STAT 512: Statistical Inference **Autumn 2022**

Quiz Session 3: Convergence of Random Variables and Hypergeometric Distribution

Yikun Zhang *October 19, 2022*

Problem 1. *Suppose that a sequence of random variables* $\{X_n\}_{n=1}^{\infty}$ *and* X *are defined on the same probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*. Consider the following types of convergence:*

(i) X_n *converge to* X *almost surely (or with probability 1) as* $n \to \infty$, i.e.,

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

(ii) X_n *converge to* X *in probability as* $n \to \infty$, i.e., *for any* $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon\right) = 0.$$

(iii) X_n *converge to* X *in* L^p*-norm as* $n \to \infty$, i.e.,

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

provided that the p-th absolute moments $\mathbb{E}|X_n|^p$ *and* $E|X|^p$ *of* $X_n, n = 1, 2, ...$ *and* X *exist. When* $p = 1$*, it reduces to the convergence in (absolute) expectation.*

(iv) X_n *converge to* X *in distribution as* $n \to \infty$, i.e., *the cumulative distribution functions (CDFs)* $F_1, F_2, ...$ *of* $X_1, X_2, ...$ *converge to the CDF* F *of* X *in the sense that*

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

for every $x \in \mathbb{R}$ *at which* F *is continuous.*

For each ordered pair of conditions (there are 12 ordered pairs), determine whether the first condition in the pair implies the other. Give a proof if yes. Give a counterexample if not.

Solution. The true implications between different types of convergence are

$$\begin{array}{ccccc} (i) & \Rightarrow & (ii) & \Rightarrow & (iv) \\ & & \Uparrow & & \\ & & (iii) & & \end{array}.$$

We prove these implications and provide counterexamples for others.

(i) $\Rightarrow$ (ii): If $X_n \to X$ almost surely, then $\mathbb{P}\left(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1$. By definition, it can also be expressed as

$$\mathbb{P}\left(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\{\omega : |X_n(\omega) - X(\omega)| \le \epsilon\}\right) = 1, \quad \text{for all } \epsilon > 0.$$

Equivalently, it indicates that $\mathbb{P}\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0$ for any $\epsilon > 0$. By the continuity of measure $\mathbb{P}$ (recall Lecture 1 notes), we know that

$$\mathbb{P}\left(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}\right) = \lim_{N\to\infty} \mathbb{P}\left(\bigcup_{n=N}^{\infty}\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0.$$

Since $\{\omega : |X_N(\omega) - X(\omega)| > \epsilon\} \subset \bigcup_{n=N}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}$,

$$\lim_{N\to\infty} \mathbb{P}\left(\{\omega : |X_N(\omega) - X(\omega)| > \epsilon\}\right) \leq \lim_{N\to\infty} \mathbb{P}\left(\bigcup_{n=N}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0$$

for any $\epsilon > 0$. Thus, X_N converge to X in probability as $N \to \infty$.

(ii) $\not\Rightarrow$ (i): **Counterexample A**: For a probability space $(\Omega, \mathcal{B}, P)$, where $\Omega = (0,1]$, $\mathcal{B}$ is the Borel set in $(0,1]$, and P is the Lebesgue measure, we consider a sequence of random variables

$$Y_{ki}(\omega) = \begin{cases} 1, & \text{if } \omega \in (\frac{i-1}{k}, \frac{i}{k}], i = 1, 2, ..., k, \\ 0, & \text{otherwise}, \end{cases} \qquad \text{for } k = 1, 2, ...$$

Let $X_1 = Y_{11}, X_2 = Y_{21}, X_3 = Y_{22}, ..., X_n = Y_{ki}, ...,$ where $n = i + \frac{k(k-1)}{2}$. On the one hand, for any fixed $\omega \in (0,1]$, there exists infinite X_{n_j}'s such that $|X_{n_j}(\omega)| > \epsilon$ for any $\epsilon \in (0,1)$, where $\{n_j\}_{j=1}^{\infty}$ is an increasing sequence of integer. On the other hand, for any fixed $\omega \in (0,1]$, there also exists infinite $X_{n'_j}$ such that $|X_{n'_j}(\omega)| \leq \epsilon$, where $\{n'_j\}_{j=1}^{\infty}$ is an increasing sequence of integer. Hence,

$$\mathbb{P}\left(\{\omega : X_n(\omega) \to 0\}\right) \neq 1,$$

showing that X_n's do not converge to $X = 0$ almost surely.
However, for any $\epsilon > 0$, $\mathbb{P}(\{\omega : |X_n(\omega)| > \epsilon\}) \leq \frac{1}{k}$. Since $n = i + \frac{k(k-1)}{2} \leq k + \frac{k(k-1)}{2}$, $k \to \infty$ whenever $n \to \infty$. Thus,

$$\lim_{n\to\infty} \mathbb{P}(\{\omega : |X_n(\omega)| > \epsilon\}) \leq \lim_{k\to\infty} \frac{1}{k} = 0,$$

which means that X_n converge to X in probability.

