Chapter 5: Convergence of Random Variables

All of Statistics, Wasserman

1. Let $X_1, X_2, \ldots$ be iid with finite mean and finite variance. Show that $S^2 \xrightarrow{P} VX_1$.

Proof. First note that (one may prove the following equation by expanding the square, and using the equalities $\sum_{i=1}^n X_i = n\overline{X}$ and $\sum_{i=1}^n X_i^2 = n\overline{X^2}$)

$$S^2 = \frac{n}{n-1}\left(\overline{X^2} - (\overline{X})^2\right).$$

Furthermore, $\overline{X^2} \xrightarrow{P} E(X_1^2)$ (by the LLN) and $(\overline{X})^2 \xrightarrow{P} (EX_1)^2$ (by part (g) of Theorem 5.5 with $g(x) = x^2$). Therefore, (since convergence in probability behaves nice with sums, products, and constant sequences)

$$S^2 \xrightarrow{P} 1 \cdot \left(E(X_1^2) - (EX_1)^2\right) = VX_1.$$

□

2. Show that $X_n \xrightarrow{L_2} b$ iff $EX_n \to b$ and $VX_n \to 0$.

Proof. Suppose $X_n \xrightarrow{L_2} b$. Since convergence in L_2 implies convergence in L_1 (cf. part (b) of Theorem 5.17)[1] then $E|X_n - b| \to 0$. In particular[2], $EX_n \to b$. Furthermore,

$$E(X_n^2) - 2bEX_n + b^2 = E\left((X_n - b)^2\right) \to 0$$

Therefore,

$$\begin{aligned} VX_n &= E(X_n^2) - (EX_n)^2 \\ &= \left(E(X_n^2) - 2bEX_n\right) + \left(2bEX_n - (EX_n)^2\right) \to -b^2 + 2bb - b^2 = 0. \end{aligned}$$

Conversely, suppose $EX_n \to b$ and $VX_n \to 0$. Then

$$\begin{aligned} E\left((X_n - b)^2\right) &= E(X_n^2) - 2bEX_n + b^2 \\ &= \left(E(X_n^2) - (EX_n)^2\right) + \left((EX_n)^2 - 2bEX_n + b^2\right) \to 0 + \left(b^2 - 2b \cdot b + b^2\right) = 0. \end{aligned}$$

□

3. Let $X_1, X_2, \ldots$ be iid with finite variance. Then $\overline{X}_n \xrightarrow{L_2} EX_1$.

Proof. By linearity of the expectation, $E\overline{X}_n = EX_1$. Also, $V\overline{X}_n = \frac{VX_1}{n} \to 0$. Therefore, the characterization of L_2 convergence given in exercise 2 yields the desired result. □

4. Let X_n be such that

$$P(X_n = 1/n) = 1 - 1/n^2 \qquad \text{and} \qquad P(X_n = n) = 1/n^2.$$

Does X_n converge in probability? In L_2?

[1] Substituting $Y = |X_n - X|$ and $Z = 1$ in Cauchy-Schwarz, yields $(E|X_n - X|)^2 \le E\left((X_n - X)^2\right)$.

[2] The converse isn't true: let X_n be such that $P(X_n = n) = 1/n$, $P(X_n = -n) = 1/n$, and $P(X_n = 0) = 1 - 2/n$. Other counterexample: let $Y_n = X_1$.

Solution. Let $\epsilon > 0$ and let n be such that $1/n < \epsilon < n$. Then

$$P(|X_n| > \epsilon) = P(X_n = n) = 1/n^2 \to 0.$$

That is, $X_n \xrightarrow{P} 0$. Alternatively, note that

$$F_n(x) = \begin{cases} 0 & \text{if } x < 1/n, \\ 1 - 1/n^2 & \text{if } 1/n \le x < n, \\ 1 & \text{if } x \ge n. \end{cases}$$

Clearly, $F_n(x) \to F(x)$ for all continuinity points $x \neq 0$ where

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \ge 0. \end{cases}$$

That is, $X_n \rightsquigarrow 0$. Furthermore, (by part (c) of Theorem 5.4) $X_n \xrightarrow{P} 0$. However, X_n does not converge in L_2: First note that if it did, then it should converge to 0 (since $X_n \xrightarrow{L_2} X \implies X_n \xrightarrow{P} X$). Moreover,

$$E\left((X_n - 0)^2\right) = E(X_n^2) = (1/n)^2 \cdot (1 - 1/n^2) + n^2 \cdot (1/n^2) = 1/n^2 - 1/n^4 + 1 \to 1 \neq 0.$$

$\Diamond$

5. Let $X_1, X_2, \cdots \sim \text{Bernoulli}(p)$. Prove that $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{L_2} p$.

