Advanced Statistics

5. Basic Asymptotics

Christian Aßmann

Chair of Survey Statistics and Data Analysis – Otto-Friedrich-Universität Bamberg

In this chapter we consider **sequences of random variables** of the form

$$Y_n = g(X_1, \dots, X_n), \quad \text{where} \quad n = 1, 2, 3, \dots.$$

A simple example for such a sequence is the average of n random variables

$$Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

The objective of asymptotic theory is to establish results relating to the stochastic behavior of such sequences Y_n when $n \to \infty$.
In particular,

- Y_n may converge to a constant in various ways,
- or the distribution of Y_n may converge to some 'limit distribution'.

What are reasons to study the 'asymptotic behavior' of sequences of random variables Y_n?

- Estimators for parameters and test statistics are typically functions such as $Y_n = g(X_1, ..., X_n)$, where n refers to the sample size (number of data observations).
- In order to evaluate/compare the quality of such estimators and test statistics it is necessary to know their probability characteristics and distributions.
- However, in many cases their actual probability density or distribution is unknown or analytically intractable, when n is finite.
- Asymptotic theory often provides tractable approximations to the distribution of functions $g(X_1, ..., X_n)$, when n is sufficiently large.

In the following section we start to repeat some basic concepts from real analysis.

Definition (Convergence of real number sequences): A sequence of real numbers $\{y_n\}$ converges to $y \in \mathbb{R}^1$ iff for every real $\epsilon > 0$ there exists an integer $N(\epsilon)$ such that

$$|y_n - y| < \epsilon \quad \forall n \geq N(\epsilon).$$

The existence of the limit is denoted by $y_n \to y$ or $\lim_{n\to\infty} y_n = y$.

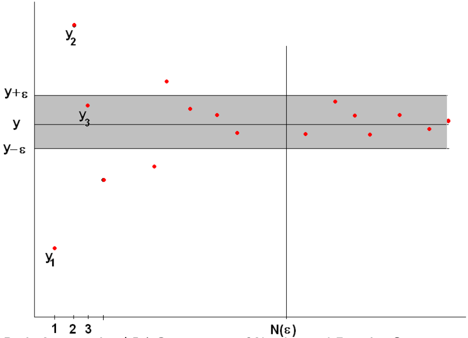

Fig. 27.

It can be shown that for the limit of a sequence of numbers to exist, it is necessary (but not sufficient) that the sequence is bounded, as defined next.

Definition (Boundedness of real number sequences): A sequence of real numbers $\{y_n\}$ is bounded iff there exists a finite number $m > 0$ such that

$$|y_n| \leq m \quad \forall n \in \mathbb{N}.$$

Example: The sequence $\quad y_n = 3 + n^{-2}, \quad n \in \mathbb{N}$

- is bounded, since $|y_n| \leq 4 \quad \forall n \in \mathbb{N}$,
- and has a limit $y_n \to 3$.

The sequence $\quad y_n = \sin(n), \quad n \in \mathbb{N}$

- is bounded, since $|\sin(x)| \leq 1 \quad \forall x$,
- but does not have a limit, since $\sin(x)$ cycles between $+1$ and -1.

||

REMARK : The concept of convergence can be extended to sequences of real valued matrices by applying the definition of convergence of real number sequences to the sequence of matrices element by element. $\diamondsuit$

A further important limit concept is the convergence of a function sequence.

DEFINITION (CONVERGENCE OF FUNCTION SEQUENCES): Let $\{f_n(x)\}$, $n \in \mathbb{N}$, be a sequence of functions having a common domain $D \subset \mathbb{R}^m$. The function sequence $\{f_n(x)\}$ converges to a function $f(x)$ with domain $D_0 \subset D$ iff for $n \to \infty$

$$f_n(x) \to f(x) \quad \forall x \in D_0.$$

f is called the limiting function of $\{f_n\}$.

REMARK : The definition implies that the values of the functions $f_n(x)$, $n = 1, 2, 3, ..$ converge to $f(x)$ pointwise for each single $x \in D_0$.

Hence, $f(x)$ can be viewed as an approximation of $f_n(x)$ when n is large. $\diamondsuit$

EXAMPLE:

- The function sequence $f_n(x) = n^{-1} + 2x^2, \quad x \in \mathbb{R}^1,\ n \in \mathbb{N}$

  has a limit function $f(x) = 2x^2$ since $\lim_{n\to\infty} f_n(x) = 2x^2 \quad \forall x \in \mathbb{R}^1$.

- For the function sequence $f_n(x) = x^n, \quad x \in [0,1],\ n \in \mathbb{N}$

  the limiting function is

$$\lim_{n\to\infty} f_n(x) = f(x) = \begin{cases} 0 & \text{for} \quad x \in [0,1) \\ 1 & \text{for} \quad x = 1 \end{cases} .$$

Note that $f_n(x)$ is continuous for each point of the domain $D = [0,1]$. In contrast, the limit function $f(x)$ is not continuous at $x = 1$. ||

In order to characterize the convergence properties of a sequence, we use the concept of the order of magnitude of a sequence.

Definition (Order of Magnitude of a Sequence): Let $\{x_n\}$ be a real number sequence.

- $\{x_n\}$ is said to be **at most of order** n^k, denoted by $O(n^k)$, if there exists a finite constant c such that
$$\left|\frac{x_n}{n^k}\right| \leq c \quad \forall n \in \mathbb{N}.$$
- $\{x_n\}$ is said to be **of order smaller than** n^k, denoted by $o(n^k)$, if
$$\frac{x_n}{n^k} \to 0.$$

Remark : Note that

if $\{x_n\}$ is $O(n^k)$, then $\{x_n\}$ is $o(n^{k+\epsilon})$ $\forall \epsilon > 0$;
if $\{x_n\}$ is $o(n^k)$, then $\{x_n\}$ is $O(n^k)$ $(c = 0)$.

Notationally, $O(n^0)$ and $o(n^0)$ is written as $O(1)$ and $o(1)$. $\diamondsuit$

EXAMPLE: Let $\{x_n\}$ be defined by $\quad x_n = 3n^3 - n^2 + 2,\ n \in \mathbb{N}$.

Since

$$\frac{x_n}{n^3} = 3 - \frac{1}{n} + \frac{2}{n^3} \quad \to \quad 3 < \infty, \qquad \text{we have} \qquad x_n = O(n^3);$$

Since for a positive ϵ

$$\frac{x_n}{n^{3+\epsilon}} = \frac{3}{n^\epsilon} - \frac{1}{n^{1+\epsilon}} + \frac{2}{n^{3+\epsilon}} \quad \to \quad 0, \qquad \text{we have} \qquad x_n = o(n^{3+\epsilon}). \qquad ||$$

In this section we extend the converge concepts for real number sequences to sequences of random variables.

For sequences of random variables, we distinguish among the following types/modes of convergence:

1.) *convergence in distribution;*
2.) *convergence in probability;*
3.) *convergence in mean square;*
4.) *almost-sure convergence.*

In the following subsection, we begin with the weakest mode of convergence, the convergence in distribution.

DEFINITION (CONVERGENCE IN DISTRIBUTION): Let $\{Y_n\}$ be a sequence of random variables with an associated sequence of cdfs $\{F_n\}$. If there exists a cdf F such that as $n \to \infty$

$$F_n(y) \to F(y) \quad \forall y \quad \text{at which } F \text{ is continuous,}$$

we say that Y_n converges in distribution to the random variable Y with cdf F.

We denote this by $Y_n \xrightarrow{d} Y$ or $Y_n \xrightarrow{d} F$. The function F is called the limiting cdf/limiting distribution of $\{Y_n\}$.

REMARK : The definition implies that if $Y_n \xrightarrow{d} Y$, then as n becomes large, the actual cdf of Y_n can be approximated by the cdf F of the random variable Y. The associated approximation error disappears as $n \to \infty$ (see Fig. 28). $\diamondsuit$

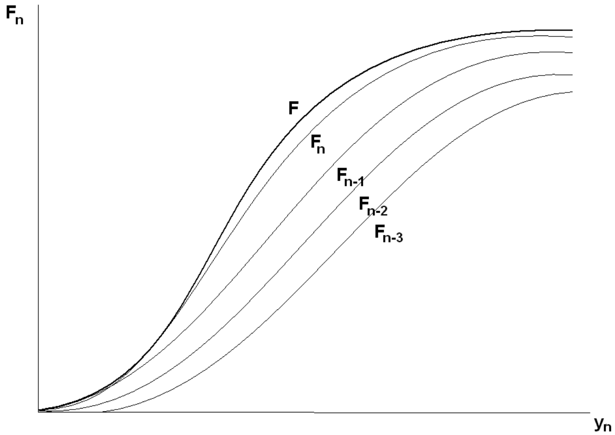

Fig. 28.

