Time Series Analysis

Nonstationary and Noninvertible Distribution Theory

Katsuto Tanaka

Chapter 3 Functional Central Limit Theorems

Weak convergence of a stochastic process defined on a function space is discussed. In doing so we explore various weak convergence results generically called functional central limit theorems or invariance principles. Emphasis is placed on how to apply those theorems to deal with statistics arising from nonstationary linear time series models. It turns out that, in most cases, the continuous mapping theorem is quite powerful for obtaining limiting random variables of statistics in the sense of weak convergence. In some cases, however, the continuous mapping theorem does not apply. In those cases limiting forms involve the Ito integral.

3.1. Function space C

As a sequel to the last chapter we continue to assume that the stochastic process $\{X(t)\}$ belongs to L_2. This assumption, however, needs to be strengthened for subsequent discussions. Let $C = C[0, 1]$ be the space of all real-valued continuous functions defined on $[0, 1]$ and $(C, \mathcal{B}(C))$ a *measurable space*, where $\mathcal{B}(C)$ is the σ-*field* generated by the subsets of C that are open with respect to the *uniform metric* ρ defined by

$$\rho(x, y) = \sup_{0 \le t \le 1} |x(t) - y(t)| \tag{3.1}$$

for any $x, y \in C$. The uniform metric $\rho(x, y)$ is a continuous function of x and y (Problem 1.1), that is, $|\rho(x, y) - \rho(\tilde{x}, \tilde{y})| \to 0$ as $\rho(x, \tilde{x}) \to 0$ and $\rho(y, \tilde{y}) \to 0$.

Then we assume that the stochastic process $X = \{X(t)\}$ to be treated below is a measurable mapping from an arbitrary *probability space* $(\Omega, \mathcal{F}, P)$ into C, that is, $X^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{B}(C)$. Thus we exclusively consider those stochastic processes which belong to L_2 and have continuous sample paths.

We do not extend the space C in this book to $D = D[0, 1]$, which is the space of all real-valued functions on [0,1] that are right continuous and have finite left limits. This is just because we avoid paying the cost of greater topological complexity associated with metrics that we equip D with.

The space C is known to be *complete* and *separable* under ρ, where completeness means that each fundamental sequence, which is a sequence $\{x_n(t)\}$ that satisfies $\rho(x_m, x_n) \to 0$ as $m,\ n \to \infty$, converges to some point of the space, while separability means that the space contains a countable, dense set (see, for more details, Billingsley (1968, p.220) and Problem 1.2). In this sense the space C is much like the real line, but each element of C is a function so that C is a function space and the distance of two elements of C is defined by the uniform metric ρ. Completeness and separability facilitate discussions concerning weak convergence of stochastic processes discussed subsequently. In particular it is because of separability that $\rho(X, Y) = \sup_{0 \le t \le 1} |X(t) - Y(t)|$ becomes a random variable when $\{X(t)\}$ and $\{Y(t)\}$ are stochastic processes in C (Billingsley (1968, p.25)).

The space C, however, is not *compact*, where compactness means that any sequence in the space contains a convergent subsequence. In fact, if we think of $\{x_n(t)\} = \{n\}$, where $x_n(t) = n$ $(n = 1, 2, \cdots)$ is a constant-valued element of C, it is clear that $\{n\}$ does not contain any convergent subsequence. The situation is again the same as for sequences on the real line.

When we discuss convergence in distribution, the lack of compactness of the space of distribution functions becomes serious. As an example consider a sequence $\{F_n\}$ of distribution functions defined by

$$F_n(x) = \begin{cases} 1 & x \geq n\,, \\ 0 & x < n\,. \end{cases}$$

It is evident that $F_n(x)$ converges to $G(x) \equiv 0$ for each x, but the limiting function $G(x)$ is not a distribution function. This example shows that the space of distribution functions is not compact. Thus we need a condition like compactness which will prevent mass from escaping to infinity (Shiryayev (1984, p.315)).

A further difficulty arises if we deal with weak convergence of stochastic processes in C. Even if a sequence $\{F_{n,t}\}$ of distribution functions corresponding to $\{X_n(t)\}$ converges properly for each t, it does not necessarily imply weak convergence of $\{X_n(t)\}$ as a whole. In the next section we shall give a sufficient condition for the proper convergence.

Problems

1.1 Prove that $\rho(x, y)$ is a continuous function of x and y.

1.2 Show that the space C is complete and separable under the uniform metric ρ.

3.2. Weak convergence of stochastic processes in C

The stochastic process $X = \{X(t) : 0 \leq t \leq 1\}$ induces a *probability measure* $Q = PX^{-1}$ on $(C, \mathcal{B}(C))$ by the relation

$$Q(A) = P(X \in A) = P(X^{-1}(A))\,, \qquad A \in \mathcal{B}(C)\,.$$

Similarly, any sequence $\{X_n\}$ of stochastic processes, where each stochastic process X_n is given by $\{X_n(t) : 0 \le t \le 1\}$, induces a sequence $\{Q_n\}$ of probability measures determined by

$$Q_n(A) = P(X_n \in A) = P(X_n^{-1}(A)), \quad A \in \mathcal{B}(C)\,.$$

We say that $\{X_n\}$ *converges in distribution* to X, and we write $\mathcal{L}(X_n) \to \mathcal{L}(X)$ if

$$\begin{aligned} \lim_{n\to\infty} E\{f(X_n)\} &= \lim_{n\to\infty} \int_C f(x) Q_n(dx) \\ &= \int_C f(x) Q(dx) \\ &= E\{f(X)\} \end{aligned}$$

for each f in the class of bounded, continuous real functions defined on C. There are some other equivalent conditions for $\mathcal{L}(X_n) \to \mathcal{L}(X)$ (Billingsley (1968, p.24)).

The difficulty with $\mathcal{L}(X_n) \to \mathcal{L}(X)$ is that, unlike random variables or vectors, the finite-dimensional distributions $\mathcal{L}(X_n(t_1), \cdots, X_n(t_k))$ for each finite k and each collection $0 \le t_1 < t_2 < \cdots < t_k \le 1$ by no means determine $\mathcal{L}(X)$. Namely the finite-dimensional sets do not form a *convergence determining class* in C, although they do form a *determining class* (Billingsley (1968, p.19)). An example is also found in Billingsley (1968, p.20). We require a condition referred to as *relative compactness* of $\{X_n\}$, which means that the sequence $\{Q_n\}$ of induced probability measures on $(C, \mathcal{B}(C))$ contains a subsequence which converges weakly to a probability measure on $(C, \mathcal{B}(C))$. The limiting measure need not be a member of $\{Q_n\}$. This is the reason why the adjective 'relative' comes in. The relative compactness condition is difficult to verify in general, but, by Prohorov's theorem (Billingsley (1968, p.35)), that condition is equivalent, under completeness and separability of C, to the more operational condition *'tightness'*. This condition says that, for each positive ε, there exists a compact set K such that $Q_n(K) > 1 - \varepsilon$ for all n. Tightness prohibits probability mass from escaping to infinity and stipulates that the X_n do not oscillate too violently.

The following is a fundamental theorem concerning weak convergence of $\{X_n\}$ in C (Billingsley (1968, p.54)).

Theorem 3.1. *If the finite-dimensional distributions $\mathcal{L}(X_n(t_1), \cdots, X_n(t_k))$ converge weakly to $\mathcal{L}(X(t_1), \cdots, X(t_k))$, and if $\{X_n\}$ is tight, then $\mathcal{L}(X_n) \to \mathcal{L}(X)$.*

The tightness condition can be further made operational. It is proved in Billingsley (1968, p.55) (see, also, Hall and Heyde (1980, p.275)) that $\{X_n\}$ is tight if and only if

i) $P(|X_n(0)| > a) \to 0$ uniformly in n as $a \to \infty$;

ii) for each $\varepsilon > 0$, $P(\sup_{|s-t|<\delta} |X_n(s) - X_n(t)| > \varepsilon) \to 0$ uniformly in n as $\delta \to 0$.