(i) $\not\Rightarrow$ (iii): **Counterexample B**: For a probability space $(\Omega, \mathcal{B}, P)$, where $\Omega = (0,1]$, $\mathcal{B}$ is the Borel set in $(0,1]$, and P is the Lebesgue measure, we define

$$X_n(\omega) = \begin{cases} n, & \text{if } \omega \in (0, \frac{1}{n^p}], \\ 0, & \text{otherwise}. \end{cases}$$

Thus, $\mathbb{P}(\{\omega : \lim_{n\to\infty} X_n(\omega) = 0\}) = 1$ and X_n converge to $X = 0$ (almost surely) as $n \to \infty$.
However, $\lim_{n\to\infty} \left(E|X_n|^p\right)^{\frac{1}{p}} = \left(n^p \cdot \frac{1}{n^p}\right)^{\frac{1}{p}} = 1 \neq 0$, so X_n does not converge to X in L^p.

(iii) $\not\Rightarrow$ (i): Consider the Counterexample A again. $\left(\mathbb{E}|X_n|^p\right)^{\frac{1}{p}} \leq \frac{1}{k^{\frac{1}{p}}}$ and thus as $n \to \infty$, we know that X_n converge to $X = 0$ in L^p. However, we have shown that X_n do not converge to $X = 0$ almost surely.

In order to prove (i) $\Rightarrow$ (iv), it suffices to prove (ii) $\Rightarrow$ (iv).
(ii) $\Rightarrow$ (iv): For any $x_1 < x$, we have that

$$\begin{aligned} \{X \leq x_1\} &= \{X_n \leq x, X \leq x_1\} \cup \{X_n > x, X \leq x_1\} \\ &\subset \{X_n \leq x\} \cup \{X_n > x, X \leq x_1\}. \end{aligned}$$

Thus,

$$F(x_1) \leq F_n(x) + \mathbb{P}\left(\{X_n > x, X \leq x_1\}\right) = F_n(x) + \mathbb{P}\left(\{|X_n - X| \geq x - x_1\}\right).$$

Since X_n converge to X in probability, we know that $\lim_{n\to\infty} \mathbb{P}(\{|X_n - X| \geq x - x_1\}) = 0$. Thus,

$$F(x_1) \leq \liminf_{n\to\infty} F_n(x).$$

On the other hand, if $x_2 > x$, we have that

$$\begin{aligned}\{X > x_2\} &= \{X_n \leq x, X > x_2\} \cup \{X_n > x, X > x_2\} \\ &\subset \{X_n \leq x, X > x_2\} \cup \{X_n > x\}\end{aligned}$$

Thus,

$$1 - F(x_2) \leq \mathbb{P}(\{X_n \leq x, X > x_2\}) + 1 - F_n(x),$$

which, in turn, shows that $1 - F(x_2) \leq 1 - \limsup_{n\to\infty} F_n(x)$.
Therefore, we conclude that

$$F(x_1) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x_2),$$

where $x_1 < x < x_2$. For every x such that F is continuous at x, we let $x_1 \to x$ and $x_2 \to x$ and obtain that

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

yielding that X_n converge to X in distribution.

(iv) $\not\Rightarrow$ (ii) and (iv) $\not\Rightarrow$ (i): **Counterexample C**: Consider $\Omega = \{\omega_1, \omega_2\}$ and $\mathbb{P}(\omega_1) = \mathbb{P}(\omega_2) = \frac{1}{2}$. Let $X(\omega_1) = -1, X(\omega_2) = 1$. Then the distribution of X is

$$\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}.$$

Now, we define $X_n = -X$ for all n. Obviously, the distributions of X_n's are identical to X, so $X_n \xrightarrow{D} X$ naturally.
However, for any $\epsilon \in (0, 2)$,

$$\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(\Omega) = 1,$$

so X_n's will not converge to X in probability, let alone converge to X almost surely.