REMARK : The limiting cdf F can be the cdf of a *degenerate random variable* with $Y = c$, where c is a constant. In this case, we say that Y_n **converges in distribution to a constant**, and we denote this by $Y_n \xrightarrow{d} c$. $\diamondsuit$

EXAMPLE: Let $\{Y_n\}$ be a sequence of random variables with an associated sequence of cdfs $\{F_n\}$ given by

$$F_n(y) = \begin{cases} 0 & \text{for} \quad y < 0 \\ (\frac{y}{\theta})^n & \text{for} \quad 0 \leq y < \theta \\ 1 & \text{for} \quad y \geq \theta \end{cases} .$$

We see that as $n \to \infty$,

$$F_n(y) \to F(x) = \begin{cases} 0 & \text{for} \quad y < \theta \\ 1 & \text{for} \quad y \geq \theta \end{cases} ,$$

which is the cdf a *degenerate random variable*, and we have $Y_n \xrightarrow{d} \theta$. $||$

REMARK : For nonnegative, integer-valued discrete random variables and continuous random variables

> the convergence of the sequence of pdfs $f_n(y)$ for the sequence of random variables Y_n to the pdf $f(y)$ for random variable Y

is sufficient for establishing convergence in distribution of Y_n to Y ($Y_n \xrightarrow{d} Y$). For more details – see Mittelhammer (1996, Theorem 5.1). $\diamondsuit$

EXAMPLE: Consider the random variable $\{Z_n\}$ with

$$Z_n = \left(3 + \frac{1}{n}\right) Y + \frac{2n}{n-1}, \quad \text{where} \qquad Y \sim \mathrm{N}(0,1) \qquad \forall n.$$

The associated sequence of pdfs $\{f_n\}$ is

$$f_n = \mathrm{N}\left(\frac{2n}{n-1}\,,\,\left(3+\frac{1}{n}\right)^2\right), \qquad \text{such that} \qquad f_n \to \mathrm{N}(2,9).$$

Hence $Z_n \xrightarrow{d} Z \sim \mathrm{N}(2,9)$. $||$

The following theorem is based upon the uniqueness of MGFs (see Theorem 3.16), and is very useful for identifying limiting distributions.

THEOREM 5.1 *Let $\{Y_n\}$ be a sequence of random variables having an associated sequence of MGFs $\{M_{Y_n}(t)\}$. Let $M_Y(t)$ be the MGF of Y. Then*

$$Y_n \xrightarrow{d} Y \qquad \textit{iff} \qquad M_{Y_n}(t) \to M_Y(t) \quad \forall t \in (-h, h), \ \textit{for some } h > 0.$$

For a proof, see Lukacs (1970, p.49-50), *Characteristic Functions,* London, Griffin.

REMARK : The theorem implies that if we can establish that $\lim_{n\to\infty} M_{Y_n}(t)$ is equal to the MGF $M_Y(t)$, then the distribution associated with $M_Y(t)$ is the limiting distribution of the sequence $\{Y_n\}$. $\diamondsuit$

EXAMPLE: Let $X_n \sim \chi^2_{(n)}$ with an MGF $M_{X_n}(t) = (1-2t)^{-\frac{n}{2}} \quad \forall n$.
Consider

$$Z_n = \frac{X_n - n}{\sqrt{2n}} = -\sqrt{\frac{n}{2}} + \frac{1}{\sqrt{2n}} X_n.$$

(Since $\mathrm{E}X_n = n$ and $\mathrm{var}(X_n) = 2n$, the variable Z_n is a standardized χ^2 variable.)

The MGF of Z_n obtains as (see Section 3.5)

$$M_{Z_n}(t) = e^{-\sqrt{\frac{n}{2}}t} \cdot M_{X_n}\left(\frac{1}{\sqrt{2n}} \cdot t\right) = e^{-\sqrt{\frac{n}{2}}t} \cdot \left(1 - \sqrt{\frac{2}{n}}t\right)^{-\frac{n}{2}}.$$

In order to establish the limit of $M_{Z_n}(t)$ as $n \to \infty$, we consider the limit of the transformation $\ln M_{Z_n}(t)$:

- First note that $\ln M_{Z_n}(t) = -\frac{n}{2} \ln\left(1 - \sqrt{\frac{2}{n}}t\right) - \sqrt{\frac{n}{2}}t.$
- A Taylor series expansion of the first term on the r.h.s. around $t = 0$ yields

$$\ln M_{Z_n}(t) = \left[\sqrt{\frac{n}{2}}t + \frac{t^2}{2} + o(1)\right] - \sqrt{\frac{n}{2}}t = \frac{t^2}{2} + o(1) \underset{n\to\infty}{\to} \frac{t^2}{2}.$$

($o(1)$ represents a term of order smaller than $n^0 = 1$ disappearing as $n \to \infty$.)

(CONTINUES)

Example (Continued):

- It follows that

$$\lim_{n\to\infty} M_{Z_n}(t) \;=\; \lim_{n\to\infty} e^{\ln M_{Z_n}(t)} \;=\; e^{\lim_{n\to\infty} \ln M_{Z_n}(t)} = e^{\frac{t^2}{2}}\,.$$

Since $\exp(t^2/2)$ is the MGF of a $\mathrm{N}(0,1)$-distribution, we know by Theorem 5.1 that $Z_n \xrightarrow{d} Z \sim \mathrm{N}(0,1)$. ||

Asymptotic Distributions

Generally speaking, the asymptotic distribution for a random variable Z_n is any distribution that provides an approximation to the true distribution of Z_n for large n.

If $\{Z_n\}$ has a limiting distribution, this limiting distribution might be considered as an asymptotic distribution, since the limiting distribution provides an approximation to the distribution of Z_n for large n.

But what should we do if the sequence $\{Z_n\}$ has *no* limiting distribution or a *degenerate* limiting distribution ($Z_n \xrightarrow{d}$ constant)?

The following definition of the asymptotic distribution generalizes the concept of approximating distributions for large n to include cases where Z_n has no limiting distribution or a degenerate limiting distribution.

Definition (**Asymptotic Distribution**): Let $\{Z_n\}$ be a sequence of random variables defined by

$$Z_n = g(X_n, \Theta_n), \quad \text{where} \quad \begin{array}{l} X_n \xrightarrow{d} X \text{ (nondegenerate)}, \\ \Theta_n : \text{sequence of numbers/parameters.} \end{array}$$

Then an *asymptotic distribution for* Z_n is the distribution of $g(X, \Theta_n)$, denoted by

$$Z_n \overset{a}{\sim} g\left(X, \Theta_n\right) \qquad \text{“}Z_n \text{ is asymptotically distributed as } g\left(X, \Theta_n\right)\text{”}.$$

EXAMPLE: In the last example, we demonstrated that if

$$X_n \sim \chi^2_{(n)}, \quad \text{then} \quad W_n = \tfrac{X_n - n}{\sqrt{2n}} \xrightarrow{d} W \sim \mathrm{N}(0,1)$$

Now consider the random variable

$$Y_n = g(W_n, n) = \sqrt{2n} \cdot W_n + n.$$

According to the definition, the asymptotic distribution of Y_n obtains as

$$Y_n = g(W_n, n) \overset{a}{\sim} g(W, n) = \sqrt{2n} \cdot W + n \sim \mathrm{N}(n, 2n),$$

and hence $Y_n \overset{a}{\sim} \mathrm{N}(n, 2n)$. Note that Y_n has - in contrast to W_n - no limiting distribution. ||

The following theorem facilitates identification of the **limiting distribution of continuous functions** of random variables with a limiting distribution.

THEOREM 5.2 *Let* $X_n \xrightarrow{d} X$, *and let* $g(X_n)$ *be a continuous function which depends on* n *only via* X_n . *Then* $g(X_n) \xrightarrow{d} g(X)$.

For a proof, see Serfling (1980, p.24-25), *Approximation Theorems,* New York, Wiley.

EXAMPLE: Consider $Z_n \xrightarrow{d} Z \sim \mathrm{N}(0,1)$. Then

- $g(Z_n) = 2Z_n + 5 \;\; \xrightarrow{d} \;\; 2Z + 5 \sim \mathrm{N}(5,4)$;
- $g(Z_n) = Z_n^2 \;\; \xrightarrow{d} \;\; Z^2 \sim \chi^2_{(1)}$. ||

If a sequence of random variable $\{Y_n\}$ converges in probability to a random variable Y, then the realizations of Y_n are arbitrarily close to the realizations of Y with probability one, as $n \to \infty$.

DEFINITION (CONVERGENCE IN PROBABILITY): The sequence of random variables $\{Y_n\}$ *converges in probability* to the random variable Y iff

$$\lim_{n\to\infty} P\left(|y_n - y| < \epsilon\right) = 1 \quad \forall\, \epsilon > 0.$$

We denote this by $Y_n \xrightarrow{p} Y$, or plim $Y_n = Y$, where Y is called the probability limit of Y_n.

REMARK : The definition implies that if n is large enough, observing outcomes of Y_n is essentially equivalent to observing outcomes of Y.

