STAT 460/560 Class 3: Convergence of random variables

Ben Bloem-Reddy

Reading: Chapter 2.1-2.2, [Vaa98]. Supplement: Chapter 5, [Was04].

1. Modes of convergence

One of the main applications of probability to statistical problems is to analyze the asymptotic (as $n \to \infty$) behavior of a sequence of random variables $X_1, X_2, \ldots$. Unlike a sequence of numbers, which essentially has only one type of convergence, a sequence of random variables may converge in different ways. This may seem counter-intuitive, but remember: a random variable is a function, and sequences of functions can converge in different ways (e.g., uniformly, pointwise, in norm, etc.) The different notions of convergence of random variables are basically expressing different types of functional convergence, which gets a little more complicated when a probability measure is thrown in.

In this course, we will work primarily with three types of convergence. Let $X_1, X_2, \ldots$ be a sequence of random vectors in $\mathbb{R}^k$, and X another random vector in $\mathbb{R}^k$.

1. **Convergence in distribution.** The sequence $X_1, X_2, \ldots$ converges in distribution to X, written $X_n \rightsquigarrow X$, if

$$\lim_{n\to\infty} P(X_n \leq t) = P(X \leq t) \tag{3.1}$$

   for all $t \in \mathbb{R}^k$ for which the function $t \mapsto P(X \leq t)$ is continuous. Note that when $k = 1$, this is just the usual CDF.

2. **Convergence in probability.** The sequence $X_1, X_2, \ldots$ converges in probability to X, written $X_n \xrightarrow{\text{p}} X$, if for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0 \ . \tag{3.2}$$

3. **Convergence in quadratic mean.** The sequence $X_1, X_2, \ldots$ converges in quadratic mean to X, written $X_n \xrightarrow{\text{qm}} X$, if

$$\lim_{n\to\infty} \mathbb{E}[(X_n - X)^2] = 0 \ . \tag{3.3}$$

   This is also known as **convergence in** L_2.

Note that the strongest mode of convergence is **almost sure convergence**, $X_n \xrightarrow{\text{a.s.}} X$, when for every $\epsilon > 0$,

$$P(\lim_{n\to\infty} |X_n - X| > \epsilon) = 0 \ . \tag{3.4}$$

Here are some examples to illustrate the differences of the modes of convergence.

Example 3.1 (Convergence in probability but not almost surely). Let $\Omega = (0, 1]$, and P the Lebesgue measure, so that $P((a, b)) = |b - a|$. Let $X_1, X_2, X_3, X_4, X_5, X_6, X_7, \ldots$ be the indicators of $(0, 1], (0, 1/2], (1/2, 1], (0, 1/3], (1/3, 2/3], (2/3, 1], (0, 1/4], \ldots$, respectively. Then for arbitrary $\epsilon \in (0, 1)$, the probabilities $P\{X_n > \epsilon\}$ are the sequence $1, 1/2, 1/2, 1/3, 1/3, 1/3, 1/4, \ldots$, whose limit is 0. Thus, $X_n \xrightarrow{\text{p}} 0$. But, for every $\omega \in \Omega$, the sequence $(X_n(\omega))_{n\geq 1}$ consists of zeros and ones without end (i.e., switching back and forth infinitely often). Therefore, $\liminf_n X_n(\omega) = 0$ and $\limsup_n X_n(\omega) = 1$, and the set of ω for which $(X_n(\omega)))_{n\geq 1}$ converges is empty.

Example 3.2 (Convergence in probability and *not* quadratic mean). Using the same setup as Example 3.1, we have that $X_n \xrightarrow{\text{qm}} 0$, as well, because $\mathbb{E}[X_n^2] \to 0$.

