STAT 460/560 Class 4: Convergence of random variables part 2

Ben Bloem-Reddy

Reading: Chapter 2.2-2.3, [van98]. Supplement: Chapter 5, [Was04].

We'll continue with our study of convergence of random variables.

1. Continuous mapping theorem(s)

Some modes of convergence are preserved by continuous functions.

Theorem 4.1 (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 4.1. Suppose that $X_n \rightsquigarrow \mathcal{N}(0,1)$ and $Y_n \xrightarrow{\text{p}} \sigma$. Show that $X_n Y_n \rightsquigarrow \mathcal{N}(0, \sigma^2)$.

This is a special case of Slutsky's lemma.

Lemma 4.2 (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 van der Vaart [van98] 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 4.1 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 4.3. *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 4.2 (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

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

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.

van der Vaart [van98] 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 4.3. The others have similar interpretations.

3. Characteristic functions

A generalization of the moment generating is the **characteristic function**, defined for a random vector $X \in \mathbb{R}^k$ as the complex function (complex in the sense of complex numbers, with $i = \sqrt{-1}$),

$$\phi_X(t) := E[e^{it^\top X}] \ , \quad t \in \mathbb{R}^k \ . \tag{4.1}$$

Learning complex integration is way beyond the scope of this class. If you've never encountered a complex integral before, don't worry. For the situations we'll encounter in this course, you can mostly ignore the i, and treat it like just a distinguished element of $\mathbb{R}$ that gets its own notation, like π.

Observe that $\phi_X(0) = 1$. Here are some other facts about the characteristic function, which we won't prove here.

- Characteristic functions are in unique correspondence with probability measures on $\mathbb{R}^k$, so X and Y are equal in distribution if and only if $E[e^{it^\top X}] = E[e^{it^\top Y}]$ for all $t \in \mathbb{R}^k$. (Lemma 2.15 in [van98]).
- $X_n \rightsquigarrow X$ if and only if $E[e^{it^\top X_n}] \to E[e^{it^\top X}]$ for all $t \in \mathbb{R}^k$. Moreover, if $E[e^{it^\top X_n}]$ converges pointwise to a function $\phi(t)$ that is continuous at 0, then ϕ is the characteristic function of a random vector X and $X_n \rightsquigarrow X$. (This is Lévy's continuity theorem, 2.13 in [van98].)
- The characteristic function is determined by the set of all linear combinations $t^\top X$, $t \in \mathbb{R}^k$. So both of the previous properties can be written in terms of $t^\top X$ for all $t \in \mathbb{R}^k$. This is called the Cramér–Wold device.

The first two properties of the characteristic function are used frequently, both to identify the distribution of sums of i.i.d. random vectors, and to characterize the limiting distribution of a sequence of random vectors.

Example 4.1. Let $X \sim \mathcal{N}_k(\mu, \Sigma)$. (The subscript k indicates that $X \in \mathbb{R}^k$.) Then

$$\begin{aligned} E[e^{z^\top X}] &= \int \frac{1}{(2\pi)^{k/2}(\det \Sigma)^{1/2}} \exp\bigg(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\bigg) e^{z^\top x} dx \\ &= \frac{1}{(2\pi)^{k/2}(\det \Sigma)^{1/2}} \int e^{-\frac{1}{2}x^\top \Sigma^{-1} x - \frac{1}{2}\mu^\top \mu + x^\top(z + \Sigma^{-1}\mu)} dx \\ &= \frac{e^{z^\top \mu + \frac{1}{2} z^\top \Sigma^{-1} z}}{(2\pi)^{k/2}(\det \Sigma)^{1/2}} \int e^{-\frac{1}{2}(x - (\mu + \Sigma z))^\top \Sigma^{-1}(x-(\mu+\Sigma z))} dx \\ &= e^{z^\top \mu + \frac{1}{2} z^\top \Sigma^{-1} z} . \end{aligned}$$

For $z \in \mathbb{R}^k$, this follows from completing the square. For $z = it$, we need some complex analysis. You can take my word that it works out. Therefore, for $X \sim \mathcal{N}_k(\mu, \Sigma)$,

$$\phi_X(t) = e^{it^\top \mu - \frac{1}{2} t^\top \Sigma t} . \tag{4.2}$$

4. Central Limit Theorem

The multivariate Central Limit Theorem (CLT) in its most basic form states that if $X_1, X_2, \ldots$ are IID random vectors in $\mathbb{R}^k$ with mean $\mathbb{E}(X)$ and covariance matrix Σ, then

$$\sqrt{n}(\bar{X} - \mathbb{E}(X)) \rightsquigarrow \mathcal{N}(0, \Sigma) . \tag{4.3}$$

This is widely used in its univariate form to obtain, for example, approximate confidence intervals for parameters estimated by the sample mean.

For brevity, we won't prove it here, but see Ch. 2.3 in [van98]. The basic idea is that for $k = 1$, we can make a Taylor expansion of $\phi_{\bar{X}_n}(t)$ around $t = 0$ to find that $\bar{X}_n \xrightarrow{\text{P}} E(X)$. That's the weak LLN, which we already proved with Chebyshev's inequality. The next step is to again make a Taylor expansion, this time of $\phi_{\sqrt{n}(\bar{X}_n - \mu)}(t)$ around $t = 0$. Because the expectation of $\sqrt{n}(\bar{X}_n - \mu)$ is 0, the expansion is

$$E[e^{it\sqrt{n}(\bar{X}_n - \mu)}] = \left(1 - \frac{1}{2}\frac{t^2}{n} E((X-\mu)^2) + o\left(\frac{1}{n}\right)\right)^n \to e^{-\frac{1}{2}t^2 E((X-\mu)^2)} = \phi_Z(t) ,$$

where $Z \sim \mathcal{N}(0, \text{Var}(X))$.

For the multivariate version, the Crámer–Wold device can be used to extend the result to $\mathbb{R}^k$.

References

[van98] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

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