Lecture 13: Weak convergence

A sequence $\{P_n\}$ of probability measures on $(\mathcal{R}^k, \mathcal{B}^k)$ is *tight* if for every $\epsilon > 0$, there is a compact set $C \subset \mathcal{R}^k$ such that $\inf_n P_n(C) > 1 - \epsilon$.
If $\{X_n\}$ is a sequence of random k-vectors, then the tightness of $\{P_{X_n}\}$ is the same as the boundedness of $\{\|X_n\|\}$ in probability ($\|X_n\| = O_p(1)$).

Proposition 1.17. Let $\{P_n\}$ be a sequence of probability measures on $(\mathcal{R}^k, \mathcal{B}^k)$.
(i) Tightness of $\{P_n\}$ is a necessary and sufficient condition that for every subsequence $\{P_{n_i}\}$ there exists a further subsequence $\{P_{n_j}\} \subset \{P_{n_i}\}$ and a probability measure P on $(\mathcal{R}^k, \mathcal{B}^k)$ such that $P_{n_j} \to_w P$ as $j \to \infty$.
(ii) If $\{P_n\}$ is tight and if each subsequence that converges weakly at all converges to the same probability measure P, then $P_n \to_w P$.
The proof can be found in Billingsley (1986, pp. 392-395).

The following result gives some useful sufficient and necessary conditions for convergence in distribution.

Theorem 1.9. Let $X, X_1, X_2, \ldots$ be random k-vectors.
(i) $X_n \to_d X$ is equivalent to any one of the following conditions:
  (a) $E[h(X_n)] \to E[h(X)]$ for every bounded continuous function h;
  (b) $\limsup_n P_{X_n}(C) \le P_X(C)$ for any closed set $C \subset \mathcal{R}^k$;
  (c) $\liminf_n P_{X_n}(O) \ge P_X(O)$ for any open set $O \subset \mathcal{R}^k$.
(ii) (Lévy-Cramér continuity theorem). Let $\phi_X, \phi_{X_1}, \phi_{X_2}, \ldots$ be the ch.f.'s of $X, X_1, X_2, \ldots$, respectively. $X_n \to_d X$ if and only if $\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$ for all $t \in \mathcal{R}^k$.
(iii) (Cramér-Wold device). $X_n \to_d X$ if and only if $c^\tau X_n \to_d c^\tau X$ for every $c \in \mathcal{R}^k$.
Proof. (i) First, we show $X_n \to_d X$ implies (a). By Theorem 1.8(iv) (Skorohod's theorem), there exists a sequence of random vectors $\{Y_n\}$ and a random vector Y such that $P_{Y_n} = P_{X_n}$ for all n, $P_Y = P_X$ and $Y_n \to_{a.s.} Y$. For bounded continuous h, $h(Y_n) \to_{a.s.} h(Y)$ and, by the dominated convergence theorem, $E[h(Y_n)] \to E[h(Y)]$. Then (a) follows from $E[h(X_n)] = E[h(Y_n)]$ for all n and $E[h(X)] = E[h(Y)]$.

Next, we show (a) implies (b). Let C be a closed set and $f_C(x) = \inf\{\|x - y\| : y \in C\}$. Then f_C is continuous. For $j = 1, 2, \ldots$, define $\varphi_j(t) = I_{(-\infty,0]} + (1 - jt)I_{(0,j^{-1}]}$. Then $h_j(x) = \varphi_j(f_C(x))$ is continuous and bounded, $h_j \ge h_{j+1}$, $j = 1, 2, \ldots$, and $h_j(x) \to I_C(x)$ as $j \to \infty$. Hence $\limsup_n P_{X_n}(C) \le \lim_{n\to\infty} E[h_j(X_n)] = E[h_j(X)]$ for each j (by (a)). By the dominated convergence theorem, $E[h_j(X)] \to E[I_C(X)] = P_X(C)$. This proves (b).

For any open set O, O^c is closed. Hence, (b) is equivalent to (c). Now, we show (b) and (c) imply $X_n \to_d X$. For $x = (x_1, \ldots, x_k) \in \mathcal{R}^k$, let $(-\infty, x] = (-\infty, x_1] \times \cdots \times (-\infty, x_k]$ and $(-\infty, x) = (-\infty, x_1) \times \cdots \times (-\infty, x_k)$. From (b) and (c), $P_X((-\infty, x)) \le \liminf_n P_{X_n}((-\infty, x)) \le \liminf_n F_{X_n}(x) \le \limsup_n F_{X_n}(x) = \limsup_n P_{X_n}((-\infty, x]) \le P_X((-\infty, x]) = F_X(x)$. If x is a continuity point of F_X, then $P_X((-\infty, x)) = F_X(x)$. This proves $X_n \to_d X$ and completes the proof of (i).