A slight alteration changes that. Let $(\hat{X}_n)_{n\geq 1} = (X_1, \sqrt{2}X_2, \sqrt{2}X_3, \sqrt{3}X_4, \sqrt{3}X_5, \sqrt{3}X_6, \sqrt{4}X_7, \ldots)$. Now for $\epsilon \in (0,1)$, $P\{\hat{X}_n > \epsilon\} = P\{X_n > \epsilon\}$ for each n, so $\hat{X}_n \xrightarrow{\text{p}} 0$. However, $\mathbb{E}[\hat{X}_n^2] = 1$ for all n, so $(\hat{X}_n)_{n\geq 1}$ does not converge to 0 in quadratic mean. But it cannot converge elsewhere, say some $a \neq 0$, otherwise Theorem 3.1(ii) would imply that $X_n \xrightarrow{\text{p}} a$. So $(\hat{X}_n)$ does not converge in quadratic mean.

In most cases, either almost sure convergence is too strong (it won't be achieved); or convergence in probability is good enough even when almost sure convergence is achieved. So we won't see it much in this course. The next theorem formalizes what implications are possible.

Theorem 3.1. *Let X_n, X and Y be random vectors, and c a constant. Then*

(i) $X_n \xrightarrow{\text{a.s.}} X$ *implies that* $X_n \xrightarrow{p} X$.

(ii) $X_n \xrightarrow{qm} X$ *implies that* $X_n \xrightarrow{p} X$.

(iii) $X_n \xrightarrow{p} X$ *implies that* $X_n \rightsquigarrow X$.

(iv) $X_n \xrightarrow{p} c$ *if and only if* $X_n \rightsquigarrow c$.

(v) If $X_n \rightsquigarrow X$ *and* $Y_n \xrightarrow{p} c$ *then* $(X_n, Y_n) \rightsquigarrow (X, c)$.

(vi) If $X_n \xrightarrow{p} X$ *and* $Y_n \xrightarrow{p} Y$ *then* $(X_n, Y_n) \xrightarrow{p} (X, Y)$.

Note that (iv) is a partial converse of (iii). The other converses are in general false without additional conditions. Note also that it in general, if $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$, $(X_n, Y_n) \not\rightsquigarrow (X, Y)$; (v) is the strongest such statement that can be made.

Example 3.3. As an example, suppose that $X_1, \ldots, X_n$ are IID random variables with finite mean $\mathbb{E}(X)$ and variance σ^2. Then

$$\mathbb{E}(\bar{X}_n^2) = \frac{1}{n^2}\mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)\left(\sum_{j=1}^n X_j\right)\right] = \frac{1}{n^2} n\mathbb{E}\left(X_1 \sum_{i=1}^n X_i\right) = \frac{1}{n^2}(n(\sigma^2 + n\mathbb{E}(X)^2)) = \frac{\sigma^2}{n} + \mathbb{E}(X)^2 \,. \tag{3.5}$$

Because $\mathbb{E}[(\bar{X}_n - \mathbb{E}(X))^2] = \mathbb{E}(\bar{X}_n^2) - \mathbb{E}(X)^2$, we see that $\bar{X}_n \xrightarrow{\text{qm}} \mathbb{E}(X)$.

Activity 3.1. Let $X_1, \ldots, X_n$ be IID random variables with finite mean and variance. Show that $\bar{X}_n \xrightarrow{\text{p}} \mathbb{E}(X)$.

Solution: By Chebyshev's inequality,

$$\begin{aligned}
\mathbb{P}(|\bar{X}_n - \mathbb{E}(\bar{X}_n)| > \epsilon) &< \mathbb{P}(|\bar{X}_n - \mathbb{E}(\bar{X}_n)| \geq \epsilon) \\
&= \mathbb{P}(|\bar{X}_n - \mathbb{E}(X)| \geq \epsilon) \\
&\leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \,,
\end{aligned}$$

then taking the limit as $n \to \infty$ on both sides,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mathbb{E}(\bar{X}_n)| > \epsilon) &\leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} \\ &= 0\end{aligned}$$

Thus, $\bar{X}_n \xrightarrow{\text{p}} \mathbb{E}(X)$.

We proved a version of the **Weak Law of Large Numbers** (WLLN).