Note that the condition i) means tightness of $X_n(0)$. A sufficient moment condition for ii) to hold is also given in Billingsley (1968, p.95).

In subsequent sections we consider various examples of $\{X_n\}$, for which weak convergence is discussed. To this end we will not go into details, but will only describe weak convergence results useful for later chapters. Details can be found in Chapter 2 of Billingsley (1968) and Chapter 4 of Hall and Heyde (1980).

3.3. The functional central limit theorem

As a sequel to the previous section we continue to consider a sequence $\{X_n\}$ of stochastic processes in C. Here we take up a typical example of $\{X_n\}$. Suppose that $u_1, u_2, \cdots,$ be random variables on $(\Omega, \mathcal{F}, P)$ and define the partial sum :

$$\begin{aligned} S_j &= S_{j-1} + u_j\,, \qquad (S_0 = 0)\,, \\ &= u_1 + \cdots + u_j\,, \quad (j = 1, \cdots, n)\,. \end{aligned} \tag{3.2}$$

We then construct, for $(j-1)/n \le t \le j/n$,

$$\begin{aligned} X_n(t) &= \frac{1}{\sqrt{n}}\, S_{j-1} + n\left(t - \frac{j-1}{n}\right)\frac{1}{\sqrt{n}}\, u_j \\ &= X_n\left(\frac{j-1}{n}\right) + n\left(t - \frac{j-1}{n}\right)\left(X_n\left(\frac{j}{n}\right) - X_n\left(\frac{j-1}{n}\right)\right) \\ &= \frac{1}{\sqrt{n}}\, S_j + n\left(t - \frac{j}{n}\right)\frac{1}{\sqrt{n}}\, u_j \\ &= X_n\left(\frac{j}{n}\right) + n\left(t - \frac{j}{n}\right)\left(X_n\left(\frac{j}{n}\right) - X_n\left(\frac{j-1}{n}\right)\right) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} u_j + (nt - [nt])\frac{1}{\sqrt{n}}\, u_{[nt]+1}\,, \end{aligned} \tag{3.3}$$

where $X_n(0) = 0$ and $X_n(1) = S_n/\sqrt{n}$. Note that, for each t, $X_n(t)$ is a random variable on $(\Omega, \mathcal{F}, P)$ and that $X_n(t)$ is a continuous function on [0,1] for each $\omega \in \Omega$. Figure 3.1 explains how to construct $X_n(t)$ on $\left[\frac{j-1}{n}, \frac{j}{n}\right]$ for fixed $\omega \in \Omega$. In any case $X_n = \{X_n(t)\}$ is a stochastic process in C and is called a *partial sum process.*

Figure 3.1

We now become more specific about random variables $u_1, u_2, \cdots$, in (3.3) to obtain the so-called *Functional Central Limit Theorem* (FCLT) or the *Invariance Principle* (IP) for $\{X_n\}$. Donsker (1951, 1952) provided the first general FCLT when $\{u_j\}$ follows i.i.d.$(0, \sigma^2)$, which we state below as Donsker's theorem. The proof starts by showing that the finite-dimensional distributions converge weakly and then goes on proving that the sequence in question is tight (see, for details, Billingsley (1968, p.68)).

Theorem 3.2 (Donsker's Theorem). *Suppose that the partial sum process $\{X_n(t)\}$ is defined in (3.3) with $\{u_j\}$ being i.i.d.$(0, \sigma^2)$, where $\sigma^2 > 0$. Then*

$$\mathcal{L}\left(\frac{X_n}{\sigma}\right) \longrightarrow \mathcal{L}(w),$$

where $w = \{w(t)\}$ is the one-dimensional standard Brownian motion on [0,1].

For later purposes we also consider the stochastic process of the following form :

$$\text{(3.4)} \qquad \begin{aligned} \bar{X}_n(t) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}(u_j - \bar{u}) + (nt - [nt])\frac{1}{\sqrt{n}}(u_{[nt]+1} - \bar{u}) \\ &= X_n(t) - tX_n(1), \end{aligned}$$

where $\bar{u} = \sum_{j=1}^{n} u_j/n$. The process $\{\bar{X}_n(t)\}$ may be referred to as the mean-corrected partial sum process, for which we have the following result.

Corollary 3.1. *Suppose that the mean-corrected partial sum process $\{\bar{X}_n(t)\}$ is defined in (3.4) with $\{u_j\}$ being i.i.d.$(0, \sigma^2)$, where $\sigma^2 > 0$. Then*

$$\mathcal{L}\left(\frac{\bar{X}_n}{\sigma}\right) \longrightarrow \mathcal{L}(\bar{w}),$$

where $\bar{w} = \{\bar{w}(t)\} = \{w(t) - tw(1)\}$ *is the one-dimensional Brownian bridge on [0,1].*

Similarly, if we consider the demeaned partial sum process :

$$\begin{aligned}(3.5) \qquad \tilde{X}_n(t) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} u_j - \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{j} u_i\right) + (nt - [nt])\frac{1}{\sqrt{n}}u_{[nt]+1} \\ &= X_n(t) - \frac{1}{n}\sum_{j=1}^{n} X_n\left(\frac{j}{n}\right),\end{aligned}$$

we have the following result.

Corollary 3.2. *Suppose that the demeaned partial sum process* $\{\tilde{X}_n(t)\}$ *is defined in (3.5) with* $\{u_j\}$ *being i.i.d.*$(0, \sigma^2)$*, where* $\sigma^2 > 0$*. Then*

$$\mathcal{L}\left(\frac{\tilde{X}_n}{\sigma}\right) \longrightarrow \mathcal{L}(\tilde{w}),$$

where $\tilde{w} = \{\tilde{w}(t)\} = \{w(t) - \int_0^1 w(t)dt\}$ *is the demeaned Brownian motion on [0,1].*

Note that the Brownian bridge and the demeaned Brownian motion are continuous functionals of the Brownian motion. Corollaries 3.1 and 3.2 may be proved from Donsker's theorem using the continuous mapping theorem discussed in the next section.

Donsker's FCLT was further developed into several directions where $\{u_j\}$ in (3.3) is a sequence of dependent random variables. Billingsley (1968, Chapter 4) established the FCLT under ϕ- *mixing conditions*, which was largely extended by McLeish (1975a, b, 1977) under the so-called *mixingale conditions*. There is now vast literature concerning mixing sequences and Yoshihara (1992, 1993) gives excellent reviews of the literature. We, however, do not take this approach because it seems difficult to accommodate the limit theory on mixing sequences to linear processes discussed subsequently. In fact, not all linear processes satisfy *strong mixing conditions* (see, for example, Withers (1981) and Athreya and Pantula (1986)). We will take an alternative approach that was advocated by Phillips and Solo (1992) and is especially designed for the case where $\{u_j\}$ follows a linear process.

3.4. Continuous mappings and related theorems

In this section we present some useful theorems to establish the FCLT for linear processes dealt with subsequently. The first theorem referred to as the *continuous mapping theorem* is quite important (see, for the proof, Billingsley (1968, p.29)).

Theorem 3.3. *Let $h(x)$ be a continuous function defined on C. If $\mathcal{L}(X_n) \to \mathcal{L}(X)$, then $\mathcal{L}(h(X_n)) \to \mathcal{L}(h(X))$.*

This theorem is well known if the X_n are random variables (see Rao (1973, p.124) and Problem 4.1).