(iv) $\Rightarrow$ (ii) under the condition that X is a constant: Let $X \equiv C$ be a constant. For any $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}(|X_n - C| > \epsilon) &= \mathbb{P}(X_n > C + \epsilon) + \mathbb{P}(X_n < C - \epsilon) \\ &\leq 1 - F_n(C + \epsilon) + F_n(C - \epsilon).\end{aligned}$$

Since X_n converge to X in distribution and the CDF of $X \equiv C$ is

$$F(x) = \begin{cases} 1, & \text{if } x \geq C, \\ 0, & \text{if } x < C, \end{cases}$$

we let $n \to \infty$ in the above inequality and derive that

$$\mathbb{P}(|X_n - C| > \epsilon) \leq 1 - \lim_{n\to\infty} F_n(C + \epsilon) + \lim_{n\to\infty} F_n(C - \epsilon) = 1 - 1 + 0 = 0.$$

Therefore, X_n converge to $X \equiv C$ in probability.

(ii) $\not\Rightarrow$ (iii): Consider the Counterexample B again. X_n converge to $X = 0$ a.s., so X_n converge to $X = 0$ in probability as shown previously. However, $(\mathbb{E}|X_n|^p)^{\frac{1}{p}} = 1 \neq 0 = (\mathbb{E}|X|^p)^{\frac{1}{p}}$.

(iii) $\Rightarrow$ (ii): We first show that given a random variable Y, for any $\epsilon > 0$,

$$\mathbb{P}(|Y| > \epsilon) = \mathbb{E}(\mathbf{1}_{\{|Y|>\epsilon\}}) \leq \mathbb{E}\left(\frac{|Y|^p}{\epsilon^p}\mathbf{1}_{\{|Y|>\epsilon\}}\right) \leq \frac{\mathbb{E}(|Y|^p)}{\epsilon^p},$$

where $\mathbf{1}_A$ is the indicator function of A.
Let $Y = X_n - X$. We thus have that

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{E(|X_n - X|^p)}{\epsilon^p}.$$

If X_n converge to X in L^p, then

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) \leq \lim_{n\to\infty} \frac{E(|X_n - X|^p)}{\epsilon^p} = 0,$$

showing that X_n converge to X in probability.

(iii) $\Rightarrow$ (iv): It is obvious since we have (iii) $\Rightarrow$ (ii) and (ii) $\Rightarrow$ (iv).

(iv) $\not\Rightarrow$ (iii): In the Counterexample C, $(\mathbb{E}|X_n|^p)^{\frac{1}{p}} = 2 \neq 0$, so X_n's do not converge to $X = 0$ in L^p. □

Problem 2 (Hypergeometric Distribution; Example 3.2 in Perlman [2020])**.** *Suppose an urn contains r red balls and w white balls. Draw n balls at random from the urn and let X denote the number of red balls obtained.*

If the balls are sampled with replacement*, then clearly $X \sim$ Binomial(n, p), where $p = \frac{r}{r+w}$, so*

$$\mathbb{E}(X) = np, \quad Var(X) = np(1-p).$$

In what follows, we suppose that the balls are sampled without replacement*. Note that we now require that $n \leq r + w$.*

- *(a) What is the probability distribution of X?*
- *(b) Calculate the expectation $\mathbb{E}(X)$.*
- *(c) Find the variance $Var(X)$. Do we expect $Var(X)$ also to be the same as for sampling with replacement, namely, $np(1-p)$? Would $Var(X)$ be larger or smaller than $np(1-p)$?*

Solution. (a) Note that the range of $X = x$ is

$$\max(0, n - w) \leq x \leq \min(r, n). \qquad \text{(Why?)}$$

Then, by combinatorics (Vandermonde's identity[1]), we know that the probability mass function (PMF) of X is

$$\mathbb{P}(X = x) = \frac{\binom{r}{x}\binom{w}{n-x}}{\binom{r+w}{n}}, \quad \max(0, n-w) \leq x \leq \min(r, n). \tag{1}$$

[1]See https://en.wikipedia.org/wiki/Vandermonde's_identity.

This distribution is called *hypergeometric* because these ratios of binomial coefficients occurs as the coefficients in the expansion of hypergeometric functions or hypergeometric series[2].

