Proof of Theorem 1.8(viii)

First, by part (iv), we may assume that $X_n \to_{a.s.} X$ (why?).
Next, for simplicity, we consider $r = 1$ and $k = 1$ only
(the general case is shown in the textbook)

UI: $\lim_{t\to\infty} \sup_n E\left(|X_n| I_{\{|X_n|>t\}}\right) = 0$
MC: $\lim_{n\to\infty} E|X_n| = E|X| < \infty$

Proof of UI implies MC

By UI, for an $\varepsilon > 0$, there is a finite $t > 0$ such that

$$\sup_n E\left(|X_n| I_{\{|X_n|>t\}}\right) < \varepsilon$$

Then

$$\sup_n E|X_n| \le \sup_n E\left(|X_n| I_{\{|X_n|>t\}}\right) + \sup_n E\left(|X_n| I_{\{|X_n|\le t\}}\right) < \varepsilon + t < \infty$$

By Fatou's lemma (Theorem 1.1(i)),

$$E|X| \le \liminf_n E|X_n| < \sup_n E|X_n| < \infty$$

Proof of UI implies MC

Hence, MC follows if we can show that

$$\limsup_n E|X_n| \le E|X|.$$

For any $\varepsilon > 0$ and $t > 0$, let $A_n = \{|X_n - X| \le \varepsilon\}$ and $B_n = \{|X_n| > t\}$. Then

$$\begin{aligned} E|X_n| &= E(|X_n| I_{A_n^c \cap B_n}) + E(|X_n| I_{A_n^c \cap B_n^c}) + E(|X_n| I_{A_n}) \\ &\le E(|X_n| I_{B_n}) + tP(A_n^c) + E|X_n I_{A_n}|. \end{aligned}$$

Since $|X_n I_{A_n}| \le (|X_n - X| + |X|) I_{A_n}$,

$$E|X_n I_{A_n}| \le E[(|X_n - X| + |X|) I_{A_n}] \le \varepsilon + E|X|.$$

Since ε is arbitrary, $\limsup_n E|X_n I_{A_n}| \le E|X|$.
This result and previous inequality imply that

$$\limsup_n E|X_n| \le \limsup_n E(|X_n| I_{B_n}) + t \lim_{n\to\infty} P(A_n^c) + E|X|,$$

Since $\lim_{n\to\infty} P(A_n^c) = 0$ and $\{|X_n|\}$ is uniformly integrable, letting $t \to \infty$ we obtain the result.

Proof of MC implies UI

Let $\xi_n = |X_n| I_{B_n^c} - |X| I_{B_n^c}$, $B_n = \{|X_n| > t\}$.
Then $\xi_n \to_{a.s.} 0$ and $|\xi_n| \le t + |X|$, which is integrable.
By the dominated convergence theorem, $E\xi_n \to 0$; this and UI imply

$$E(|X_n| I_{B_n}) - E(|X| I_{B_n}) \to 0.$$

Since $E|X| < \infty$, by the dominated convergence theorem,

$$\lim_{n\to\infty} E(|X| I_{\{|X_n - X| > t/2\}}) = 0$$

From the definition of B_n,

$$|X| I_{B_n} \le |X| I_{\{|X_n - X| > t/2\}} + |X| I_{\{|X| > t/2\}}.$$

Hence

$$\limsup_n E(|X_n| I_{B_n}) \le \limsup_n E(|X| I_{B_n}) \le E(|X| I_{\{|X| > t/2\}}).$$

Letting $t \to \infty$, it follows from the dominated convergence theorem that

$$\lim_{t\to\infty} \limsup_n E(|X_n| I_{B_n}) \le \lim_{t\to\infty} E(|X| I_{\{|X| > t/2\}}) = 0.$$

This proves UI.

Example 1.27.

As an application of Theorem 1.8(viii) and Proposition 1.15, we consider again the prediction problem in Example 1.22.

Suppose that we predict a random variable X by a random n-vector $Y = (Y_1, ..., Y_n)$, all random variables are defined on $(\Omega, \mathscr{F})$

It is shown in Example 1.22 that $X_n = E(X|Y_1, ..., Y_n)$ is the best predictor in terms of the mean squared prediction error, when $EX^2 < \infty$.

We now show that $X_n \to_{a.s.} X$ when $n \to \infty$ under the assumption that $\mathscr{F} = \sigma(Y_1, Y_2, ...)$ (i.e., $Y_1, Y_2, ...$ provide all information).

From the discussion in §1.4.4, $\{X_n\}$ is a martingale.

Also, $\sup_n E|X_n| \leq \sup_n E[E(|X||Y_1, ..., Y_n)] = E|X| < \infty$.

Hence, by Proposition 1.15, $X_n \to_{a.s.} Z$ for some random variable Z.

We now need to show $Z = X$ a.s.

Since $EX_n^2 \leq EX^2 < \infty$ (why?), $\{|X_n|\}$ is uniformly integrable (why?).

Example 1.27 (continued)

By Theorem 1.8(viii), $E|X_n - Z| \to 0$, which implies $\int_A X_n dP \to \int_A Z dP$ for any event A.

Note that if $A \in \sigma(Y_1, ..., Y_n)$, then $\int_A X_n dP = \int_A X dP$.

Also, $\sigma(Y_1, ..., Y_n) \subset \sigma(Y_1, ..., Y_m)$ if $m > n$

Therefore, for any $A \in \cup_{j=1}^{\infty} \sigma(Y_1, ..., Y_j)$, $\int_A X dP = \int_A Z dP$.