Theorem 3.2. *Let $X_1, \dots, X_n$ be* IID *random variables with finite mean $\mathbb{E}(X)$ and variance. Then $\bar{X}_n \xrightarrow{p} \mathbb{E}(X)$.*

The interpretation of the WLLN is that the distribution of $\bar{X}_n$ concentrates around $\mathbb{E}(X)$ as $n \to \infty$.

Exercise 3.1. Exercise 5.2, [Was04].

Solution: Suppose $\lim_{n\to\infty} \mathbb{E}\left(X_n - b\right)^2 = 0$. Then for all n,

$$\begin{aligned}\mathbb{E}\left(X_n - b\right)^2 &= \mathbb{E}\left(X_n^2 - 2bX_n + b^2\right) \\ &= \mathbb{E}(X_n^2) - \mathbb{E}(2bX_n) + \mathbb{E}(b^2) \qquad \text{by linearity of expectation} \\ &= \mathbb{E}(X_n^2) - 2b\mathbb{E}(X_n) + b^2 \qquad \text{by expectation of constants} \\ &= \mathbb{E}(X_n^2) - 2b\mathbb{E}(X_n) + b^2 - \left(\mathbb{E}(X_n)\right)^2 + \left(\mathbb{E}(X_n)\right)^2 \\ &= \mathbb{E}(X_n^2) + \Big(\left(\mathbb{E}(X_n)\right)^2 - 2b\mathbb{E}(X_n) + b^2\Big) - \left(\mathbb{E}(X_n)\right)^2 \\ &= \mathbb{E}(X_n^2) + \left(\mathbb{E}(X_n) - b\right)^2 - \left(\mathbb{E}(X_n)\right)^2 \\ &= \mathrm{Var}(X_n) + \left(\mathbb{E}(X_n) - b\right)^2 \\ &= \mathrm{Var}(X_n) + \mathsf{bias}_b^2(X_n)\ .\end{aligned}$$

Since both of these terms are non-negative and by assumption, their sum in the limit $n \to \infty$ is equal to zero, they must both converge to zero as $n \to \infty$. Therefore,

$$\lim_{n\to\infty} \mathbb{E}(X_n) = b\ , \quad \lim_{n\to\infty} \mathrm{Var}(X_n) = 0\ . \tag{3.6}$$

Conversely, if both of those convergence statements hold then by the identity

$$\mathbb{E}\left(X_n - b\right)^2 = \mathrm{Var}(X_n) + \left(\mathbb{E}(X_n) - b\right)^2\ , \tag{3.7}$$

clearly $\mathbb{E}\left(X_n - b\right)^2 \to 0$, which implies $X_n \xrightarrow{\text{qm}} b$.

Some modes of convergence are preserved by continuous functions.

Theorem 3.3 (Continuous mapping theorem)**.** *Let X_n, X be random vectors in $\mathbb{R}^k$ and let $g \colon \mathbb{R}^k \to \mathbb{R}^m$ be continuous at every point of a set C such that $P\{X \in C\} = 1$. Then*

(i) If $X_n \rightsquigarrow X$ then $g(X_n) \rightsquigarrow g(X)$.

(ii) If $X_n \xrightarrow{p} X$ then $g(X_n) \xrightarrow{p} g(X)$.

(iii) If $X_n \xrightarrow{\text{a.s.}} X$ then $g(X_n) \xrightarrow{\text{a.s.}} g(X)$.

Activity 3.2. Suppose that $X_n \rightsquigarrow \mathcal{N}(0,1)$ and $Y_n \xrightarrow{\text{p}} \sigma$. Show that $X_nY_n \rightsquigarrow \mathcal{N}(0, \sigma^2)$.

This is a special case of Slutsky's lemma.

Lemma 3.4 (Slutsky)**.** *Let X_n, X, Y_n be random vectors, and c a constant. If $X_n \rightsquigarrow X$ and $Y_n \xrightarrow{p} c$, then* *(i)* $X_n + Y_n \rightsquigarrow X + c$.

(ii) $Y_n X_n \rightsquigarrow cX$.

(iii) $X_n / Y_n \rightsquigarrow X/c$, *provided* $c \neq 0$.