(ii) From (a) of part (i), $X_n \to_d X$ implies $\phi_{X_n}(t) \to \phi_X(t)$, since $e^{\sqrt{-1}t^\tau x} = \cos(t^\tau x) + \sqrt{-1}\sin(t^\tau x)$ and $\cos(t^\tau x)$ and $\sin(t^\tau x)$ are bounded continuous functions for any fixed t.

Suppose now that $k = 1$ and that $\phi_{X_n}(t) \to \phi_X(t)$ for every $t \in \mathcal{R}$.
We want to show that $\{P_{X_n}\}$ is tight. By Fubini's theorem,

$$\begin{aligned}\frac{1}{u}\int_{-u}^{u}[1-\phi_{X_n}(t)]dt &= \int_{-\infty}^{\infty}\left[\frac{1}{u}\int_{-u}^{u}(1-e^{\sqrt{-1}tx})dt\right]dP_{X_n}(x)\\ &= 2\int_{-\infty}^{\infty}\left(1-\frac{\sin ux}{ux}\right)dP_{X_n}(x)\\ &\geq 2\int_{\{|x|>2u^{-1}\}}\left(1-\frac{1}{|ux|}\right)dP_{X_n}(x)\\ &\geq P_{X_n}\left((-\infty,-2u^{-1})\cup(2u^{-1},\infty)\right)\end{aligned}$$

for any $u > 0$. Since ϕ_X is continuous at 0 and $\phi_X(0) = 1$, for any $\epsilon > 0$ there is a $u > 0$ such that $u^{-1}\int_{-u}^{u}[1-\phi_X(t)]dt < \epsilon/2$. Since $\phi_{X_n} \to \phi_X$, by the dominated convergence theorem, $\sup_n\{u^{-1}\int_{-u}^{u}[1-\phi_{X_n}(t)]dt\} < \epsilon$. Hence,

$$\inf_n P_{X_n}\left([-2u^{-1}, 2u^{-1}]\right) \geq 1 - \sup_n\left\{\frac{1}{u}\int_{-u}^{u}[1-\phi_{X_n}(t)]dt\right\} \geq 1-\epsilon,$$

i.e., $\{P_{X_n}\}$ is tight.
Let $\{P_{X_{n_j}}\}$ be any subsequence that converges to a probability measure P.
By the first part of the proof, $\phi_{X_{n_j}} \to \phi$, which is the ch.f. of P.
By the convergence of ϕ_{X_n}, $\phi = \phi_X$. By the uniqueness theorem, $P = P_X$.
By Proposition 1.17(ii), $X_n \to_d X$.

Consider now the case where $k \geq 2$ and $\phi_{X_n} \to \phi_X$.
Let Y_{nj} be the jth component of X_n and Y_j be the jth component of X.
Then $\phi_{Y_{nj}} \to \phi_{Y_j}$ for each j.
By the proof for the case of $k = 1$, $Y_{nj} \to_d Y_j$.
By Proposition 1.17(i), $\{P_{Y_{nj}}\}$ is tight, $j = 1, ..., k$. This implies that $\{P_{X_n}\}$ is tight (why?).
Then the proof for $X_n \to_d X$ is the same as that for the case of $k = 1$.

(iii) Note that $\phi_{c^\tau X_n}(u) = \phi_{X_n}(uc)$ and $\phi_{c^\tau X}(u) = \phi_X(uc)$ for any $u \in \mathcal{R}$ and any $c \in \mathcal{R}^k$. Hence, convergence of ϕ_{X_n} to ϕ_X is equivalent to convergence of $\phi_{c^\tau X_n}$ to $\phi_{c^\tau X}$ for every $c \in \mathcal{R}^k$. Then the result follows from part (ii).