To utilize fully the continuous mapping theorem 3.3 we need to define the Riemann integral in terms of convergence with probability 1. To see this consider

$$h(X) = \int_0^1 X^2(t)dt\,, \tag{3.6}$$

where $X(t)$ belongs to C. Although we naturally require $X(t) \in L_2$ for every t, it does not necessarily imply $X^2(t) \in L_2$. Thus the integral in (3.6) cannot be defined as the m.s. Riemann integral unless $X^2(t) \in L_2$. In the present case, however, $X^2(t) = X^2(t,\omega)$ is a continuous function of t with $\omega \in \Omega$ fixed so that the Riemann integral $\int_0^1 X^2(t,\omega)dt$ is well defined at each ω. On the other hand the collection of these integrals at all ω can be defined as limits of sequences of $\mathcal{F}$-measurable Riemann sums :

$$\sum_j X^2(t_j', \omega)(t_j - t_{j-1})\,, \qquad \left(t_j' \in [\,t_{j-1}, t_j\,)\right)\,,$$

since a uniform sequence of partitions of $[0,1]$ and uniform values t_j' can be used in obtaining $\int_0^1 X^2(t,\omega)dt$ at all ω. Thus $h(X)$ in (3.6) is $\mathcal{F}$-measurable and is independent of the sequences of partitions and the values t_j' involved in the limiting procedures. Thus the integral in (3.6) is well defined in the sense described above, which we call the Riemann sample integral. More details can be found in Soong (1973).

The continuity of $h(x)$ in (3.6) at a point $x \in C$ can be proved as follows. For x fixed consider

$$|h(y) - h(x)| \;=\; \left|\int_0^1 \left\{(x(t)-y(t))^2 - 2x(t)(x(t)-y(t))\right\} dt\right|$$

$$\leq \rho^2(x,y) + 2\rho(x,y)\int_0^1 |x(t)|dt\,,$$

which evidently tends to 0 as $y \to x$ ($\rho(x,y) \to 0$).

Three more examples of $h(x)$ follow (Problem 4.2).

$$h_1(x) = \sup_{0\leq t\leq 1} x(t)\,, \quad h_2(x) = \sup_{0\leq t\leq 1} |x(t)|\,, \quad h_3(x) = (h_1(x), h_2(x))\,. \tag{3.7}$$

Note that $h(x)$ may be vector-valued, as in $h_3(x)$. As a special case of Theorem 3.3, suppose that $\mathcal{L}(X_n) \to \mathcal{L}(c)$, where c is a constant on [0,1], which belongs to C. Then X_n converges in probability to c (Problem 4.3) in the sense that $\rho(X_n, c) \to 0$ in probability, that is,

$$P(\rho(X_n, c) \geq \varepsilon) = P(\sup_{0\leq t\leq 1} |X_n(t) - c| \geq \varepsilon) \longrightarrow 0$$

for each positive ε. Therefore we have the following theorem.

Theorem 3.4. *Let $h(x)$ be a continuous function defined on C. If X_n converges in probability to a constant c, then $h(X_n)$ converges in probability to $h(c)$.*

This theorem is also standard if the X_n are random variables (see Rao (1973, p.124) and Problem 4.4). The continuous mapping theorem 3.3 can be extended to the case where h is not necessarily continuous. The following theorem is proved in Billingsley (1968, p.31).

Theorem 3.5. *Let $h(x)$ be a measurable mapping of C into another metric space S' (with metric ρ' and σ-field $\mathcal{B}(S')$) and let D_h be the set of discontinuities of h. If $\mathcal{L}(X_n) \to \mathcal{L}(X)$ and $P(X \in D_h) = 0$, then $\mathcal{L}(h(X_n)) \to \mathcal{L}(h(X))$.*

As an example consider $h(w) = 1/\int_0^1 w^2(t)dt$, where $\{w(t)\}$ is the standard Brownian motion. Then $h(w)$ is measurable and $P(w = 0) = 0$ so that $P(w \in D_h) = 0$; hence the above theorem applies.

The next theorem relates convergence in probability on C with weak convergence. The proof is given in Billingsley (1968, p.25).

Theorem 3.6. *If $\mathcal{L}(X_n) \to \mathcal{L}(X)$ and $\rho(X_n, Y_n) \to 0$ in probability, then $\mathcal{L}(Y_n) \to \mathcal{L}(X)$.*

This theorem is also well known if X_n and Y_n are random variables (see Rao (1973, p.122) and Problem 4.5).

We now explore some applications of theorems presented above. Let us consider a model :

$$(3.8) \qquad y_j = \rho y_{j-1} + \varepsilon_j\,, \quad y_0 = 0\,, \quad (j = 1, \cdots, T),$$

where the true value of ρ is 1 and $\{\varepsilon_j\} \sim$ i.i.d.$(0, \sigma^2)$. The statistics dealt with here are

$$\begin{aligned}
S_{1T} &= \frac{1}{T^2}\sum_{j=1}^{T} y_j^2\,, \\
S_{2T} &= \frac{1}{T^2}\sum_{j=1}^{T} (y_j - \bar{y})^2\,, \\
S_{3T} &= T(\hat{\rho} - 1)\,, \\
S_{4T} &= \frac{\hat{\rho} - 1}{\hat{\sigma} \Big/ \sqrt{\sum_{j=2}^{T} y_{j-1}^2}}\,,
\end{aligned}$$

where

$$\hat{\rho} = \sum_{j=2}^{T} y_{j-1} y_j \Big/ \sum_{j=2}^{T} y_{j-1}^2\,, \qquad \hat{\sigma}^2 = \frac{1}{T-1}\sum_{j=2}^{T} (y_j - \hat{\rho} y_{j-1})^2\,.$$

The above statistics S_{1T} through S_{4T} are functions of $\{y_j\}$. Noting that $y_j = \varepsilon_1 + \cdots + \varepsilon_j$ we define, for $(j-1)/T \le t \le j/T$,

$$(3.9) \qquad X_T(t) = \frac{1}{\sqrt{T}}\sum_{i=1}^{j} \varepsilon_i + T\left(t - \frac{j}{T}\right)\frac{1}{\sqrt{T}}\varepsilon_j$$

so that $X_T(j/T) = y_j/\sqrt{T}$ and $\mathcal{L}(X_T/\sigma) \to \mathcal{L}(w)$.

As for S_{1T} we have

$$\begin{aligned}
S_{1T} &= \frac{1}{T}\sum_{j=1}^{T} X_T^2\left(\frac{j}{T}\right) \\
&= h_1(X_T) + R_{1T}\,,
\end{aligned}$$

where

$$
\begin{aligned}
h_1(x) &= \int_0^1 x^2(t)dt\,, \quad x \in C\,, \\
R_{1T} &= \frac{1}{T}\sum_{j=1}^{T} X_T^2\left(\frac{j}{T}\right) - \int_0^1 X_T^2(t)dt \qquad (3.10)\\
&= \sum_{j=1}^{T}\int_{\frac{j-1}{T}}^{\frac{j}{T}} \left[X_T^2\left(\frac{j}{T}\right) - X_T^2(t)\right]dt\,.
\end{aligned}
$$

Since $\mathcal{L}(h_1(X_T/\sigma)) \to \mathcal{L}(h_1(w))$ by the continuous mapping theorem, $\mathcal{L}(S_{1T}/\sigma^2) \to \mathcal{L}(h_1(w))$ follows from Theorem 3.6 if R_{1T} converges in probability to 0. In fact the integrand in (3.10) has the following bound:

$$
\left|X_T^2\left(\frac{j}{T}\right) - X_T^2(t)\right| \le 2 \sup_{0\le t\le 1}|X_T(t)| \max_{1\le j\le T}\frac{|\varepsilon_j|}{\sqrt{T}}\,,
$$

where we have $\mathcal{L}(\sup_{0\le t\le 1}|X_T(t)/\sigma|) \to \mathcal{L}(\sup_{0\le t\le 1}|w(t)|)$. Here $P(\sup_{0\le t\le 1}|w(t)| \ge b)$ is known as a boundary-crossing probability (Shorack and Wellner (1986, p.34)). We also have (Problem 4.6)

$$
\max_{1\le j\le T}\frac{|\varepsilon_j|}{\sqrt{T}} \longrightarrow 0 \qquad \text{in probability.} \qquad (3.11)
$$

Thus it must hold (Problem 4.7) that

$$
\sup_{0\le t\le 1}|X_T(t)| \max_{1\le j\le T}\frac{|\varepsilon_j|}{\sqrt{T}} \longrightarrow 0 \qquad \text{in probability.} \qquad (3.12)
$$

Therefore we obtain

$$
\mathcal{L}\left(\frac{S_{1T}}{\sigma^2}\right) = \mathcal{L}\left(\frac{1}{T^2\sigma^2}\sum_{j=1}^{T} y_j^2\right) \to \mathcal{L}\left(\int_0^1 w^2(t)dt\right)\,.
$$

Note that this last integral expression is well defined both in the m.s. sense and in the sense of convergence with probability 1, while $h_1(X_T)$ is not necessarily m.s. integrable.