Also note that the probability limit Y can be a *degenerate random variable* with $Y = c$, where c is a constant. We denote this by $Y_n \xrightarrow{p} c$. $\diamondsuit$

EXAMPLE: Consider the random variable Y_n with pdf

$$f_n(y) = \tfrac{1}{n}\mathbb{I}_{\{0\}}(y) + \left(1 - \tfrac{1}{n}\right)\mathbb{I}_{\{1\}}(y) \quad \underset{n\to\infty}{\rightarrow} \quad \mathbb{I}_{\{1\}}(y).$$

Hence we have $P(|Y_n - 1| = 0) \to 1$ as $n \to \infty$, so that

$$\lim_{n\to\infty} P(|y_n - 1| < \epsilon) = 1 \quad \forall\, \epsilon > 0, \qquad \text{and} \qquad \operatorname{plim} Y_n = 1. \qquad ||$$

EXAMPLE: Let $Y \sim \mathrm{N}(0,1)$ and $Z_n \sim \mathrm{N}(0, \frac{1}{n})$, assume Y and Z_n are independent. Consider

$$Y_n = Z_n + Y \quad \sim \quad \mathrm{N}(0\,,\ [1 + \tfrac{1}{n}]).$$

Since $Y_n - Y = Z_n$, we obtain

$$\lim_{n\to\infty} P(|Y_n - Y| < \epsilon) = \underbrace{\lim_{n\to\infty} P(|Z_n| < \epsilon) \geq \lim_{n\to\infty}(1 - \frac{\mathrm{var}(Z_n)}{\epsilon^2})}_{\text{Chebyshev's Ineq.}} = 1.$$

Hence $\operatorname{plim} Y_n = Y$. ||

Difference Between Convergence in Probability and Convergence in Distribution

$Y_n \xrightarrow{d} Y$ means that the random variables $Y_{n=\infty}$ and Y have the same probability distribution.

However, it is immaterial whether outcomes of Y_n and Y are related in any way. This results from the fact that random variables with the same distribution are not necessarily the same random variables.

$Y_n \xrightarrow{p} Y$ involves the outcomes of $Y_{n=\infty}$ and Y, and not merely their distributions.

That is, for large enough n, observing outcomes of Y_n is essentially equivalent to observing outcomes of Y. Of course, this implies that $Y_{n=\infty}$ and Y must have the same probability distribution.

The following theorem facilitates identification of the **probability limit of continuous functions** of sequences of random variables.

THEOREM 5.3 *Let* $X_n \xrightarrow{p} X$, *and let* $g(X_n)$ *be a continuous function which depends on* n *only via* X_n . *Then* $plim\, g(X_n) = g(plim\, X_n) = g(X)$.

For a proof, see Serfling (1980, p.24-25), *Approximation Theorems,* New York, Wiley.

REMARK: The theorem implies that the plim operator acts analogously to the standard lim operator of real analysis. $\diamondsuit$

EXAMPLE: Let $X_n \xrightarrow{p} 3$. Then the probability limit of $Y_n = \ln(X_n) + \sqrt{X_n}$ is

$$\text{plim}\, Y_n = \ln(\text{plim} X_n) + \sqrt{\text{plim} X_n} = \ln(3) + \sqrt{3}. \qquad ||$$

The following theorem establishes further useful properties of the plim operator, which obtain as special cases of Theorem 5.3.

THEOREM 5.4 *For the sequences of random variables X_n, Y_n, and the constant a.*

a. *plim* $(aX_n) = a$ *(plim* X_n*);*
b. *plim* $(X_n + Y_n) =$ *plim* X_n+*plim* Y_n *(the plim of a sum = the sum of the plims);*
c. *plim* $(X_n Y_n) =$ *plim* X_n *plim* Y_n *(the plim of a product = the product of the plims);*
d. *plim* $(X_n/Y_n) =$ *(plim* X_n*)/(plim* Y_n*).*

PROOF: All results follow from Theorem 5.3 and from the fact that the functions being considered are continuous and depend on n only via X_n and Y_n. $\square$

REMARK: The results of Theorem 5.4 extend to matrices by applying them to matrices element-by-element – see Mittelhammer (1996, p. 244-245). $\diamondsuit$

Relationship Between Convergence in Probability and Convergence in Distribution

The following theorem indicates that convergence in probability implies convergence in distribution.

THEOREM 5.5 $\quad Y_n \xrightarrow{p} Y \Rightarrow Y_n \xrightarrow{d} Y.$

PROOF: $Y_n \xrightarrow{p} Y$ implies that observing outcomes of $Y_{n=\infty}$ is equivalent to observing outcomes of Y. Hence $Y_{n=\infty}$ and Y must have the same probability distribution. For a formal proof, see Mittelhammer (1996, p. 246). $\square$

EXAMPLE: Let $Y_n = (2 + \frac{1}{n})X + 3$, where $X \sim \mathrm{N}(1, 2)$. Then the plim of Y_n is

$$\mathrm{plim}\, Y_n = \mathrm{plim}(2 + \tfrac{1}{n})\mathrm{plim}X + \mathrm{plim}3 = 2X + 3 = Y \sim \mathrm{N}(5, 8).$$

The fact that $Y_n \xrightarrow{p} Y \sim \mathrm{N}(5, 8)$ implies according to Theorem 5.5 that $Y_n \xrightarrow{d} Y \sim \mathrm{N}(5, 8)$. ||

The converse of Theorem 5.5 is generally not true. However, in the special case where we have convergence in distribution to a constant, the converse of Theorem 5.5 does hold as stated in the following theorem.

THEOREM 5.6 $Y_n \xrightarrow{d} c \Rightarrow Y_n \xrightarrow{p} c.$

PROOF: Suppose that $Y_n \xrightarrow{d} c$. This implies for the cdf F_n of the sequence Y_n that $F_n(y) \to F(y) = \mathbb{I}_{[c,\infty)}(y)$. Then as $n \to \infty$,

$$P(|y_n - c| < \epsilon) \geq \underbrace{F_n(c+\tau)}_{\to 1} - \underbrace{F_n(c-\tau)}_{\to 0} \to 1, \quad \text{for } \tau \in (0,\epsilon) \text{ and } \forall \epsilon > 0.$$

It follows that $Y_n \xrightarrow{p} c$. $\square$

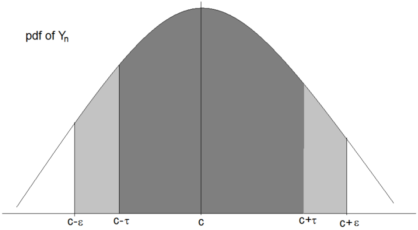

Fig. 29.

The following theorem combines the concepts of convergence in distribution and in probability to produce an extension of Theorem 5.2 that is useful for obtaining the limiting distribution of a wider variety of functions of random variables.

Theorem 5.7 (Slutsky's Theorems) *Let* $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{p} c$. *Then,*

a. $X_n + Y_n \xrightarrow{d} X + c$;
b. $X_n \cdot Y_n \xrightarrow{d} X \cdot c$;
c. $X_n / Y_n \xrightarrow{d} X / c$.

For a formal proof, see Rohatgi and Saleh (2001, p. 270).

Definition (Convergence in Mean Square): The sequence of random variables $\{Y_n\}$ **converges in mean square** to the random variable Y, iff

$$\lim_{n\to\infty} \mathrm{E}(Y_n - Y)^2 = 0.$$

We denote this by $Y_n \stackrel{m}{\to} Y$.

Remark: Since $\mathrm{E}(Y_n - Y)^2$ measures the squared distance, convergence in mean squared error implies that of Y_n and Y are arbitrarily close to one another when $n \to \infty$.

In particular, first- and second order moments of Y_n and Y converge to one another as indicated in the following theorem which provides necessary and sufficient conditions for convergence in mean square. $\diamondsuit$

Theorem 5.8 $Y_n \xrightarrow{m} Y$ *iff*

a. $EY_n \to EY$,

b. $var(Y_n) \to var(Y)$,

c. $cov(Y_n, Y) \to var(Y)$.

Proof: *Necessity of the conditions (a)-(c);*

a. $[\mathrm{E}Y_n \to \mathrm{E}Y]$ Note that

$$|\mathrm{E}Y_n - \mathrm{E}Y| = |\mathrm{E}(Y_n - Y)|\Big|_{(|x| \geq x \ \forall x)} \leq \mathrm{E}|Y_n - Y| = \mathrm{E}\left[(|Y_n - Y|^2)^{1/2}\right],$$

where

$$\underbrace{\mathrm{E}\left[(|Y_n - Y|^2)^{1/2}\right] \leq \left[\mathrm{E}(|Y_n - Y|^2)\right]^{1/2}}_{\text{Jensen's inequality for the convex function } x^{1/2}} = \left[\mathrm{E}(Y_n - Y)^2\right]^{1/2}.$$

Hence the fact that $\mathrm{E}(Y_n - Y)^2 \to 0$ implies that $|\mathrm{E}Y_n - \mathrm{E}Y| \to 0$ which in turn implies that $\mathrm{E}Y_n \to \mathrm{E}Y$.