(b) One can directly use (1) to obtain $\mathbb{E}(X)$, which requires complicated combinatorial calculations. We instead consider the following representation of X as:

$$X = X_1 + \cdots + X_n,$$

where, as in the binomial case, $X_i = 1$ (or 0) if a red (or white) ball is obtained on the i-th trial. However, different from the binomial random variable, which is a sum of i.i.d. Bernoulli random variables, $X_1, ..., X_n$ here are *not mutually independent*. (Why?) Nevertheless, the joint distribution of $(X_1, ..., X_n)$ is *exchangeable≡symmetric≡permutation-invariant*, *i.e.*,

$$(X_1, ..., X_n) \sim (X_{i_1}, ..., X_{i_n})$$

for every permutation $(i_1, ..., i_n)$ of $(1, ..., n)$. This is intuitively evident as one can compute, for instance,

$$\mathbb{P}(X_1 = 0, X_2 = 1) = \frac{w}{r+w} \cdot \frac{r}{r+w-1} = \frac{r}{r+w} \cdot \frac{w}{r+w-1} = \mathbb{P}(X_1 = 1, X_2 = 0).$$

By induction, it can be shown that $\mathbb{P}(X_1 = x_1, ..., X_n = x_n) = \mathbb{P}(X_{i_1} = x_1, ..., X_{i_n} = x_n)$. Moreover, notice that

$$\begin{aligned}
\mathbb{P}(X_2 = 1) &= \mathbb{P}(X_2 = 1|X_1 = 1) \cdot \mathbb{P}(X_1 = 1) + \mathbb{P}(X_2 = 1|X_1 = 0) \cdot \mathbb{P}(X_1 = 0) \\
&= \frac{r-1}{r+w-1} \cdot \frac{r}{r+w} + \frac{r}{r+w-1} \cdot \frac{w}{r+w} \\
&= \frac{r}{r+w} \equiv \mathbb{P}(X_1 = 1),
\end{aligned}$$

showing that X_1 and X_2 have the same distribution (but they are not independent). Therefore, by linearity of expectation and exchangeability, $\mathbb{E}(X_i) = \frac{r}{r+w}$ and therefore,

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n\left(\frac{r}{r+w}\right) = np.$$

(c) From (b), we know that X_1X_2 has its range as $\{0, 1\}$, so

$$\begin{aligned}
\mathbb{E}(X_1X_2) &= \mathbb{P}(X_1X_2 = 1) \\
&= \mathbb{P}(X_1 = 1, X_2 = 1) \\
&= \mathbb{P}(X_2 = 1|X_1 = 1) \cdot \mathbb{P}(X_1 = 1) \\
&= \frac{r-1}{r+w-1} \cdot \frac{r}{r+w}
\end{aligned}$$

and

$$\begin{aligned}
\text{Cov}(X_1, X_2) &= \mathbb{E}(X_1X_2) - (\mathbb{E}X_1)(\mathbb{E}X_2) \\
&= \frac{r-1}{r+w-1} \cdot \frac{r}{r+w} - \left(\frac{r}{r+w}\right)^2 \\
&= -\frac{rw}{(r+w)^2(r+w-1)} < 0.
\end{aligned} \tag{2}$$

[2]See https://en.wikipedia.org/wiki/Hypergeometric_function.

Thus, X_1 and X_2 are *negatively correlated*, which is intuitively clear (because $\frac{r-1}{r+w-1} < \frac{r}{r+w}$). By (2) and exchangeability,

$$\begin{aligned}\operatorname{Var}(X) &= \sum_{i=1}^{n} \operatorname{Var}(X_i) + 2 \sum_{1\le i<j\le n} \operatorname{Cov}(X_i, X_j) \\ &= np(1-p) + n(n-1)\operatorname{Cov}(X_1, X_2) \\ &= n\left(\frac{r}{r+w}\right)\left(1 - \frac{r}{r+w}\right) + n(n-1)\left[-\frac{rw}{(r+w)^2(r+w-1)}\right] \\ &= \frac{nrw}{(r+w)^2}\left[1 - \frac{n-1}{r+w-1}\right],\end{aligned}$$

which is smaller than $np(1-p)$ when $n > 1$. It suggests that *samping without replacement from a finite population reduces the variability of the outcome*. This is expected from the representation $X = X_1 + \cdots + X_n$ and the fact that each pair (X_i, X_j) is negatively correlated by (2) and exchangeability. □

References

M. Perlman. Probability and Mathematical Statistics I (STAT 512 Lecture Notes), 2020. URL https://sites.stat.washington.edu/people/mdperlma/STAT%20512%20MDP%20Notes.pdf.