Since $\cup_{j=1}^{\infty} \sigma(Y_1, ..., Y_j)$ generates $\sigma(Y_1, Y_2, ...) = \mathscr{F}$, we conclude that $\int_A X dP = \int_A Z dP$ for any $A \in \mathscr{F}$ and thus $Z = X$ a.s.

In the proof above, the condition $EX^2 < \infty$ is used only for showing the uniform integrability of $\{|X_n|\}$.

But by Exercise 120, $\{|X_n|\}$ is uniformly integrable as long as $E|X| < \infty$.

Hence $X_n \to_{a.s.} X$ is still true if the condition $EX^2 < \infty$ is replaced by $E|X| < \infty$.

Tightness

A sequence $\{P_n\}$ of probability measures on $(\mathscr{R}^k, \mathscr{B}^k)$ is *tight* if for every $\varepsilon > 0$, there is a compact set $C \subset \mathscr{R}^k$ such that $\inf_n P_n(C) > 1 - \varepsilon$.

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 ($\|X_n\| = O_p(1)$), i.e., for any $\varepsilon > 0$, there is a constant $C_\varepsilon > 0$ such that $\sup_n P(\|X_n\| \geq C_\varepsilon) < \varepsilon$.

Proposition 1.17

Let $\{P_n\}$ be a sequence of probability measures on $(\mathscr{R}^k, \mathscr{B}^k)$.

(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 $(\mathscr{R}^k, \mathscr{B}^k)$ 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 $P_n \to_w P$.

Proof: See Billingsley (1995, pp. 336-337)

Theorem 1.9 (useful sufficient and necessary conditions for convergence in distribution)

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 $C \subset \mathscr{R}^k$;
  (c) $\liminf_n P_{X_n}(O) \ge P_X(O)$ for any open set $O \subset \mathscr{R}^k$.

(ii) (Lévy-Cramér continuity theorem). Let $\phi_X, \phi_{X_1}, \phi_{X_2}, ...$ be the ch.f.'s of $X, X_1, X_2, ...$, respectively.
$X_n \to_d X$ iff $\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$ for all $t \in \mathscr{R}^k$.

(iii) (Cramér-Wold device). $X_n \to_d X$ iff $c^\tau X_n \to_d c^\tau X$ for every $c \in \mathscr{R}^k$.

Proof of Theorem 1.9(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, $E[h(Y_n)] \to E[h(Y)]$.

(a) follows from $E[h(X_n)] = E[h(Y_n)]$ for all n and $E[h(X)] = E[h(Y)]$.
The proof of (b) implies (c) is in the textbook.
For any open set O, O^c is closed: hence, (b) is equivalent to (c).

To complete the proof we now show that (b) and (c) imply $X_n \to_d X$.
For $x = (x_1, ..., x_k) \in \mathscr{R}^k$, 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),

$$P_X\big((-\infty, x)\big) \le \liminf_n P_{X_n}\big((-\infty, x)\big) \le \liminf_n F_{X_n}(x)$$

$$\le \limsup_n F_{X_n}(x) = \limsup_n P_{X_n}\big((-\infty, x]\big) \le P_X\big((-\infty, x]\big) = F_X(x).$$

If x is a continuity point of F_X, then $P_X\big((-\infty, x)\big) = F_X(x)$.
This proves $X_n \to_d X$.

Proof of Theorem 1.9(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.

Proof of Theorem 1.9(ii) (continued)

Suppose that $k = 1$ and that $\phi_{X_n}(t) \to \phi_X(t)$ for every $t \in \mathscr{R}$.
We want to show that $X_n \to_d X$.
We first 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 $\varepsilon > 0$ there is a $u > 0$ such that $u^{-1}\int_{-u}^{u}[1-\phi_X(t)]dt < \varepsilon/2$.

Proof of Theorem 1.9(ii) (continued)

Since $\phi_{X_n} \to \phi_X$, by the dominated convergence theorem,

$$\sup_n \{ u^{-1} \int_{-u}^{u} [1 - \phi_{X_n}(t)] dt \} < \varepsilon.$$

Hence,

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

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), $X_n \to_d X$.

Proof of Theorem 1.9(ii) (continued)

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 $k = 1$.

Proof of Theorem 1.9(iii)

Note that $\phi_{c^\tau X_n}(u) = \phi_{X_n}(uc)$ and $\phi_{c^\tau X}(u) = \phi_X(uc)$ for any $u \in \mathscr{R}$ and any $c \in \mathscr{R}^k$.
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 $c \in \mathscr{R}^k$.
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

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

for any $t \in \mathscr{R}$ as $n \to \infty$, because $(1 + c_n/n)^n \to e^c$ for any complex sequence $\{c_n\}$ satisfying $c_n \to c$.
$e^{\sqrt{-1}\mu t}$ is the ch.f. of 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 $T_n/n \to_p \mu$.

Example 1.28 (continued)

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

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

for any $t \in \mathscr{R}$ as $n \to \infty$.
$e^{-\sigma^2 t^2/2}$ is 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 $\Sigma = \text{var}(X_1)$ are finite.
For any fixed $c \in \mathscr{R}^k$, 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

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

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 next 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 $\mathscr{R}^k$ 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 $g_n(x) = [f(x) - f_n(x)] I_{\{f \geq f_n\}}(x)$, $n = 1, 2, \ldots$.
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), $n = 1, 2, \ldots$.
One can show (exercise) that $f_n \to f$, where f is the p.d.f. of $N(0,1)$.
This is an important result in statistics.