(continues)

PROOF (CONTINUED):

b. [$\mathrm{Var}(Y_n) \to \mathrm{var}(Y)$] By expansion of $\mathrm{E}Y_n^2$ we obtain

$$\begin{aligned} \mathrm{E}Y_n^2 &= \mathrm{E}(Y_n^2 - 2YY_n + Y^2) + \mathrm{E}Y^2 + 2\mathrm{E}(YY_n) - 2\mathrm{E}Y^2 \\ &= \mathrm{E}(Y_n - Y)^2 \quad + \quad \mathrm{E}Y^2 + 2\mathrm{E}[Y(Y_n - Y)], \end{aligned}$$

where the last term is bounded by

$$\underbrace{|\mathrm{E}[Y(Y_n - Y)]| \leq \left[\mathrm{E}Y^2\,\mathrm{E}(Y_n - Y)^2\right]^{1/2}}_{\text{Cauchy-Schwartz inequality } (\mathrm{E}WZ)^2 \leq \mathrm{E}W^2\mathrm{E}Z^2, \text{ i.e., } |\mathrm{E}WZ| \leq (\mathrm{E}W^2\mathrm{E}Z^2)^{1/2}}.$$

It follows that

$$\mathrm{E}Y_n^2 \in \left[\mathrm{E}(Y_n - Y)^2 + \mathrm{E}Y^2 \;\pm\; 2\left[\mathrm{E}Y^2\,\mathrm{E}(Y_n - Y)^2\right]^{1/2}\right].$$

Hence the fact that $\mathrm{E}(Y_n - Y)^2 \to 0$ implies that $\mathrm{E}Y_n^2 \in [0 + \mathrm{E}Y^2 \pm 2 \cdot 0]$, which together with (a) implies that $\mathrm{var}(Y_n) \to \mathrm{var}(Y)$.

(CONTINUES)

PROOF (CONTINUED):

c. $[\text{cov}(Y_n, Y) \to \text{var}(Y)]$ Note that

$$\begin{aligned} \text{E}(Y_n - Y)^2 &= \text{E}Y_n^2 - 2\text{E}(YY_n) + \text{E}Y^2 \\ &= \text{var}(Y_n) + (\text{E}Y_n)^2 - 2[\text{cov}(Y_n, Y) + \text{E}Y\text{E}Y_n] + \text{var}(Y) + (\text{E}Y)^2. \end{aligned}$$

If $\text{E}(Y_n - Y)^2 \to 0$, with (a) and (b) we obtain from the preceding equality that

$$0 = 2\text{var}(Y) - 2\lim \text{cov}(Y_n, Y),$$

which implies $\text{cov}(Y_n, Y) \to \text{var}(Y)$.

Sufficiency of the conditions (a)-(c);

As shown above, we have

$$\text{E}(Y_n - Y)^2 = \text{var}(Y_n) + (\text{E}Y_n)^2 - 2[\text{cov}(Y_n, Y) + \text{E}Y\text{E}Y_n] + \text{var}(Y) + (\text{E}Y)^2.$$

This shows directly that conditions (a)-(c), imply that $\text{E}(Y_n - Y)^2 \to 0$. $\square$

The necessary and sufficient conditions in Theorem 5.8 simplify, when Y is a constant, as stated in the following corrolary.

Corrolary 5.1 $Y_n \stackrel{m}{\rightarrow} c$ *iff* $EY_n \rightarrow c$ *and* $var(Y_n) \rightarrow 0$.

Proof: This follows directly from Theorem 5.8, upon letting $Y = c$ and noting that $\mathrm{E}(c) = c$ and $\mathrm{var}(c) = \mathrm{cov}(Y_n, c) = 0$. $\square$

The following theorem indicates that convergence in mean square implies convergence in probability.

This result can be useful as a tool for establishing the convergence in probability in cases where convergence in mean square is relatively easy to demonstrate.

THEOREM 5.9 $Y_n \stackrel{m}{\to} Y \Rightarrow Y_n \stackrel{p}{\to} Y.$

PROOF: Note that $(Y_n - Y)^2$ is a non-negative random variable, and letting $a = \epsilon^2 > 0$, we have by Markov's inequality that

$$P([y_n - y]^2 \geq \epsilon^2) \leq \frac{E(Y_n - Y)^2}{\epsilon^2},$$

or for the complementary event

$$P([y_n - y]^2 < \epsilon^2) \geq 1 - \frac{E(Y_n - Y)^2}{\epsilon^2}.$$

Thus, mean square convergence, which means $E(Y_n - Y)^2 \to 0$, implies

$$\lim_{n \to \infty} P(|y_n - y| < \epsilon) = 1.$$

Thus, $\operatorname{plim} Y_n = Y$. $\square$

EXAMPLE: Let $Y_n = \frac{1}{n}\sum_{i=1}^n X_i$, where the X_is are independently N(0, 1)-distributed. We have

$$\mathrm{E}Y_n = 0 \qquad \text{and} \qquad \mathrm{var}(Y_n) = \tfrac{1}{n} \;\to\; 0,$$

which implies that $Y_n \xrightarrow{m} 0$, so that according to Theorem 5.9, $Y_n \xrightarrow{p} 0$. ||

The following example demonstrates that the converse of Theorem 5.9 is generally not true. That means that convergence in probability *does not* imply mean square convergence.

EXAMPLE: Let the pdf of Y_n be given by,

$$f_n(y) = \begin{cases} 1 - \frac{1}{n^2} & \text{for} \quad y_n = 0 \\ \frac{1}{n^2} & \text{for} \quad y_n = n \end{cases}.$$

It immediately follows that $P(y_n = 0) \to 1$ so that $\operatorname{plim} Y_n = 0$. However,

$$\mathrm{E}(Y_n - 0)^2 \;=\; \mathrm{E}Y_n^2 \;=\; 0^2 \cdot (1 - \tfrac{1}{n^2}) + n^2 \cdot \tfrac{1}{n^2} \;=\; 1 \quad \forall n,$$

so that $Y_n \overset{m}{\nrightarrow} 0$. ||

In this section, we consider the convergence behavior of a specific sequence of random variable, namely, the **mean of a sequence of random variables**

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad n = 1, 2, 3, \ldots.$$

The convergence behavior of such a specific sequence deserves special attention, since a large number of parameter-estimation and hypothesis-testing procedures in econometrics can be defined in terms of averages of random variables.

Hence, the analysis of the convergence behavior of averages is useful for analyzing the asymptotic behavior of these procedures.

Definition (Weak Law of Large Numbers): Let $\{X_n\}$ be a sequence of random variables with finite expected values $EX_n = \mu_n$. We say that $\{X_n\}$ obeys the weak law of large numbers (WLLN), if

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) = \bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0.$$

Remark: For $EX_i = \mu_i = \mu \ \forall i$ the WLLN implies that

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \xrightarrow{p} 0 \qquad \text{and} \qquad \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu,$$

such that the (sample) average converges in probability to the expectation of the random variables. $\diamondsuit$

There are a variety of conditions that can be placed on the stochastic behavior of the variables in the sequence $\{X_n\}$ that ensure that they obey the WLLN. These conditions typically relate to *independence*, *variance* and *covariances* of the variables in $\{X_n\}$.

A WLLN which is based upon the condition that the variables in the sequence $\{X_n\}$ are iid is **Khinchin's WLLN**, as follows.

THEOREM 5.10 (KHINCHIN'S WLLN) *Let $\{X_n\}$ be a sequence of iid random variables with finite expectations $EX_n = \mu\ \forall i$. Then $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$.*

PROOF: Our proof is based on the additional assumption that the MGF of X_i, denoted by $M_{X_i}(t)$, exists[1]. The MGF of $\bar{X}_n$ obtains as (see Section 3.5)

$$\begin{aligned}
M_{\bar{X}_n}(t) &\overset{\text{(def.)}}{=} \mathrm{E}\left[\exp\left\{t \cdot \tfrac{1}{n}\textstyle\sum_{i=1}^{n} X_i\right\}\right] = \mathrm{E}\left[\textstyle\prod_{i=1}^{n} \exp\left\{\tfrac{t}{n} X_i\right\}\right] \\
&= \prod_{i=1}^{n} \mathrm{E}\left[\exp\left\{\frac{t}{n} X_i\right\}\right] \qquad \text{(by independence of the } X_i\text{s)} \\
&= \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right) \qquad \text{(by def.)} \\
&= \left[M_{X_i}\left(\frac{t}{n}\right)\right]^n . \qquad \text{(identical distr. of the } X_i\text{s)}
\end{aligned}$$

(CONTINUES)

[1] For a more general proof, see Rao (1973), *Linear Statistical Inference and Its Applications.* New York: John Wiley & Sons

PROOF (CONTINUED): For $n \to \infty$, we get for $M_{\bar{X}_n}(t) = [M_{X_i}(\frac{t}{n})]^n$

$$\begin{aligned}
\lim_{n\to\infty} M_{\bar{X}_n}(t) &= \lim_{n\to\infty} \left[1 + \frac{n \cdot M_{X_i}(\frac{t}{n}) - n}{n}\right]^n \qquad \text{(by extending } M_{X_i}(t/n)) \\
&= \exp\left\{\lim_{n\to\infty} \left[n \cdot \left(M_{X_i}(\frac{t}{n}) - 1\right)\right]\right\} \qquad (\text{since } \lim_{n\to\infty}[1 + \frac{a_n}{n}]^n = exp\{\lim_{n\to\infty} a_n\}).
\end{aligned}$$

For the limit in the exponent we have

$$\lim_{n\to\infty} \left[\frac{M_{X_i}(\frac{t}{n}) - 1}{n^{-1}}\right] = \frac{0}{0}.$$

By L'Hospital's rule we obtain

$$\begin{aligned}
\lim_{n\to\infty} \left[\frac{M_{X_i}(\frac{t}{n}) - 1}{n^{-1}}\right] &= \lim_{n\to\infty} \left[\frac{d(M_{X_i}(\frac{t}{n}) - 1)/(dn)}{d(n^{-1})/(dn)}\right] \\
&= \lim_{n\to\infty} \left[\frac{\left[dM_{X_i}(\frac{t}{n})/d(\frac{t}{n})\right] \cdot (-\frac{t}{n^2})}{-n^{-2}}\right] \\
&= \lim_{n\to\infty} \left[dM_{X_i}\left(\frac{t}{n}\right)/d\left(\frac{t}{n}\right) \cdot t\right] = \mu t,
\end{aligned}$$

since the first derivative of $M_{X_i}(t^*) \to \mu$ as $t^* = \frac{t}{n} \to 0$. Thus $\lim_{n\to\infty} M_{\bar{X}_n}(t) = e^{t\mu}$.