Proof. First note that if X is Bernoulli, then $X^2 = X$. Hence, $\frac{1}{n}\sum_{i=1}^n X_i^2 = \overline{X}_n$. With this, the result is an immediate consequence of exercise 3. $\square$

6. Suppose that the height of men has mean 68 inches and standard deviation 2.6 inches. We draw 100 men at random. Find (approximately) the probability that the average height of men in our sample will be at least 68 inches.

Solution. We will solve it in general: that is, we will approximate $P(\overline{X}_n \ge \mu)$. Calculating directly,

$$\begin{aligned} P(\overline{X}_n \ge \mu) &= P\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \ge \frac{\sqrt{n}(\mu - \mu)}{\sigma}\right) \\ &\approx P(Z \ge 0) && \text{(CLT)} \\ &= 1/2. \end{aligned}$$

(Recall that CLT requires iid with finite variance.) $\Diamond$

7. Let $X_n \sim \text{Poisson}(1/n)$.

(a) Show that $X_n \xrightarrow{P} 0$.

(b) Let $Y_n = nX_n$. Show that $Y_n \xrightarrow{P} 0$.

Proof. (a) Let $\epsilon > 0$. By Markov, $P(|X_n| > \epsilon) < E(X_n)/\epsilon = 1/(n\epsilon) \to 0$.

(b) Let $\epsilon > 0$ and choose n such that $\epsilon/n < 1$. Then

$$P(|Y_n| > \epsilon) = P(|X_n| > \epsilon/n) = P(|X_n| \ge 1) = 1 - P(X_n = 0) = 1 - e^{-1/n} \to 0.$$

where the second equality is because $\epsilon/n < 1$ and $X_n \in \mathbb{Z}_{\ge 0}$.

$\square$

8. Let $X_1, \ldots, X_{100} \overset{\text{iid}}{\sim} \text{Poisson}(1)$ and let $Y = \sum_{i=1}^n X_i$. Use the CLT to approximate $P(Y < 90)$.

Solution. First note that in general,

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} = \frac{Y - n\mu}{\sqrt{n}\sigma}.$$

In this case, $n = 100$, $\mu = 1$, and $\sigma = 1$. Therefore,

$$P(Y < 90) = P\left(\frac{Y - 100}{10} < \frac{90 - 100}{10}\right) \approx P(Z < -1) \approx 0.1587.$$

Furthermore, correcting for continuity[3] and proceeding analogously

$$P(Y < 90) = P(Y \leq 89) \approx P\left(\frac{Y - 100}{10} < \frac{89.5 - 100}{10}\right) = P(Z < -1.05) \approx 0.1469.$$

$\Diamond$

9. Suppose that $P(X = 1) = P(X = -1) = 1/2$. Let

$$X_n = \begin{cases} X & \text{with probability } 1 - 1/n, \\ e^n & \text{with probability } 1/n. \end{cases}$$

Does X_n converge to X in probability? In L^2?

Solution. Let $\epsilon > 0$ and n be large enough so that $e^n - 1 > \epsilon$. By the law of total probability

$$\begin{aligned} P(|X_n - X| > \epsilon) &= P(X_n = X)P\left(|X_n - X| > \epsilon \mid X_n = X\right) + P(X_n = e^n)P\left(|X_n - X| > \epsilon \mid X_n = e^n\right) \\ &= 0 + (1/n)P(|e^n - X| > \epsilon) = 1/n \to 0. \end{aligned}$$

On the other hand[4],

$$\begin{aligned} E\left((X_n - X)^2\right) &= P(X_n = X)E\left((X_n - X)^2 \mid X_n = X\right) + P(X_n = e^n)E\left((X_n - X)^2 \mid X_n = e^n\right) \\ &= 0 + (1/n)E\left((X_n - X)^2 \mid Y = 1\right) = (1/n)E\left((e^n - X)^2\right) \\ &= (1/n)(e^{2n} - 2e^n EX + E(X^2)) = (1/n)(e^{2n} - 0 + 1) \to \infty. \end{aligned}$$

That is, X_n does not converge to X in L^2. $\Diamond$

10. Show that for any $t > 0$ and $k > 0$,

$$P(|X| > t) \leq \frac{E|X|^k}{t^k}.$$

Proof. By Markov, $P(|X| > t) = P(|X|^k > t^k) \leq \frac{E|X|^k}{t^k}$. $\square$

11. Let $X_n \sim N(0, 1/n)$. Is $X_n \xrightarrow{P} 0$ true?

Solution. Yes: By Chebyshev, $P(|X_n| > \epsilon) \leq \frac{1/n}{\epsilon^2} \to 0$. $\Diamond$

12. Let $X, X_1, X_2, \ldots$ be positive and integer valued. Show that $X_n \rightsquigarrow X$ if and only if

$$\lim_{n \to \infty} P(X_n = k) = P(X = k)$$

for every integer k.