Similarly we can show (Problem 4.8) that

$$
\begin{aligned}
\mathcal{L}\left(\frac{S_{2T}}{\sigma^2}\right) &= \mathcal{L}\left(\frac{1}{T^2\sigma^2}\sum_{j=1}^{T}(y_j - \bar{y})^2\right) \qquad (3.13)\\
&\longrightarrow \mathcal{L}\left(\int_0^1 w^2(t)dt - \left(\int_0^1 w(t)dt\right)^2\right)\\
&= \mathcal{L}\left(\int_0^1 (w(t) - tw(t))^2 dt\right)\,.
\end{aligned}
$$

Establishing the last equality in (3.13), however, is not straightforward so far as $\{\varepsilon_j\} \sim$ i.i.d.$(0, \sigma^2)$. Under this assumption weak convergence can be established as in the second line in (3.13). Thus we may consider a special case where $\{\varepsilon_j\} \sim$ NID$(0, \sigma^2)$, from which the last expression in (3.13) results. This is a reason why Donsker's FCLT is also called the invariance principle. Once the weak convergence is established, we can find the limiting distribution in an easy special case.

We next consider

$$
\begin{aligned}
S_{3T} &= T(\hat{\rho} - 1) \\
&= \frac{1}{T}\sum_{j=2}^{T} y_{j-1}(y_j - y_{j-1}) \Bigg/ \left[\frac{1}{T^2}\sum_{j=2}^{T} y_{j-1}^2\right] \\
&= U_T / V_T \,,
\end{aligned}
$$

where

$$
\begin{aligned}
U_T &= \frac{1}{T}\sum_{j=2}^{T} y_{j-1}(y_j - y_{j-1}) \\
&= \frac{1}{2}X_T^2(1) - \frac{1}{2T}\sum_{j=1}^{T}\varepsilon_j^2 \,,
\end{aligned}
\tag{3.14}
$$

$$
\begin{aligned}
V_T &= \frac{1}{T^2}\sum_{j=2}^{T} y_{j-1}^2 \\
&= \frac{1}{T}\sum_{j=1}^{T} X_T^2\left(\frac{j}{T}\right) - \frac{1}{T^2}y_T^2 \,.
\end{aligned}
\tag{3.15}
$$

Let us define a continuous function $h_3(x) = (h_{31}(x),\, h_{32}(x))$ for $x \in C$, where

$$
h_{31}(x) = \frac{1}{2}x^2(1)\,, \qquad h_{32}(x) = \int_0^1 x^2(t)dt\,.
$$

Then we have

$$
\begin{aligned}
U_T &= h_{31}(X_T) - \frac{1}{2T}\sum_{j=1}^{T}\varepsilon_j^2 \,, \\
V_T &= h_{32}(X_T) + R_{1T} - \frac{1}{T^2}y_T^2 \,,
\end{aligned}
$$

where R_{1T} is defined in (3.10).

It is now easy to deduce that

$$\mathcal{L}\left(\frac{U_T}{\sigma^2},\, \frac{V_T}{\sigma^2}\right) \longrightarrow \mathcal{L}\left(h_{31}(w) - \frac{1}{2},\, h_{32}(w)\right)$$

and Theorem 3.5 yields

$$\begin{aligned}
\mathcal{L}(S_{3T}) &= \mathcal{L}(T(\hat{\rho}-1)) \\
&\longrightarrow \mathcal{L}\left(\frac{h_{31}(w) - \frac{1}{2}}{h_{32}(w)}\right) \\
&= \mathcal{L}\left(\frac{\frac{1}{2}(w^2(1)-1)}{\int_0^1 w^2(t)dt}\right) \\
&= \mathcal{L}\left(\frac{\int_0^1 w(t)dw(t)}{\int_0^1 w^2(t)dt}\right) .
\end{aligned}$$

Finally we deal with the t-ratio like statistic defined by

$$\begin{aligned}
S_{4T} &= \frac{\hat{\rho}-1}{\hat{\sigma}\Big/\sqrt{\sum_{j=2}^{T} y_{j-1}^2}} = \frac{U_T/V_T}{\hat{\sigma}/\sqrt{V_T}} \\
&= \frac{U_T}{\hat{\sigma}\sqrt{V_T}} ,
\end{aligned}$$

where U_T and V_T are defined in (3.14) and (3.15), respectively. Since it can be shown (Problem 4.9) that

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{j=2}^{T}(y_j - \hat{\rho}y_{j-1})^2 \longrightarrow \sigma^2 \tag{3.16}$$

in probability, Theorem 3.5 again yields

$$\mathcal{L}(S_{4T}) \longrightarrow \mathcal{L}\left(\frac{\int_0^1 w(t)dw(t)}{\sqrt{\int_0^1 w^2(t)dt}}\right) .$$

As is seen above, the FCLT combined with the continuous mapping theorem is powerful and plays an important role in deriving weak convergence results for various statistics. Since the present approach always starts with constructing a partial sum process in C, while our concern is a statistic, the approach may be referred to as the stochastic process approach. The terminology will also be used in Chapter 4 for another purpose. If the statistic under consideration is a quadratic form or the ratio of quadratic forms, we need not make such a detour as is involved in the present approach. An alternative approach will be presented in Section 6 of Chapter 5.

Problems

4.1 Prove Theorem 3.3 when the X_n are random variables.

4.2 Show that the functions $h_1(x)$, $h_2(x)$ and $h_3(x)$ in (3.7) are continuous in C.

4.3 Prove that $\mathcal{L}(X_n) \to \mathcal{L}(c)$ in C implies $\rho(X_n, c) \to 0$ in probability, where c is a constant.

4.4 Prove Theorem 3.4 when the X_n are random variables.

4.5 Prove Theorem 3.6 when X_n and Y_n are random variables.

4.6 Show that (3.11) holds.

4.7 Prove that, if $\mathcal{L}(X_n) \to \mathcal{L}(X)$ and $Y_n \to 0$ in probability, then $X_n Y_n \to 0$ in probability, where X, X_n and Y_n are random variables.

4.8 Derive the weak convergence results in (3.13).

4.9 Establish (3.16).

3.5. FCLT for linear processes: case 1

In this section we consider a sequence $\{Y_n\}$ of stochastic processes in C defined by

$$Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{j} u_i + n\left(t - \frac{j}{n}\right) \frac{1}{\sqrt{n}}\, u_j\,, \quad \left(\frac{j-1}{n} \le t \le \frac{j}{n}\right), \tag{3.17}$$

where $\{u_j\}$ is assumed to be generated by

$$u_j = \sum_{l=0}^{\infty} \alpha_l\, \varepsilon_{j-l}\,, \qquad \alpha_0 = 1\,. \tag{3.18}$$

Here $\{\varepsilon_j\}$ is a sequence of random variables defined on $(\Omega, \mathcal{F}, P)$, while $\{\alpha_l\}$ is a sequence of constants for which we assume

$$\sum_{l=0}^{\infty} l\,|\alpha_l| < \infty\,. \tag{3.19}$$

The condition (3.19) may be replaced, for example, by $\sum_{l=0}^{\infty} l^2\alpha_l^2 < \infty$, but we do assume the stronger condition (3.19) only for simplicity (Problem 5.1).