(CONTINUES)

Proof (continued): Note that $\lim_{n\to\infty} M_{\bar{X}_n}(t) = e^{t\mu}$ is the MGF of a random variable that is degenerated at the expected value μ.

Therefore, by Theorem 5.1 (Convergence of MGFs) and Theorem 5.5 ($X_n \xrightarrow{d} c \;\Rightarrow\; X_n \xrightarrow{p} c$), we have $\bar{X}_n \xrightarrow{p} \mu$. $\square$

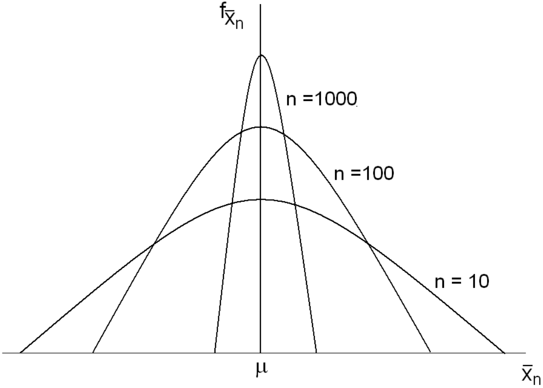

Fig. 30.

EXAMPLE: Let $\{X_n\}$ be a sequence of iid random variables, with $X_n \sim \text{Gamma}(\alpha, \beta)$ such that $\mathrm{E}X_n = \alpha\beta$. Khinchin's WLLN implies that

$$\bar{X}_n \xrightarrow{p} \mathrm{E}X_n = \alpha\beta.$$

Hence, for large enough n, the outcome of the random variable $\bar{X}_n$ can be taken as a close approximation of $\alpha\beta$.

This is *the* property of a *consistent estimator* for $\alpha\beta$ as we shall discuss in the course *Advanced Statistics II*. $||$

REMARK: Note that Khinchin's WLLN does not require the existence of the variance of the variables in the sequence $\{X_n\}$.

On the other hand, it requires that the variables are iid. This is too restrictive for many situations we encounter, where the variables are not iid.

WLLNs that relax the iid assumption can be defined by imposing various other conditions on the (co)variances of the X_ns.

The next theorem states necessary and sufficient conditions for a sequence $\{X_n\}$ to satisfy the WLLN. $\diamondsuit$

Theorem 5.11 *Let $\{X_n\}$ be a sequence of random variables with finite variances, and let $\{\mu_n\}$ be the corresponding sequence of their expectations, Then*

$$\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0 \qquad iff \qquad E\left[\frac{(\bar{X}_n - \bar{\mu}_n)^2}{1 + (\bar{X}_n - \bar{\mu}_n)^2}\right] \to 0.$$

Proof: *Sufficiency*; For any two positive numbers a and b, the fact that

$$a \geq b \qquad \text{implies that} \qquad \frac{a}{1+a} \geq \frac{b}{1+b}.$$

Hence the event

$$(\bar{X}_n - \bar{\mu}_n)^2 \geq \epsilon^2 \qquad \text{implies the event} \qquad \frac{(\bar{X}_n - \bar{\mu}_n)^2}{1 + (\bar{X}_n - \bar{\mu}_n)^2} \geq \frac{\epsilon^2}{1+\epsilon^2}.$$

It follows that[2]

$$\begin{aligned} P\{(\bar{X}_n - \bar{\mu}_n)^2 \geq \epsilon^2\} &\leq P\left\{\underbrace{\frac{(\bar{X}_n - \bar{\mu}_n)^2}{1 + (\bar{X}_n - \bar{\mu}_n)^2}}_{\text{pos. random variable}} \geq \frac{\epsilon^2}{1+\epsilon^2}\right\} \\ &\leq \mathrm{E}\left[\frac{(\bar{X}_n - \bar{\mu}_n)^2}{1 + (\bar{X}_n - \bar{\mu}_n)^2}\right] \Big/ \left[\frac{\epsilon^2}{1+\epsilon^2}\right] \qquad \text{(by Markov's inequality)} \end{aligned}$$

If the expectation of the term in brackets $\to 0$ as $n \to \infty$, then, $\forall \epsilon > 0$, the probability on the l.h.s. $\to 0$. That is

(continues)

[2] Note that the event of tossing a '2' implies the event of tossing an 'even number'.

PROOF (CONTINUED):

$$P\{(\bar{X}_n - \bar{\mu}_n)^2 \geq \epsilon^2\} = P\{|\bar{X}_n - \bar{\mu}_n| \geq \epsilon\} \to 0,$$

and for the complementary event

$$P\{|\bar{X}_n - \bar{\mu}_n| < \epsilon\} \to 1, \qquad \text{so that} \quad \bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0.$$

Necessity; See Rohatgi and Saleh (2001, p. 276). $\square$

REMARK: The theorem states that $\{X_n\}$ obeys a WLLN such that $\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0$ iff the condition that $\mathrm{E}[(\bar{X}_n - \bar{\mu}_n)^2/(1 + [\bar{X}_n - \bar{\mu}_n]^2)] \to 0$ is satisfied .

But since this condition applies not to the individual variables in $\{X_n\}$, but to their average, Theorem 5.11 is of limited use.

However, any condition placed on the individual variables in $\{X_n\}$ that ensures that $\mathrm{E}[(\bar{X}_n - \bar{\mu}_n)^2/(1 + [\bar{X}_n - \bar{\mu}_n]^2)] \to 0$ is sufficient to guarantee that $\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0$.

The following theorem identifies one such condition. $\diamondsuit$

Theorem 5.12 *Let $\{X_n\}$ be a sequence of random variables with respective expectations given by $\{\mu_n\}$. If*

$$var(\bar{X}_n) \to 0, \quad then \quad \bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0.$$

Proof: Note that

$$0 \leq \mathrm{E}\left[\frac{(\bar{X}_n - \bar{\mu}_n)^2}{1 + (\bar{X}_n - \bar{\mu}_n)^2}\right] \leq \mathrm{E}\left[(\bar{X}_n - \bar{\mu}_n)^2\right] = \mathrm{var}(\bar{X}_n).$$

Hence, $\mathrm{var}(\bar{X}_n) \to 0$ implies that $\mathrm{E}[(\bar{X}_n - \bar{\mu}_n)^2/(1 + [\bar{X}_n - \bar{\mu}_n]^2)] \to 0$, and it follows from Theorem 5.11 that $\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0$. $\square$

EXAMPLE: Let $\{X_n\}$ be a sequence random variables, with

$$\mathrm{E}X_i = \mu_i = \frac{1}{2^i}\ , \qquad \mathrm{var}(X_i) = 4, \quad \text{and} \quad \mathrm{cov}(X_i, X_j) = 0 \quad \forall i \neq j.$$

The mean and variance for the average $\bar{X}_n$ are

$$\mathrm{E}\bar{X}_n \;=\; \frac{1}{n}\sum_{i=1}^{n}\mu_i \;=\; \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2^i} \;=\; \frac{1-(\frac{1}{2})^n}{n}, \qquad \mathrm{var}(\bar{X}_n) \;=\; \frac{4}{n}.$$

Since $\mathrm{var}(\bar{X}_n) \to 0$, it follows by Theorem 5.12, that

$$\bar{X}_n - \bar{\mu}_n \;=\; \bar{X}_n - \frac{1-(\frac{1}{2})^n}{n} \quad \xrightarrow{p} \quad 0.$$

Also note that

$$\bar{\mu}_n \;=\; \frac{1-(\frac{1}{2})^n}{n} \quad \to \quad 0, \qquad \text{so that} \qquad \bar{X}_n \quad \xrightarrow{p} \quad 0.$$

||

Central limit theorems (CLTs) are concerned with the conditions under which sequences of random variables **converge in distribution** to known families of distribution.

Let $\{X_n\}$ be a sequence of random variables, and let $S_n = \sum_{i=1}^{n} X_i$, $n = 1, 2, ..$ Here we focus on the convergence in distribution of sequences of random variables of the following form

$$b_n^{-1}(S_n - a_n) \xrightarrow{d} Y \sim \mathrm{N}(0, \Sigma),$$

where $\{a_n\}$ and $\{b_n\}$ are sequences of appropriately chosen real constants.