[3] The histogram of an integer valued variable uses bins of the form $[x - 0.5, x + 0.5]$. This explains the 89.5.

[4] If X is defined piecewise on events A_i, then $E\left(f(X)\right) = \sum_i P(A_i)\, E\left(f(X) \mid A_i\right)$. Moreover, if $X = Y$ on the conditioning event, then $E\left(f(X) \mid X = Y\right) = E(f(Y))$.

Proof. Suppose $X_n \rightsquigarrow X$. Let $0 < \epsilon < 1$. Then[5]

$$P(X_n = k) = F_n(k+\epsilon) - F_n(k-\epsilon) \to F(k+\epsilon) - F(k-\epsilon) = P(X = k).$$

Conversely, suppose $\lim_{n\to\infty} P(X_n = k) = P(X = k)$. Then

$$P(X_n \le k) = \sum_{x \le k} P(X_n = x) \to \sum_{x \le k} P(X = k) = P(X \le k).$$

$\square$

13. Let $X_1, X_2, \ldots$ be iid with density f. Suppose $X_i > 0$ a.s. and $\lambda = \lim_{x \downarrow 0} f(x)$. Let $X_n = n \min_{i=1}^n X_i$. Show that $X_n \rightsquigarrow X$ where $X \sim \text{Exponential}(1/\lambda)$ with $0 < \lambda < \infty$.

Proof. Recall that if $Y_1, \ldots, Y_n$ are independent and $Y = \min_{i=1}^n Y_i$, then

$$F_Y(y) = 1 - \prod_{i=1}^n (1 - F_{Y_i}(y)).$$

Therefore,

$$F_{X_n}(x) = 1 - (1 - F(x/n))^n.$$

Since $F_X(x) = 1 - e^{-\lambda x}$ (recall $X \sim \text{Exponential}(1/\lambda)$), it suffices to show

$$(1 - F(x/n))^n \to e^{-\lambda x} \text{ for all } x > 0.$$

For this, note that the hypothesis $X_i > 0$ a.s. and $\lambda = \lim_{x \downarrow 0} f(x)$ imply that for all $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\lambda - \epsilon \le f(u) \le \lambda + \epsilon \qquad \text{for all } 0 < u < \delta.$$

Then for $0 < t < \delta$,

$$(\lambda - \epsilon)t \le F(t) = \int_0^t f(u)\, du \le (\lambda + \epsilon)t.$$

Fix $x \ge 0$. For sufficiently large n, we have $x/n < \delta$, hence

$$1 - (\lambda + \epsilon)\frac{x}{n} \le 1 - F(x/n) \le 1 - (\lambda - \epsilon)\frac{x}{n}.$$

Raising to the n-th power[6] and recalling $(1 + a/n)^n \to e^a$ we obtain

$$e^{-(\lambda+\epsilon)x} \le \liminf_{n\to\infty} (1 - F(x/n))^n \le \limsup_{n\to\infty} (1 - F(x/n))^n \le e^{-(\lambda-\epsilon)x}.$$

Letting $\epsilon \to 0$, we obtain the desired result. $\square$

16. Show that $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$ doesn't imply $X_n + Y_n \rightsquigarrow X + Y$.

Proof. Let $Z \sim N(0, 1)$. Let $X_n = Z$ and $Y_n = Z$ for n even and $Y_n = -Z$ for n odd. Then $X_n \rightsquigarrow Z$ and $Y_n \rightsquigarrow Z$ (by symmetry of the normal). However, $X_n + Y_n = 2Z$ for n even and $X_n + Y_n = 0$ for n odd. Therefore, the distribution of $X_n + Y_n$ does not converge. $\square$

[5] By definition, convergence in distribution means $F_n(x) \to F(x)$ at all *continuity* points of F. Since X is integer valued, the continuity points of F are all non-integer values, i.e. $k \pm \epsilon$ with $k \in \mathbb{Z}$ and $0 < \epsilon < 1$.

[6] For all sufficiently large n, we have $(\lambda + \varepsilon)x/n < 1$. This ensures the bounds are between 0 and 1.