To establish the FCLT for $\{Y_n\}$ in (3.17) we need additional assumptions on $\{\varepsilon_j\}$. In this section we assume that $\{\varepsilon_j\} \sim \text{i.i.d.}(0, \sigma^2)$ with $\sigma^2 > 0$ so that $\{u_j\}$ in (3.18) belongs to L_2 discussed in Chapter 2 and becomes ergodic and strictly stationary as well as second-order stationary (Hannan (1970, p.204)). Following Phillips and Solo (1992) let us decompose $\{u_j\}$ into

$$(3.20) \qquad u_j = \alpha\varepsilon_j + \tilde{\varepsilon}_{j-1} - \tilde{\varepsilon}_j \, ,$$

where it is easy to check (Problem 5.2) that

$$(3.21) \qquad \alpha = \sum_{l=0}^{\infty} \alpha_l \, ,$$

$$(3.22) \qquad \tilde{\varepsilon}_j = \sum_{l=0}^{\infty} \tilde{\alpha}_l \, \varepsilon_{j-l} \, , \qquad \tilde{\alpha}_l = \sum_{k=l+1}^{\infty} \alpha_k \, .$$

The sequence $\{\tilde{\varepsilon}_j\}$ also becomes stationary (Problem 5.3) with $|\alpha| < \infty$ and $\sum_{l=0}^{\infty} |\tilde{\alpha}_l| < \infty$. The decomposition (3.20) is known in the econometrics literature as the Beveridge-Nelson (1981) or BN decomposition. A similar decomposition was already used by Fuller (1976) when the MA representation is of finite order (Problem 5.4).

We now have

$$(3.23) \qquad Y_n(t) = \alpha X_n(t) + R_n(t) \, ,$$

where, for $(j-1)/n \le t \le j/n$,

$$(3.24) \qquad X_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{j} \varepsilon_i + n\left(t - \frac{j}{n}\right) \frac{1}{\sqrt{n}} \varepsilon_j \, ,$$

$$(3.25) \qquad R_n(t) = \frac{1}{\sqrt{n}} (\tilde{\varepsilon}_0 - \tilde{\varepsilon}_j) + n\left(t - \frac{j}{n}\right) \frac{1}{\sqrt{n}} (\tilde{\varepsilon}_{j-1} - \tilde{\varepsilon}_j) \, .$$

By Donsker's theorem 3.2, we have $\mathcal{L}(X_n/\sigma) \to \mathcal{L}(w)$, where $w = \{w(t)\}$ is the standard Brownian motion, so that $\mathcal{L}(\alpha X_n/\sigma) \to \mathcal{L}(\alpha w)$ by the continuous mapping theorem. We now show that $\rho(Y_n, \alpha X_n) \to 0$ in probability so that, by Theorem 3.6, we have $\mathcal{L}(Y_n/\sigma) \to \mathcal{L}(\alpha w)$. Consider

$$\rho(Y_n, \alpha X_n) = \sup_{0\le t\le 1} |R_n(t)| \le \frac{4}{\sqrt{n}} \max_{0\le j\le n} |\tilde{\varepsilon}_j| \,, \tag{3.26}$$

which converges in probability to 0 if

$$\frac{1}{\sqrt{n}} \max_{0\le j\le n} |\tilde{\varepsilon}_j| \longrightarrow 0 \qquad \text{in probability.} \tag{3.27}$$

This last condition is equivalent (Problem 5.5) to

$$J_n = \frac{1}{n}\sum_{j=0}^{n} \tilde{\varepsilon}_j^2\, I(\tilde{\varepsilon}_j^2 > n\delta) \longrightarrow 0 \qquad \text{in probability} \tag{3.28}$$

for any $\delta > 0$, where $I(A)$ is the indicator function of A. Since $E(J_n) \to 0$ because of strict and second-order stationarity of $\{\tilde{\varepsilon}_j\}$, (3.28) follows from Markov's inequality.

The above arguments are summarized in the following theorem.

Theorem 3.7. *Let $\{Y_n\}$ be defined by (3.17) with $\{u_j\}$ being generated by the linear process (3.18) under the summability condition (3.19). If $\{\varepsilon_j\}$ in (3.18) is an i.i.d.$(0,\sigma^2)$ sequence, then $\mathcal{L}(Y_n/\sigma) \to \mathcal{L}(\alpha w)$.*

As an application of this theorem consider an integrated process :

$$y_j = y_{j-1} + u_j \,, \qquad y_0 = 0 \,, \qquad u_j = \sum_{l=0}^{\infty} \alpha_l\, \varepsilon_{j-l} \,, \qquad \alpha_0 = 1 \,, \tag{3.29}$$

where $\{\varepsilon_j\} \sim$ i.i.d.$(0,1)$ and $\{\alpha_l\}$ satisfies (3.19). Then we obtain (Problem 5.6)

$$\mathcal{L}\left(\frac{1}{T^2}\sum_{j=1}^{T} y_j^2\right) \longrightarrow \mathcal{L}\left(\alpha^2 \int_0^1 w^2(t)dt\right) , \tag{3.30}$$

$$\mathcal{L}\left(\frac{1}{T^2}\sum_{j=1}^{T} (y_j - \bar{y})^2\right) \longrightarrow \mathcal{L}\left(\alpha^2 \int_0^1 \left(w(t) - \int_0^1 w(s)ds\right)^2 dt\right) , \tag{3.31}$$

where $\alpha = \sum_{l=0}^{\infty} \alpha_l$.

Problems

5.1 Show that $\sum_{l=0}^{\infty} l\,|\alpha_l| < \infty$ implies $\sum_{l=0}^{\infty} l^2 \alpha_l^2 < \infty$.

5.2 Derive the BN decomposition (3.20) from (3.18).

5.3 Show that $\{\tilde{\varepsilon}_j\}$ in (3.22) is second-order stationary.

5.4 Derive the BN decomposition as in (3.20) when the order of the MA representation in (3.18) is finite.

5.5 Prove that

$$P(\max_{1\le j\le n} |Z_j| > \delta) = P\left(\sum_{j=1}^{n} Z_j^2\, I(|Z_j| > \delta) > \delta^2\right)$$

so that (3.27) and (3.28) are equivalent.

5.6 Derive the weak convergence results (3.30) and (3.31).

3.6. FCLT for martingale differences

This section serves as a basis of discussions concerning FCLT's for linear processes presented in the next section. The FCLT that we have seen in Section 3 assumes the basic sequence that forms the partial sum as in (3.2) to be i.i.d., and this FCLT has been extended in Section 5 to the dependent case where the basic sequence is stationary and is represented by an infinite, weighted sum of i.i.d. random variables. In these FCLT's the sequence of stochastic processes has a constant variance, while covariances depend only on time differences. This is referred to as the *homogeneous case*. The present section deals with a *nonhomogeneous case*, where the basic sequence has nonconstant variances, while covariances are assumed to be zero. This last assumption is relaxed in the next section so that the basic sequence has nonconstant covariances.

For the above purpose we assume $\{\varepsilon_j\}$ to be a sequence of *martingale differences*, that is, $E(|\varepsilon_j|) < \infty$ and $E(\varepsilon_j|\mathcal{F}_{j-1}) = 0$ (a.s.), where $\{\mathcal{F}_j\}$ is an increasing sequence of *sub σ-fields* of $\mathcal{F}$. Note that $\{\varepsilon_j\}$ is defined on $(\Omega, \mathcal{F}, P)$ and that each ε_j is measurable with respect to $\mathcal{F}_j$. Then $\{\varepsilon_j\}$ is said to be *adapted* to $\{\mathcal{F}_j\}$. In subsequent discussions $\{\varepsilon_j\}$ is assumed to belong to L_2, that is, $E(\varepsilon_j^2) < \infty$, which is said to be *square integrable*. Note that square integrability of $\{\varepsilon_j\}$ does not necessarily imply $\sup_j E(\varepsilon_j^2) < \infty$, much less $E(\sup_j \varepsilon_j^2) < \infty$.