A statement of conditions on $\{X_n\}$, $\{a_n\}$, and $\{b_n\}$ that ensure the convergence in distribution result constitutes a CLT.

What are the reasons to study the particular problem concerning the convergence in distribution specified above?

- Many procedures for parameter estimation and hypothesis testing are specified as functions of sums of random variables such as $S_n = \sum_{i=1}^{n} X_i$.
- CLTs are then often useful for establishing the asymptotic distributions for those procedures.

  Recall that asymptotic distributions provide approximations to the exact but often unknown distribution.

Similar to the case of the WLLN, there are a variety of conditions that can be placed on the stochastic behavior of the variables in the sum $S_n = \sum_{i=1}^{n} X_i$ that ensure the convergence in distribution according to a CLT.

The simplest, but least general, of all CLTs is the **Lindberg-Levy CLT**, which assumes iid variables.

THEOREM 5.13 (LINDBERG-LEVY) *Let $\{X_n\}$ be a sequence of iid random variables with $EX_i = \mu$ and $var(X_i) = \sigma^2 \in (0, \infty)\ \forall i$. Then*

$$Y_n = \frac{1}{\sqrt{n}\sigma}\left(\sum_{i=1}^{n} X_i - n\mu\right) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1).$$

PROOF: Our proof is based on the additional assumption that the MGF of X_i, denoted by $M_{X_i}(t)$, exists[3].

Consider the standardized variable

$$Z_i = (X_i - \mu)/\sigma, \qquad \text{which is iid with} \qquad \mathrm{E}Z_i = 0, \quad \mathrm{var}(X_i) = 1.$$

Now, rewrite Y_n as

$$Y_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu)/\sigma = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i.$$

(CONTINUES)

[3] For a more general proof, see Rao (1973), *Linear Statistical Inference and Its Applications.* New York: John Wiley & Sons

PROOF (CONTINUED): The MGF of $Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i$ obtains as (see Section 3.5)

$$M_{Y_n}(t) \overset{\text{(def.)}}{=} \mathrm{E}\left[\exp\{t \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i\}\right] \overset{\text{(independ.)}}{=} \prod_{i=1}^{n} M_{Z_i}\left(\frac{t}{\sqrt{n}}\right) \overset{\text{(ident. distr.)}}{=} \left[M_{Z_i}\left(\frac{t}{\sqrt{n}}\right)\right]^n,$$

and its log-transformation is

$$\ln M_{Y_n}(t) = \frac{\ln M_{Z_i}\left(\frac{t}{\sqrt{n}}\right)}{n^{-1}} = \frac{L\left(\frac{t}{\sqrt{n}}\right)}{n^{-1}}, \qquad \text{where} \quad L(\cdot) = \ln M_{Z_i}(\cdot).$$

Since $\ln M_{Z_i}(0) = \ln(1) = 0$, the limit $\lim_{n\to\infty} \ln M_{Y_n}(t)$ has the indeterminate form $0/0$. Applying L'Hospital's rule yields

$$\lim_{n\to\infty} \ln M_{Y_n}(t) = \lim_{n\to\infty} \left[\frac{\left[dL(\frac{t}{\sqrt{n}})/d(\frac{t}{\sqrt{n}})\right](-\frac{1}{2})(\frac{t}{n^{3/2}})}{-n^{-2}}\right] = \lim_{n\to\infty} \left[\frac{\left[dL(\frac{t}{\sqrt{n}})/d(\frac{t}{\sqrt{n}})\right] t}{2/\sqrt{n}}\right],$$

$$\text{with} \qquad dL(0)/d(\tfrac{t}{\sqrt{n}}) = \frac{1}{M_{Z_i}(0)} \frac{dM_{Z_i}(0)}{d(\frac{t}{\sqrt{n}})} = 1 \cdot \mathrm{E}Z_i = 0$$

A second application of L'Hospital's rule yields

(CONTINUES)

Proof (continued):

$$\lim_{n\to\infty} \ln M_{Y_n}(t) = \lim_{n\to\infty} \left[\frac{\left[d^2 L(\frac{t}{\sqrt{n}})/d(\frac{t}{\sqrt{n}})^2\right](-\frac{1}{2})(\frac{t}{n^{3/2}})t}{-\frac{1}{n^{3/2}}} \right] = \lim_{n\to\infty} \left[\frac{\left[d^2 L(\frac{t}{\sqrt{n}})/d(\frac{t}{\sqrt{n}})^2\right] t^2}{2} \right].$$

Since

$$\begin{aligned} d^2 L(0)/d(\tfrac{t}{\sqrt{n}})^2 &= \frac{1}{M_{Z_i}(0)} \frac{d^2 M_{Z_i}(0)}{d(\frac{t}{\sqrt{n}})^2} - \frac{1}{[M_{Z_i}(0)]^2} \left[\frac{dM_{Z_i}(0)}{d(\frac{t}{\sqrt{n}})} \right]^2 \\ &= 1 \cdot \mathrm{E} Z_i^2 - 1 \cdot (\mathrm{E} Z_i)^2 = 1, \end{aligned}$$

we obtain

$$\lim_{n\to\infty} \ln M_{Y_n}(t) = \frac{1}{2} t^2.$$

Thus

$$\lim_{n\to\infty} M_{Y_n}(t) = \exp\left\{ \lim_{n\to\infty} \ln M_{Y_n}(t) \right\} = \exp\left\{ \frac{1}{2} t^2 \right\}.$$

Since $M(t) = e^{t^2/2}$ is the MGF of a $\mathrm{N}(0,1)$ distribution, we have that $Y_n \xrightarrow{d} \mathrm{N}(0,1)$. □

REMARK: In order to understand the 'mechanism' behind the Lindberg-Levy CLT, consider the random variable

$$\sum_{i=1}^n X_i, \quad \text{with} \quad \mathrm{E}(\sum_{i=1}^n X_i) = n\mu \quad \text{and} \quad \mathrm{var}(\sum_{i=1}^n X_i) = n\sigma^2.$$

Note that $\sum_{i=1}^n X_i$ does not have a limiting distribution as its mean and variance diverge to ∞ as n increases.

Hence, some form of centering and scaling of $\sum_{i=1}^n X_i$ is necessary, for obtaining some limiting distribution.

An appropriate centering and scaling which stabilize the mean and variance, obtains by defining the random variable

$$Y_n = \underbrace{\frac{1}{\sqrt{n}\sigma}}_{\text{scaling}} [\sum_{i=1}^n X_i - \underbrace{n\mu}_{\text{centering}}] = \tfrac{1}{\sqrt{n}} \sum_{i=1}^n Z_i,$$

$$\text{such that} \qquad \mathrm{E}Y_n = 0 \quad \text{and} \quad \mathrm{var}(Y_n) = 1 \quad \forall n \text{ (!)}.$$

An additional effect of this centering and scaling is that it removes any tendencies for higher order moments of the X_i's to cause the moments of Y_n to deviate form those of a $N(0,1)$ distribution as $n \to \infty$.

(CONTINUES)

REMARK (CONTINUED): To see this consider the third moment of Y_n, which is (see Mittelhammer, 1996, p. 271)

$$\mathrm{E}Y_n^3 = \frac{1}{n^{3/2}\sigma^2}\mathrm{E}Z_i^3 = \frac{1}{n^{3/2}\sigma^2}\mathrm{E}\left[\left(\frac{X_i-\mu}{\sigma}\right)^3\right].$$

Thus $\mathrm{E}Y_n^3 \to 0$ as $n \to \infty$ regardless of the specific value of $\mathrm{E}Z_i^3 = \mathrm{E}(\frac{X_i-\mu}{\sigma})^3$. Recall that $\mathrm{E}W^3 = 0$ if $W \sim \mathrm{N}(0,1)$.

Now consider the fourth moment of Y_n, which is

$$\mathrm{E}Y_n^4 = \frac{1}{n^2\sigma^4}[n\mathrm{E}Z_i^4 + 3n(n-1)\sigma^4],$$

so that $\mathrm{E}Y_n^4 \to 3$ as $n \to \infty$ regardless of the specific value of $\mathrm{E}Z_i^4 = \mathrm{E}(\frac{X_i-\mu}{\sigma})^4$. Recall that $\mathrm{E}W^4 = 3$ if $W \sim \mathrm{N}(0,1)$.

This type of argument can be continued ad infinitum to show that all moments of Y_n converge to their Gaussian values[4]. $\diamondsuit$

[4] This logic can be exploited in order to proof the Lindberg-Levy CLT by means of a Taylor series expansion of the MGF of Y_n (see Rohatgi and Saleh, 2001, p. 297).

REMARK: The Lindberg-Levy CLT states that

$$Y_n = \frac{1}{\sqrt{n}\sigma}\Big(\sum_{i=1}^n X_i - n\mu\Big) \xrightarrow{d} Y \sim \mathrm{N}(0,1).$$

This **limiting distribution** can be used to obtain **asymptotic distributions** for functions of Y_n.