Example 1.28. Let $X_1, ..., X_n$ be independent random variables having a common c.d.f. and $T_n = X_1 + \cdots + X_n$, $n = 1, 2, ...$ Suppose that $E|X_1| < \infty$. It follows from a result in calculus that the ch.f. of X_1 satisfies

$$\phi_{X_1}(t) = \phi_{X_1}(0) + \sqrt{-1}\mu t + o(|t|)$$

as $|t| \to 0$, where $\mu = EX_1$. Then, the ch.f. of T_n/n is

$$\phi_{T_n/n}(t) = \left[\phi_{X_1}\left(\frac{t}{n}\right)\right]^n = \left[1 + \frac{\sqrt{-1}\mu t}{n} + o\left(\frac{t}{n}\right)\right]^n$$

for any $t \in \mathcal{R}$, as $n \to \infty$. Since $(1 + c_n/n)^n \to e^c$ for any complex sequence $\{c_n\}$ satisfying $c_n \to c$, we obtain that $\phi_{T_n/n}(t) \to e^{\sqrt{-1}\mu t}$, which is the ch.f. of the distribution degenerated at μ (i.e., the point mass probability measure at μ). By Theorem 1.9(ii), $T_n/n \to_d \mu$. From Theorem 1.8(vii), this also shows that $T_n/n \to_p \mu$.

Similarly, $\mu = 0$ and $\sigma^2 = \mathrm{Var}(X_1) < \infty$ imply

$$\phi_{T_n/\sqrt{n}}(t) = \left[1 - \frac{\sigma^2 t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n$$

for any $t \in \mathcal{R}$, which implies that $\phi_{T_n/\sqrt{n}}(t) \to e^{-\sigma^2 t^2/2}$, the ch.f. of $N(0, \sigma^2)$. Hence $T_n/\sqrt{n} \to_d N(0, \sigma^2)$. If $\mu \neq 0$, a transformation of $Y_i = X_i - \mu$ leads to $(T_n - n\mu)/\sqrt{n} \to_d N(0, \sigma^2)$.

Suppose now that $X_1, ..., X_n$ are random k-vectors and $\mu = EX_1$ and $\Sigma = \mathrm{Var}(X_1)$ are finite. For any fixed $c \in \mathcal{R}^k$, it follows from the previous discussion that $(c^\tau T_n - nc^\tau \mu)/\sqrt{n} \to_d N(0, c^\tau \Sigma c)$. From Theorem 1.9(iii) and a property of the normal distribution (Exercise 81), we conclude that $(T_n - n\mu)/\sqrt{n} \to_d N_k(0, \Sigma)$.

Example 1.29. Let $X_1, ..., X_n$ be independent random variables having a common Lebesgue p.d.f. $f(x) = (1 - \cos x)/(\pi x^2)$. Then the ch.f. of X_1 is $\max\{1 - |t|, 0\}$ (Exercise 73) and the ch.f. of $T_n/n = (X_1 + \cdots + X_n)/n$ is

$$\left(\max\left\{1 - \frac{|t|}{n}, 0\right\}\right)^n \to e^{-|t|}, \qquad t \in \mathcal{R}.$$

Since $e^{-|t|}$ is the ch.f. of the Cauchy distribution $C(0, 1)$ (Table 1.2), we conclude that $T_n/n \to_d X$, where X has the Cauchy distribution $C(0, 1)$.

Does this result contradict the first result in Example 1.28?

Other examples are given in Exercises 135-140.

The following result can be used to check whether $X_n \to_d X$ when X has a p.d.f. f and X_n has a p.d.f. f_n.

Proposition 1.18 (Scheffé's theorem). Let $\{f_n\}$ be a sequence of p.d.f.'s on $\mathcal{R}^k$ w.r.t. a measure ν. Suppose that $\lim_{n\to\infty} f_n(x) = f(x)$ a.e. ν and $f(x)$ is a p.d.f. w.r.t. ν. Then $\lim_{n\to\infty} \int |f_n(x) - f(x)| d\nu = 0$.
Proof. Let $g_n(x) = [f(x) - f_n(x)] I_{\{f \geq f_n\}}(x)$, $n = 1, 2, ...$ Then

$$\int |f_n(x) - f(x)| d\nu = 2 \int g_n(x) d\nu.$$

Since $0 \leq g_n(x) \leq f(x)$ for all x and $g_n \to 0$ a.e. ν, the result follows from the dominated convergence theorem.

As an example, consider the Lebesgue p.d.f. f_n of the t-distribution t_n (Table 1.2), $n = 1, 2, ...$ One can show (exercise) that $f_n \to f$, where f is the standard normal p.d.f. This is an important result in statistics.