We now describe the FCLT due to Brown (1971) (see, also, Hall and Heyde (1980, p.99)). Let us define

$$\xi_n(t) = \frac{1}{s_n}\sum_{i=1}^{j-1}\varepsilon_i + \frac{ts_n^2 - s_{j-1}^2}{s_j^2 - s_{j-1}^2}\frac{1}{s_n}\varepsilon_j\,, \quad \left(\frac{s_{j-1}^2}{s_n^2} \le t \le \frac{s_j^2}{s_n^2}\right), \tag{3.32}$$

where

$$s_n^2 = \sum_{j=1}^{n} E(\varepsilon_j^2) = E\left[\left(\sum_{j=1}^{n}\varepsilon_j\right)^2\right]. \tag{3.33}$$

It is noticed that the construction of the partial sum process $\{\xi_n(t)\}$ is different from Donsker's. This is because of the nonhomogeneous nature of variances of $\{\varepsilon_j\}$. A geometrical interpretation of paths of $\xi_n(t)$ is that $\xi_n(t)$ in the interval $[s_{j-1}^2/s_n^2,\ s_j^2/s_n^2]$ is on the line joining $(s_{j-1}^2/s_n^2,\ \sum_{i=1}^{j-1}\varepsilon_i/s_n)$ and $(s_j^2/s_n^2,\ \sum_{i=1}^{j}\varepsilon_i/s_n)$ (Problem 6.1).

Theorem 3.8. *Let $\{\varepsilon_j\}$ be a sequence of square integrable martingale differences satisfying*

$$\frac{1}{s_n^2}\sum_{j=1}^{n}\varepsilon_j^2 \longrightarrow 1 \qquad \text{in probability,} \tag{3.34}$$

$$\frac{1}{s_n^2}\sum_{j=1}^{n}E\left[\varepsilon_j^2 I(|\varepsilon_j| > \delta s_n)\right] \longrightarrow 0 \qquad \text{for every } \delta > 0\,. \tag{3.35}$$

Then $\mathcal{L}(\xi_n) \to \mathcal{L}(w)$.

It is easy to check that the present theorem does imply Donsker's FCLT when $\{\varepsilon_j\}$ is i.i.d.$(0,\sigma^2)$ (Problem 6.2). It is not trivial if what class of martingale differences satisfies the conditions (3.34) and (3.35). A set of sufficient conditions for these to hold will be given later. The condition (3.35) is referred to as the *Lindeberg condition*, which implies (Problem 6.3) that

$$\max_{1\le j\le n}\frac{E(\varepsilon_j^2)}{s_n^2} \longrightarrow 0\,, \tag{3.36}$$

and thus, by Chebyshev's inequality,

$$\max_{1\le j\le n} P\left(\frac{|\varepsilon_j|}{s_n} > \delta\right) \longrightarrow 0 \qquad \text{for every} \quad \delta > 0\,. \tag{3.37}$$

Moreover, since it holds (Problem 6.4) that

$$P\left(\max_{1\le j\le n} \frac{|\varepsilon_j|}{s_n} > \delta\right) = P\left(\frac{1}{s_n^2}\sum_{j=1}^n \varepsilon_j^2 I(|\varepsilon_j| > \delta s_n) > \delta^2\right)\,, \tag{3.38}$$

the Lindeberg condition also implies

$$\max_{1\le j\le n} \frac{|\varepsilon_j|}{s_n} \longrightarrow 0 \qquad \text{in probability.} \tag{3.39}$$

Phillips and Solo (1992) prove that, if (3.34) holds, (3.39) implies the Lindeberg condition (3.35). Therefore the conditions (3.35) and (3.39) are equivalent under (3.34).

We now give a set of sufficient conditions for Theorem 3.8 to hold. We first assume that there exists a random variable η with $E(\eta^2) < \infty$ such that

$$P(|\varepsilon_j| > x) \le cP(|\eta| > x) \tag{3.40}$$

for each $x \ge 0$, $j \ge 1$ and for some positive constant c. In general, the sequence $\{\varepsilon_j\}$ satisfying (3.40) with $E(|\eta|) < \infty$ is said to be *strongly uniformly integrable (s.u.i.)*. Since we assume $E(\eta^2) < \infty$, $\{\varepsilon_j^2\}$ also becomes s.u.i., which implies that

$$\begin{aligned} E(\varepsilon_j^2) &= \int_0^\infty P(\varepsilon_j^2 > x)dx \\ &\le c\int_0^\infty P(\eta^2 > x)dx \\ &= cE(\eta^2) < \infty \end{aligned}$$

so that $\sup_j E(\varepsilon_j^2) < \infty$. It also implies *uniform integrability* of $\{\varepsilon_j^2\}$, that is

$$\lim_{\delta\to\infty} \sup_j E\left[\varepsilon_j^2 I(|\varepsilon_j| > \delta)\right] = 0\,, \tag{3.41}$$

since it holds (Problem 6.5) that

$$E\left[\varepsilon_j^2 I(|\varepsilon_j| > \delta)\right] = \delta^2 P(|\varepsilon_j| > \delta) + \int_{\delta^2}^\infty P\left(|\varepsilon_j| > \sqrt{x}\right) dx\,. \tag{3.42}$$

It can be shown (Problem 6.6) that uniform integrability of $\{\varepsilon_j^2\}$ implies $\sup_j E(\varepsilon_j^2) < \infty$.

We next assume that

$$(3.43) \qquad \frac{1}{n}\sum_{j=1}^{n} E\left(\varepsilon_j^2|\mathcal{F}_{j-1}\right) \longrightarrow \sigma_\varepsilon^2 \qquad \text{in probability,}$$

where σ_ε^2 is a positive constant. Since Hall and Heyde (1980, p.36) proved that, if $\{\varepsilon_j^2\}$ is s.u.i.,

$$\frac{1}{n}\sum_{j=1}^{n} \left[\varepsilon_j^2 - E\left(\varepsilon_j^2|\mathcal{F}_{j-1}\right)\right] \longrightarrow 0 \qquad \text{in probability,}$$

we necessarily have that

$$\frac{1}{n}\sum_{j=1}^{n} \varepsilon_j^2 \longrightarrow \sigma_\varepsilon^2 \qquad \text{in probability,}$$

Moreover it is known (Chow and Teicher (1988, p.102)) that $\{\sum_{j=1}^{n} \varepsilon_j^2/n\}$ is s.u.i. if $\{\varepsilon_j^2\}$ is. This fact together with convergence in probability implies (Chow and Teicher (1988, p.100)) that

$$\frac{1}{n}\sum_{j=1}^{n} E\left(\varepsilon_j^2\right) = \frac{s_n^2}{n} \longrightarrow \sigma_\varepsilon^2 \,.$$

Thus the first condition (3.34) in Theorem 3.8 is clearly satisfied if $\{\varepsilon_j^2\}$ is s.u.i. and (3.43) holds.

As for the Lindeberg condition (3.35) we can deduce from (3.42) that, if $\{\varepsilon_j^2\}$ is s.u.i.,

$$E\left[\varepsilon_j^2 I(|\varepsilon_j| > \sqrt{n}\delta\sigma_\varepsilon)\right] \le cE\left[\eta^2 I(|\eta| > \sqrt{n}\delta\sigma_\varepsilon)\right] \,.$$

Thus the Lindeberg condition (3.35) is ensured because $E(\eta^2) < \infty$.

We conclude the above arguments by the following corollary.

Corollary 3.3. *Let $\{\varepsilon_j\}$ be a sequence of square integrable martingale differences that satisfies (3.40) with $E(\eta^2) < \infty$, and (3.43) with $\sigma_\varepsilon^2 > 0$. Then the two conditions in Theorem 3.8 are satisfied so that $\mathcal{L}(\xi_n) \to \mathcal{L}(w)$.*

The strong uniform integrability condition plays an important role in the above corollary. That condition was also used by Hannan and Heyde (1972) in a different context. If we can impose quite a restrictive assumption:

$$(3.44) \qquad E(\sup_j \varepsilon_j^2) < \infty\,,$$

we necessarily have $P(\varepsilon_j^2 > x) \le P(\sup_j \varepsilon_j^2 > x)$ so that (3.44) ensures strong uniform integrability of $\{\varepsilon_j^2\}$. Thus (3.44) implies the Lindeberg condition (3.35). Then, if (3.43) holds, the condition (3.34) in Theorem 3.8 is also satisfied.