- For the variable $S_n = \sum_{i=1}^n X_i$, for example, we obtain

$$S_n = \sqrt{n}\sigma Y_n + n\mu\Big|_{\text{def.}} \overset{a}{\sim} \sqrt{n}\sigma Y + n\mu \qquad \Rightarrow \qquad S_n \overset{a}{\sim} \mathrm{N}(n\mu, n\sigma^2).$$

- For the average $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ we have

$$\bar{X}_n = \frac{\sigma}{\sqrt{n}} Y_n + \mu\Big|_{\text{def.}} \overset{a}{\sim} \frac{\sigma}{\sqrt{n}} Y + \mu \qquad \Rightarrow \qquad \bar{X}_n \overset{a}{\sim} \mathrm{N}(\mu, \tfrac{\sigma^2}{n}).$$

Note that according to the WLLN $\bar{X}_n \xrightarrow{p} \mu$, and hence $\bar{X}_n \xrightarrow{d} \mu$. Because the limiting distribution of $\bar{X}_n$ is *degenerate* it provides no information about the variability of $\bar{X}_n$ for finite n and is useless for establishing an approximative distribution. $\diamondsuit$

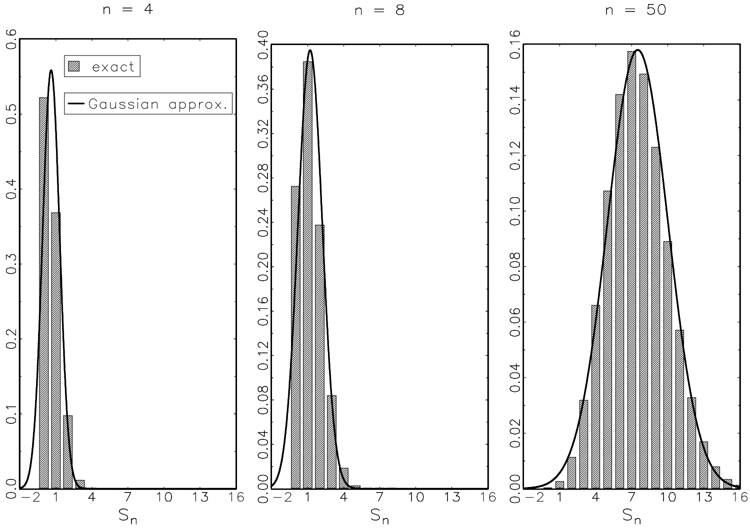

Fig. 31. Exact and asymptotic distribution of $S_n = \sum_{i=1}^{n} X_i$ for $x_i \sim$ iid Bernoulli with $p = 0.15$. The exact distribution is a binomial(n, p); the asymptotic distribution is a $\mathrm{N}(np, np(1-p))$.

EXAMPLE: Let $\{X_n\}$ be a sequence of iid $\chi^2_{(1)}$ random variables with $\mathrm{E}X_i = 1$ and $\mathrm{var}(X_i) = 2$.

- Remember that by the additivity property of χ^2-distribution (Theorem 4.2), $\sum_{i=1}^n X_i \sim \chi^2_{(n)}$.
- By the CLT of Lindberg-Levy, we have

$$Y_n = \frac{\sum_{i=1}^n X_i - n \cdot 1}{\sqrt{n2}} \xrightarrow{d} \mathrm{N}(0,1) \qquad \Rightarrow \qquad \sum_{i=1}^n X_i \overset{a}{\sim} \mathrm{N}(n, 2n).$$

This implies, that we can approximate the pdf of a $\chi^2_{(n)}$-distribution by a Gaussian pdf if the d.o.f. is large. ||

REMARK: The CLT of Lindberg-Levy requires that the **random variables are iid**.

- However, in many applications the assumption that the variables are iid is violated since we have variables which are correlated and/or have different distributions.
- Fortunately, there are various other CLTs, which do not need the iid condition. Instead, they place alternative conditions on the stochastic behavior of the random variables in the sequence $\{X_n\}$.

  In the following, we consider such **CLTs for variables which are not iid**. The results are presented without proofs[5]. $\diamondsuit$

[5] For the proofs, see Mittelhammer, (1996, p. 274-282)

A CLT for **non-identically distributed** random variables is that of **Lindberg**. It is based on the condition that the contribution that each variable X_i makes to the variance of $\sum_{i=1}^{n} X_i$ is negligible as $n \to \infty$.

THEOREM 5.14 (LINDBERG'S CLT) *Let $\{X_n\}$ be a sequence of independent random variables with $EX_i = \mu_i$ and $var(X_i) = \sigma_i^2 < \infty \, \forall \, i$. Define $b_n^2 = \sum_{i=1}^{n} \sigma_i^2$, $\bar{\sigma}_n^2 = n^{-1} \sum_{i=1}^{n} \sigma_i^2$, $\bar{\mu}_n = n^{-1} \sum_{i=1}^{n} \mu_i$, and let f_i be the PDF of X_i. If $\forall \, \varepsilon > 0$,*

$$\text{(continuous case:)} \qquad \lim_{n \to \infty} \frac{1}{b_n^2} \sum_{i=1}^{n} \int_{(x_i - \mu_i)^2 \geq \varepsilon b_n^2} (x_i - \mu_i)^2 f_i(x_i) \, dx_i = 0,$$

$$\text{(discrete case:)} \qquad \lim_{n \to \infty} \frac{1}{b_n^2} \sum_{i=1}^{n} \sum_{\substack{(x_i - \mu_i)^2 \geq \varepsilon b_n^2 \\ f_i(x_i) > 0}} (x_i - \mu_i)^2 f_i(x_i) = 0,$$

then

$$\frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^{1/2}} = \frac{n^{1/2} \left(\bar{X}_n - \bar{\mu}_n\right)}{\bar{\sigma}_n} \xrightarrow{d} N(0, 1).$$

A further CLT for **non-identically distributed** random variables relies on the condition that the range of the variables are bounded.

THEOREM 5.15 (CLT FOR BOUNDED VARIABLES) *Let $\{X_n\}$ be a sequence of independent random variables such that $P(|x_i| \leq m) = 1 \ \forall \ i$ for some $m \in (0, \infty)$, and suppose $EX_i = \mu_i$ and $var(X_i) = \sigma_i^2 < \infty \, \forall \, i$. If $\sum_{i=1}^{n} var(X_i) = \sum_{i=1}^{n} \sigma_i^2 \to \infty$ as $n \to \infty$, then*

$$\frac{n^{1/2}\left(\bar{X}_n - \bar{\mu}_n\right)}{\bar{\sigma}_n} \xrightarrow{d} N(0, 1).$$

Multivariate Central Limit Theorems

The CLTs presented so far are applicable to sequences of **random scalars**.

In order to discuss CLTs for a sequence of **random vectors** a result of Cramér and Wold, termed the **Cramér-Wold device** is very useful.

The Cramér-Wold device allows to reduce the question of convergence in distribution for multivariate random vectors to the question of convergence in distribution for random scalars.

Thus it facilitates the use of CLTs for random scalars in order to obtain multivariate CLTs.

THEOREM 5.16 (CRAMÉR-WOLD DEVICE) *The sequence of* $(k \times 1)$*-dim. random vectors* $\{\boldsymbol{X}_n\}$ *converges in distribution to the* $(k \times 1)$*-dim. random vector* $\boldsymbol{X}$ *iff*

$$\boldsymbol{\ell}'\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{\ell}'\boldsymbol{X} \qquad \forall \ \boldsymbol{\ell} \in \mathbb{R}^k.$$

PROOF: *Sufficiency*; Our proof of sufficiency assumes the existence of MGFs. By Theorem 5.1 (Convergence of MGFs), the fact that

$$\boldsymbol{\ell}'\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{\ell}'\boldsymbol{X} \qquad \text{implies that} \qquad M_{\boldsymbol{\ell}'\boldsymbol{X}_n}(t) \to M_{\boldsymbol{\ell}'\boldsymbol{X}}(t).$$

The MGFs obtain as

$$\begin{aligned} M_{\boldsymbol{\ell}'\boldsymbol{X}_n}(t) &\overset{\text{(def.)}}{=} \underbrace{\mathsf{E}e^{t\cdot\boldsymbol{\ell}'\boldsymbol{X}_n} = \mathsf{E}e^{(t^*)'\cdot\boldsymbol{X}_n}}_{t\,\cdot\,\boldsymbol{\ell}' \,=\, (t^*)'} = M_{\boldsymbol{X}_n}(t^*), \\ M_{\boldsymbol{\ell}'\boldsymbol{X}}(t) &= M_{\boldsymbol{X}}(t^*). \end{aligned}$$

Hence, if $M_{\boldsymbol{\ell}'\boldsymbol{X}_n}(t) \to M_{\boldsymbol{\ell}'\boldsymbol{X}}(t)$, then $M_{\boldsymbol{X}_n}(t^*) \to M_{\boldsymbol{X}}(t^*)$, which implies that $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$.

(CONTINUES)

PROOF (CONTINUED): *Necessity;* Since $\ell' \boldsymbol{X}_n$ is a continuous function g of $\boldsymbol{X}_n$, Theorem 5.2 (limiting distributions of continuous function) implies that

$$\ell' \boldsymbol{X}_n = g(\boldsymbol{X}_n) \xrightarrow{d} g(\boldsymbol{X}) = \ell' \boldsymbol{X}. \qquad \square$$

REMARK: The Cramér-Wold Device implies that to establish convergence in distribution of a vector $\boldsymbol{X}_n$ to a vector $\boldsymbol{X}$ it suffices to demonstrate that **any linear combination of $\boldsymbol{X}_n$** converges in distribution to the corresponding linear combination of $\boldsymbol{X}$.

