_i$ and continuously differentiable. Let $Y = g(X)$. Then we have*

$$f_Y(y) = \sum_{i=1}^{K} f_X(g_i^{-1}(y))|\frac{dg_i^{-1}(y)}{dy}|1_{\{y \in g(A_i)\}}(y)$$

**Example.** ($\chi^2(1)$ random variable) Let $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$. Let $Y = X^2$. Then we have

$$f_Y(y) = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} + f_X(\sqrt{y})\frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}} \exp(-\frac{1}{2}y), \quad y \in (0, \infty)$$

and get the $\chi^2(1)$ density.

(log-normal variable) Now let $Y = \exp(X)$. Then we have

$$f_Y(y) = f_X(\log y)\frac{1}{y} = \frac{1}{y\sqrt{2\pi}} \exp(-\frac{1}{2}(\log y)^2), \quad y \in (0, \infty)$$

and get the standard log-normal density.

## 3.2 Short Note on Order Statistics

**Definition.** Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample. (which means $X_i$'s are i.i.d random variables) Order statistic $(X^{(1)}, X^{(2)}, \ldots X^{(n)})$ is the ascending ordering of the random sample. i.e. $X^{(1)} \leq X^{(2)} \leq X^{(3)} \leq \ldots \leq X^{(n)}$. $r$th order statistic is $X^{(r)}$.

*Remark.* Note that the order statistic is a significant data reduction. There are $n!$ different random samples that can generate the same order statistic. In general, it is often the most we can get without losing information. For parametric families, we can do much better than order statistic. (See sufficiency part.)

**Theorem.** *Let $X_i$ has distribution function $F(x)$ and $X_i$'s are continuous random variable. Then we have*

$$f_{X^{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} f_X(x)F(x)^{r-1}(1 - F(x))^{n-r}$$

*Proof.* We consider $F_{X^{(r)}}(x + \Delta) - F_{X^{(r)}}(x) = P(X^{(r)} \in (x, x + \Delta])$. Note that it is equal to the probability that $r - 1$ $X_i$'s are smaller than $x$, 1 $X_i$ is in $(x, x + \Delta]$ and $n - r$ $X_i$'s are greater than $x + \Delta$. That is,

$$P(X^{(r)} \in (x, x + \Delta]) = \frac{n!}{(r-1)!1!(n-r)!} F(x)^{r-1}(F(x + \Delta) - F(x))(1 - F(x + \Delta))^{n-r}$$

Therefore, dividing by $\Delta$ and taking $\Delta \to 0$, we can have the desired result. □

**Example.** Let $X_1, X_2, \ldots, X_n$ be a random sample from $U[0, \lambda n]$. Then the density function of $X^{(1)}$ is

$$f_{X^{(1)}}(x) = \frac{n!}{(n-1)!} f_X(x)(1 - F(x))^n = n \cdot \frac{1}{\lambda n}(1 - \frac{x}{\lambda n})^n = (1 - \frac{x}{\lambda n})^n$$

Note that as $n \to \infty$, the density function converges to $\frac{1}{\lambda} \exp(-\frac{x}{\lambda})$. Think of each $X_i$ to be life span of an independent component which can fail at any time before $\lambda n$. Then it makes sense to think $X^{(1)}$ as the life span(or time before failure) of the machine built with those components. This explains why exponential distribution is often used for duration models.