As Vaart [Vaa98] notes on p. 11, these are particular instantiations of the following: since $X_n \rightsquigarrow X$ and $Y_n \xrightarrow{\text{P}} c$ implies $(X_n, Y_n) \rightsquigarrow (X, c)$, Theorem 3.3 tells us that $g(X_n, Y_n) \rightsquigarrow g(X, c)$ for any function g that is continuous on the subset of $\mathbb{R}^k \times \{c\}$ in which (X, c) takes its values.

We can apply similar reasoning to show the following.

Theorem 3.5. *If* $X_n \xrightarrow{p} X$ *and* $Y_n \xrightarrow{p} Y$*, then* $g(X_n, Y_n) \xrightarrow{p} g(X, Y)$ *for* g *continuous on a set* C *such that* $P\{(X, Y) \in C\} = 1$.

Activity 3.3 (Asymptotically valid approximate confidence interval). Suppose that we have sequences of estimators T_n and S_n such that

$$\sqrt{n}(T_n - \theta) \rightsquigarrow \mathcal{N}(0, \sigma^2) \quad \text{and} \quad S_n^2 \xrightarrow{\text{P}} \sigma^2 \ ,$$

for some parameters θ and $\sigma^2 > 0$. Form the confidence interval

$$[T_n - z_{\alpha/2} S_n/\sqrt{n}, T_n + z_{\alpha/2} S_n/\sqrt{n}] \ ,$$

where $z_{\alpha/2}$ is the value of the quantile function of the standard normal distribution evaluated at $\alpha/2$. Show that

$$\lim_n P\{T_n - z_{\alpha/2} S_n/\sqrt{n} \leq \theta \leq T_n + z_{\alpha/2} S_n/\sqrt{n}\} = 1 - \alpha \ .$$

Solution: By (iii) of Slutsky's lemma, $\sqrt{n}(T_n - \theta)/S_n \rightsquigarrow \mathcal{N}(0, 1)$. Hence,

$$\lim_n P\{-z_{\alpha/2} \leq \sqrt{n}(T_n - \theta)/S_n \leq z_{\alpha/2}\} = 1 - \alpha \ .$$

Rearranging the terms in the probability argument gives us the confidence interval.

2. Stochastic o and O symbols

Much of our work will require analyzing the behavior of sequences of random variables, and it's convenient to have some compact notation for sequences that converge in certain ways.

A sequence of random vectors X_n is **bounded in probability** if there is some finite number M such that, for each $\epsilon > 0$,

$$\sup_n P(\|X_n\| > M) < \epsilon \ .$$

(Note that the same M can be used for each n. Another name for this is uniform tightness.) In this case, we write $X_n = O_P(1)$ and say that X_n is "big oh-P-one". More generally, for a sequence R_n, we write $X_n = O_P(R_n)$ if $X_n = Y_n R_n$ and $Y_n \xrightarrow{\text{P}} 0$, which indicates that X_n is bounded at "rate" R_n. For example, when $R_n \neq 0$, $X_n = O_P(R_n)$ implies that $X_n/R_n \xrightarrow{\text{P}} 0$.

If, instead, $X_n \xrightarrow{\text{P}} 0$, we write $o_P(1)$ and say that X_n is "little oh-P-one". More generally, we write $X_n = o_P(R_n)$ if $X_n = Y_n R_n$ and $Y_n \xrightarrow{\text{P}} 0$. Again, we interpret R_n as the rate at which X_n converges in probability to zero.

Vaart [Vaa98] lists some simple rules of calculus that the o_P and O_P symbols obey. For example,

$$o_P(1) + o_P(1) = o_P(1) \ ,$$

which should be interpreted as: if $X_n \xrightarrow{\text{P}} 0$ and $Y_n \xrightarrow{\text{P}} 0$, then $X_n + Y_n \xrightarrow{\text{P}} 0$. This is just a special case of Theorem 3.5. The others have similar interpretations.

References

[Vaa98] A. W. v. d. Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[Was04] L. Wasserman. *All of Statistics*. Springer New York, 2004.