The following corollary summarizes the above arguments.

Corollary 3.4. *Let $\{\varepsilon_j\}$ be a sequence of square integrable martingale differences such that*

i) $E(\sup_j \varepsilon_j^2) < \infty$;

ii) $\dfrac{1}{n}\sum_{j=1}^{n} E(\varepsilon_j^2|\mathcal{F}_{j-1})$ *converges in probability to* $\sigma_\varepsilon^2 > 0$.

Then the two conditions in Theorem 3.8 are satisfied so that $\mathcal{L}(\xi_n) \to \mathcal{L}(w)$.

As a final remark to this section we mention that the above results also apply to a triangular array $\{\varepsilon_{jn},\ 1 \le j \le n,\ n \ge 1\}$ of square integrable martingale differences, where $\{\varepsilon_{jn}\}$ is adapted to $\{F_{jn}\}$ which is a triangular array of sub σ-fields of $\mathcal{F}$ such that $\mathcal{F}_{j-1,n} \subset \mathcal{F}_{jn}$ for all n. This means that we may put $\varepsilon_{jn} = \varepsilon_j/s_n$ in Theorem 3.8. A simple example of $\{\varepsilon_{jn}\}$ is $\{j\eta_j/n\}$ with $\{\eta_j\}$ being i.i.d.$(0, \sigma^2)$. Here $\{j\eta_j/n\}$ is not identically distributed, although independent. It can be checked (Problem 6.7) that Theorem 3.8 still holds with ε_j replaced by $\varepsilon_{jn} = j\eta_j/n$.

Problems

6.1 Explain the geometrical interpretation of paths of $\xi_n(t)$ described below (3.33).

6.2 Check that the two conditions in Theorem 3.8 are satisfied if $\{\varepsilon_j\}$ is i.i.d.$(0, \sigma^2)$.

6.3 Derive (3.36) from the Lindeberg condition (3.35).

6.4 Prove the relation in (3.38).

6.5 Show that strong uniform integrability of $\{\varepsilon_j^2\}$ implies uniform integrability of $\{\varepsilon_j^2\}$, proving the formula (3.42).

6.6 Prove that uniform integrability of $\{\varepsilon_j^2\}$ implies $\sup_j E(\varepsilon_j^2) < \infty$.

6.7 Show that the two conditions in Theorem 3.8 are satisfied if ε_j is replaced by $\varepsilon_{jn} = j\eta_j/n$ with $\{\eta_j\}$ being i.i.d.$(0, \sigma^2)$.

3.7. FCLT for linear processes : case 2

This section deals with the linear process generated by a sequence $\{\varepsilon_j\}$ of square integrable martingale differences discussed in the previous section. More specifically we consider

$$Y_n(t) = \frac{1}{s_n}\sum_{i=1}^{j-1} u_i + \frac{ts_n^2 - s_{j-1}^2}{s_j^2 - s_{j-1}^2}\frac{1}{s_n}u_j\,, \qquad \left(\frac{s_{j-1}^2}{s_n^2} \le t \le \frac{s_j^2}{s_n^2}\right)\,, \tag{3.45}$$

where

$$s_n^2 = \sum_{j=1}^{n} E(\varepsilon_j^2) = E\left[\left(\sum_{j=1}^{n}\varepsilon_j\right)^2\right]\,, \tag{3.46}$$

$$u_j = \sum_{l=0}^{\infty}\alpha_l\,\varepsilon_{j-l}\,, \qquad \alpha_0 = 1\,, \tag{3.47}$$

$$\sum_{l=0}^{\infty} l\,|\alpha_l| < \infty\,. \tag{3.48}$$

In Section 5 we saw that the FCLT for the linear process generated by an i.i.d.$(0, \sigma^2)$ sequence did not require any additional assumptions except (3.47) and (3.48). In the present case it seems necessary to impose a slightly stronger moment condition on $\{\varepsilon_j\}$ than required in the previous section to establish the FCLT for $\{Y_n\}$ in (3.45), which we now discuss.

Using the BN decomposition as in Section 5 we obtain (Problem 7.1)

$$Y_n(t) = \alpha\xi_n(t) + R_n(t)\,, \tag{3.49}$$

where, for $s_{j-1}^2/s_n^2 \le t \le s_j^2/s_n^2$,

$$\xi_n(t) = \frac{1}{s_n}\sum_{i=1}^{j-1}\varepsilon_i + \frac{ts_n^2 - s_{j-1}^2}{s_j^2 - s_{j-1}^2}\frac{1}{s_n}\varepsilon_j\,, \tag{3.50}$$

$$R_n(t) = \frac{1}{s_n}(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{j-1}) + \frac{ts_n^2 - s_{j-1}^2}{s_j^2 - s_{j-1}^2}\frac{1}{s_n}(\tilde{\varepsilon}_{j-1} - \tilde{\varepsilon}_j)\,, \tag{3.51}$$

$$\alpha = \sum_{l=0}^{\infty}\alpha_l\,, \tag{3.52}$$

$$\tilde{\varepsilon}_j = \sum_{l=0}^{\infty}\tilde{\alpha}_l\,\varepsilon_{j-l}\,, \qquad \tilde{\alpha}_l = \sum_{k=l+1}^{\infty}\alpha_k\,. \tag{3.53}$$

Suppose, for a moment, that $\{\varepsilon_j\}$ is a sequence of square integrable martingale differences satisfying that $\{\varepsilon_j^2\}$ is strongly uniformly integrable, that is, there exists η with $E(\eta^2) < \infty$ such that $P(|\varepsilon_j| > x) \le cP(|\eta| > x)$ for each $x \ge 0$, each integer j and some $c > 0$. We also assume that $\sum_{j=1}^{n} E(\varepsilon_j^2|\mathcal{F}_{j-1})/n$ converges in probability to $\sigma_\varepsilon^2 > 0$. Then it follows from Corollary 3.3 and the continuous mapping theorem 3.3 that $\mathcal{L}(\alpha\xi_n) \to \mathcal{L}(\alpha w)$.

We next deal with the remainder term $R_n(t)$ defined in (3.51), for which we have (Problem 7.2)

$$\sup_{0\le t\le 1}|R_n(t)| = \rho(Y_n, \alpha\xi_n) \le \frac{4}{s_n}\max_{0\le j\le n}|\tilde{\varepsilon}_j|\,. \tag{3.54}$$

Note that $\{\tilde{\varepsilon}_j\}$ satisfies (Problem 7.3) that

$$E(\tilde{\varepsilon}_j^2) \le cE(\eta^2)\sum_{l=0}^{\infty}\tilde{\alpha}_l^2 < \infty\,. \tag{3.55}$$

As was explained in Section 5, the last quantity in (3.54) converges in probability to 0 if

$$\frac{1}{s_n}\max_{0\le j\le n}|\tilde{\varepsilon}_j| \longrightarrow 0 \qquad \text{in probability} \tag{3.56}$$

or, equivalently,

$$J_n = \frac{1}{s_n^2}\sum_{j=0}^{n}\tilde{\varepsilon}_j^2 I(\tilde{\varepsilon}_j^2 > s_n^2\delta) \longrightarrow 0 \qquad \text{in probability} \tag{3.57}$$

for any $\delta > 0$. The condition (3.57) holds if $E(J_n) \to 0$, which was automatically satisfied in Section 5 since $\{\tilde{\varepsilon}_j^2\}$ was uniformly integrable because of strict and second-order stationarity of $\{\tilde{\varepsilon}_j\}$. In the present case, however, $\{\tilde{\varepsilon}_j\}$ is not stationary. Nonetheless we assume $\{\tilde{\varepsilon}_j^2\}$ to be uniformly integrable, as in Phillips and Solo (1992). A sufficient condition for this is $\sup_j E(|\tilde{\varepsilon}_j|^{2+\gamma}) < \infty$ for some $\gamma > 0$ (Billingsley (1968, p.32) and Problem 7.4). This, in turn, holds if $\sup_j E(|\varepsilon_j|^{2+\gamma}) < \infty$, which is implied by $E(|\eta|^{2+\gamma}) < \infty$ because of Hölder's inequality (Problem 7.5). We now have $\rho(Y_n, \alpha\xi_n) \to 0$ in probability so that $\mathcal{L}(Y_n) \to \mathcal{L}(\alpha w)$ by Theorem 3.6.