The implications of the Cramér-Wold Device in the context of normal distributions is formalized in the following Corollary. $\diamondsuit$

COROLLARY 5.2 $\quad \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad$ *iff* $\quad \boldsymbol{\ell}'\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{\ell}'\boldsymbol{X} \sim N(\boldsymbol{\ell}'\boldsymbol{\mu}\,,\ \boldsymbol{\ell}'\boldsymbol{\Sigma}\boldsymbol{\ell}).$

REMARK: The Corollary implies that to establish convergence in distribution of a vector $\boldsymbol{X}_n$ to a **multivariate Normal distribution** it suffices to demonstrate that any linear combination of $\boldsymbol{X}_n$ converges in distribution to a **univariate Normal distribution** by using an appropriate univariate CLT.

Hence the Cramér-Wold Device allows us to extend CLTs for scalar variables to multivariate CLTs.

The following CLT is a multivariate extension of the univariate CLT of Lindberg-Levy.
$\diamondsuit$

THEOREM 5.17 (MULTIVARIATE LINDBERG-LEVY) *Let $\{\boldsymbol{X}_n\}$ be a sequence of iid $(k \times 1)$ random vectors with $E\boldsymbol{X}_i = \boldsymbol{\mu}$ and $\mathbf{Cov}(\boldsymbol{X}_i) = \boldsymbol{\Sigma}\ \forall\, i$, where $\boldsymbol{\Sigma}$ is a $(k \times k)$ positive definite matrix. Then*

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i - \boldsymbol{\mu}\Big) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}\right).$$

PROOF: Consider the linear combination of $\boldsymbol{X}_i$ given by $Z_i = \boldsymbol{\ell}'\boldsymbol{X}_i$, $(\boldsymbol{\ell} \neq \mathbf{0})$. Note that

$$Z_i \sim \text{iid}, \qquad \text{with} \qquad \mathrm{E}Z_i = \boldsymbol{\ell}'\boldsymbol{\mu} = \mu_z \quad \text{and} \quad \mathrm{var}(Z_i) = \boldsymbol{\ell}'\boldsymbol{\Sigma}\boldsymbol{\ell} = \sigma_z^2.$$

Applying the Lindberg-Levy CLT for random scalars to the iid sequence $\{Z_i\}$ yields

$$\begin{aligned} \frac{1}{\sqrt{n}\sigma_z}\Big(\sum_{i=1}^{n} Z_i - n\mu_z\Big) &= \frac{1}{\sqrt{n}\sqrt{\boldsymbol{\ell}'\boldsymbol{\Sigma}\boldsymbol{\ell}}}\boldsymbol{\ell}'\Big(\sum_{i=1}^{n}\boldsymbol{X}_i - n\boldsymbol{\mu}\Big) \\ &= \frac{\boldsymbol{\ell}'\sqrt{n}}{\sqrt{\boldsymbol{\ell}'\boldsymbol{\Sigma}\boldsymbol{\ell}}}\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i - \boldsymbol{\mu}\Big) \qquad \xrightarrow{d} \quad \mathrm{N}(0,1). \end{aligned}$$

(CONTINUES)

PROOF (CONTINUED): Then, by Slutsky's Theorem, we get

$$\underbrace{\left(\sqrt{\boldsymbol{\ell}'\boldsymbol{\Sigma}\boldsymbol{\ell}}\right)}_{\xrightarrow{p}\sqrt{\boldsymbol{\ell}'\boldsymbol{\Sigma}\boldsymbol{\ell}}} \cdot \underbrace{\frac{\boldsymbol{\ell}'\sqrt{n}}{\sqrt{\boldsymbol{\ell}'\boldsymbol{\Sigma}\boldsymbol{\ell}}}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i - \boldsymbol{\mu}\right)}_{\xrightarrow{d} \mathrm{N}(0,1)} = \boldsymbol{\ell}'\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i - \boldsymbol{\mu}\right) \quad \xrightarrow{d} \quad \mathrm{N}(0, \boldsymbol{\ell}'\boldsymbol{\Sigma}\boldsymbol{\ell}).$$

By the Cramér-Wold Device the last 'equality' is sufficient to conclude that $\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$. $\square$

REMARK:

- It follows from the multivariate Lindberg-Levy CLT that $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i \overset{a}{\sim} \mathrm{N}(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$.
- Various other multivariate CLTs can be constructed using the Cramér-Wold Device, including the multivariate extension of the Lindberg CLT (Theorem 5.14) and the CLT for bounded variables (Theorem 5.15). $\diamondsuit$

5.5 Asymptotic Distributions of Functions of Asymptotically Normally Distributed Random Variables

In this section we consider the asymptotic distribution of differentiable functions of asymptotically normally distributed random variables.

The practical use of results about this asymptotic distribution is that once the asymptotic distribution of a sequence $\{X_n\}$ is known, the asymptotic distribution of interesting functions of X_n need not be derived anew.

We will apply those results in order to obtain the asymptotic distribution for parameter estimators and test statistics (*Adv. Statistics II*).

THEOREM 5.18 (ASYMPTOTIC DISTRIBUTION OF $g(\boldsymbol{X}_n)$) *Let* $\{\boldsymbol{X}_n\}$ *be a sequence of* $(k \times 1)$ *random vectors such that* $\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu}) \xrightarrow{d} \boldsymbol{Z} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

Let $g(\boldsymbol{x})$ *be a function that has first-order partial derivatives in a neighborhood of the point* $\boldsymbol{x} = \boldsymbol{\mu}$ *that are continuous at* $\boldsymbol{\mu}$*, and suppose the gradient vector of* $g(\boldsymbol{x})$ *evaluated at* $\boldsymbol{x} = \boldsymbol{\mu}$*,*

$$\boldsymbol{G}_{(1\times k)} = \left[\partial g(\boldsymbol{\mu})/\partial x_1 \ldots \partial g(\boldsymbol{\mu})/\partial x_k\right]',$$

is not the zero vector. Then

$$\sqrt{n}\left(g\left(\boldsymbol{X}_n\right) - g\left(\boldsymbol{\mu}\right)\right) \xrightarrow{d} N(0, \boldsymbol{G\Sigma G}') \quad \text{and} \quad g\left(\boldsymbol{X}_n\right) \overset{a}{\sim} N\left(g(\boldsymbol{\mu}), n^{-1}\boldsymbol{G\Sigma G}'\right).$$

PROOF: The proof is based upon a first-order Taylor series expansion of the function $g(\boldsymbol{x})$ around the point $\boldsymbol{\mu}$. This yields (see Mittelhammer, 1996, Lemma 5.6)

(CONTINUES)

PROOF (CONTINUED):

$$g(\boldsymbol{X}_n) = g(\boldsymbol{\mu}) + \boldsymbol{G} \cdot (\boldsymbol{X}_n - \boldsymbol{\mu}) + \underbrace{[(\boldsymbol{X}_n - \boldsymbol{\mu})'(\boldsymbol{X}_n - \boldsymbol{\mu})]^{1/2} R(\boldsymbol{X}_n)}_{\text{remainder term}},$$

$$\text{where} \qquad \lim_{\boldsymbol{X}_n \to \boldsymbol{\mu}} R(\boldsymbol{X}_n) = R(\boldsymbol{\mu}) = 0.$$

Multiplying by $\sqrt{n}$ and rearranging terms obtains

$$\sqrt{n}[g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})] = \boldsymbol{G} \cdot \underbrace{[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})]}_{\xrightarrow{d} \boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma})} + \underbrace{\left\{[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})]'[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})]\right\}^{1/2} R(\boldsymbol{X}_n)}_{\xrightarrow{p} 0},$$

To see that the second term converges in probability to 0 first note that

$$\left\{[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})]'[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})]\right\}^{1/2} \xrightarrow{d} \left\{[\mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma})]'[\mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma})]\right\}^{1/2},$$

(note that this follows from Theorem 5.2 (limiting distribution of continuous functions)).

Then, the application of Slutsky's theorem to the second term yields

$$\underbrace{\left\{[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})]'[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})]\right\}^{1/2}}_{\xrightarrow{d} \left\{[\mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma})]'[\mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma})]\right\}^{1/2}} \underbrace{R(\boldsymbol{X}_n)}_{\xrightarrow{p} 0} \quad \xrightarrow{p} \quad 0.$$

(CONTINUES)

PROOF (CONTINUED): By application of Slutsky's theorem to

$$\sqrt{n}[g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})] = \boldsymbol{G} \cdot \underbrace{[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})]}_{\xrightarrow{d} \boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma})} + \underbrace{\left\{[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})]'[\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\mu})]\right\}^{1/2} R(\boldsymbol{X}_n)}_{\xrightarrow{p} 0},$$

it follows that

$$\sqrt{n}[g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})] \quad \xrightarrow{d} \quad \boldsymbol{G} \cdot \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}) = \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}'\right). \quad \square$$