The above arguments are summarized in the following theorem.

Theorem 3.9. *Let $\{Y_n\}$ be defined by (3.45) with $\{u_j\}$ being generated by the linear process (3.47) under the summability condition (3.48). If $\{\varepsilon_j\}$ in (3.47) is a sequence of square integrable martingale differences that satisfies (3.40) with $E(|\eta|^{2+\gamma}) < \infty$ for some $\gamma > 0$, and (3.43) with $\sigma_\varepsilon^2 > 0$, then $\mathcal{L}(Y_n) \to \mathcal{L}(\alpha w)$.*

Note that, in comparison with the condition $E(\eta^2) < \infty$ in the previous section, we have imposed a stronger moment condition $E(|\eta|^{2+\gamma}) < \infty$ for some $\gamma > 0$ so that $\sup_j E(|\varepsilon_j|^{2+\gamma}) < \infty$. This condition may be dispensed with if we can assume $E(\sup_j \varepsilon_j^2) < \infty$ (Problem 7.6). In fact we have the following corollary.

Corollary 3.5. *Let $\{Y_n\}$ be defined as in Theorem 3.9. Suppose that $\{\varepsilon_j\}$ in (3.47) is a sequence of square integrable martingale differences such that*

i) $E(\sup_j \varepsilon_j^2) < \infty$ *;*

ii) $\frac{1}{n}\sum_{j=1}^{n} E(\varepsilon_j^2|\mathcal{F}_{j-1})$ *converges in probability to* $\sigma_\varepsilon^2 > 0$ *.*

Then $\mathcal{L}(Y_n) \to \mathcal{L}(\alpha w)$ *.*

Problems

7.1 Derive the expression in (3.49).

7.2 Prove the inequality in (3.54).

7.3 Show that $\{\tilde{\varepsilon}_j\}$ in (3.53) satisfies the relation (3.55) if $\{\varepsilon_j^2\}$ is s.u.i. and (3.48) holds.

7.4 Show that $\sup_j E(|\tilde{\varepsilon}_j|^{2+\gamma})$ for some $\gamma > 0$ implies uniform integrability of $\{\tilde{\varepsilon}_j^2\}$.

7.5 Prove that, if there exists η with $E(|\eta|^{2+\gamma}) < \infty$ for some $\gamma > 0$ such that $P(|\varepsilon_j| > x) \leq cP\ (|\eta| > x)$ for each $x \geq 0$, each integer j and some $c > 0$, then $\sup_j E(|\varepsilon_j|^{2+\gamma}) < \infty$ and $\sup_j E(|\tilde{\varepsilon}_j|^{2+\gamma}) < \infty$.

7.6 Prove that, if $E(\sup_j \varepsilon_j^2) < \infty$ and (3.48) holds, then (3.57) is ensured.

3.8. Weak convergence to the integrated Brownian motion

In Section 4 of Chapter 2 we introduced the integrated Brownian motion and indicated that the so-called I(d) process is essentially the $(d-1)$-fold integrated Brownian motion. In this section we refine this fact on the basis of results obtained so far in this chapter.

Let us first discuss weak convergence to the one-fold integrated Brownian motion $\{F_1(t)\}$ defined by

$$F_1(t) = \int_0^t w(s)ds\,, \tag{3.58}$$

where $\{w(s)\}$ is the one-dimensional standard Brownian motion. Let us construct the I(2) process $\{y_j^{(2)}\}$ generated by

$$(1-L)^2 y_j^{(2)} = \varepsilon_j\,, \qquad y_{-1}^{(2)} = y_0^{(2)} = 0\,, \qquad (j = 1, \cdots, n)\,, \tag{3.59}$$

where we assume, for the time being, that $\{\varepsilon_j\}$ is i.i.d.$(0, \sigma^2)$ with $\sigma^2 > 0$. Note that (3.59) can be rewritten (Problem 8.1) as

$$y_j^{(2)} = y_{j-1}^{(2)} + y_j^{(1)} = y_1^{(1)} + \cdots + y_j^{(1)}\,, \tag{3.60}$$

where $\{y_j^{(1)}\}$ is the I(1) process or the random walk following $y_j^{(1)} = y_{j-1}^{(1)} + \varepsilon_j$, $y_0^{(1)} = 0$. Define two sequences $\{Y_n^{(1)}\}$ and $\{Y_n^{(2)}\}$ of stochastic processes in C by

$$\begin{aligned} Y_n^{(1)}(t) &= \frac{1}{\sqrt{n}} y_{[nt]}^{(1)} + (nt - [nt]) \frac{1}{\sqrt{n}} \varepsilon_{[nt]+1} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{j} \varepsilon_i + n\left(t - \frac{j}{n}\right) \frac{1}{\sqrt{n}} \varepsilon_j\,, \qquad \left(\frac{j-1}{n} \leq t \leq \frac{j}{n}\right)\,, \end{aligned} \tag{3.61}$$

$$(3.62) \qquad Y_n^{(2)}(t) = \frac{1}{n\sqrt{n}} y_{[nt]}^{(2)} + (nt-[nt]) \frac{1}{n\sqrt{n}} y_{[nt]+1}^{(1)}$$
$$= \frac{1}{n}\sum_{i=1}^{j} Y_n^{(1)}\left(\frac{i}{n}\right) + n\left(t-\frac{j}{n}\right)\frac{1}{n\sqrt{n}} y_j^{(1)}, \qquad \left(\frac{j-1}{n} \le t \le \frac{j}{n}\right).$$

It follows from Donsker's theorem that $\mathcal{L}(Y_n^{(1)}/\sigma) \to \mathcal{L}(w)$.

We now show that $\mathcal{L}(Y_n^{(2)}/\sigma) \to \mathcal{L}(F_1)$. For this purpose define the integral version of (3.62) by

$$G_{1n}(t) = \int_0^t Y_n^{(1)}(s)ds .$$

Note that $\mathcal{L}(G_{1n}/\sigma) \to \mathcal{L}(F_1)$ by the continuous mapping theorem. Then it holds (Problem 8.2) that, for $(j-1)/n \le t \le j/n$,

$$(3.63) \qquad |Y_n^{(2)}(t) - G_{1n}(t)| \le \left|\sum_{i=1}^{j} \int_{\frac{i-1}{n}}^{\frac{i}{n}} Y_n^{(1)}\left(\frac{i}{n}\right) ds - \int_0^t Y_n^{(1)}(s)ds\right| + \frac{1}{n\sqrt{n}} \left|y_j^{(1)}\right|$$
$$\le \frac{2}{\sqrt{n}} \max_{1\le j\le n} |\varepsilon_j| + \frac{1}{n} \sup_{0\le t\le 1} \left|Y_n^{(1)}(t)\right| .$$

Now it can be shown (Problem 8.3) that

$$(3.64) \qquad \sup_{0\le t\le 1} |Y_n^{(2)}(t) - G_{1n}(t)| \longrightarrow 0 \qquad \text{in probability.}$$

This fact together with $\mathcal{L}(G_{1n}/\sigma) \to \mathcal{L}(F_1)$ establishes that $\mathcal{L}(Y_n^{(2)}/\sigma) \to \mathcal{L}(F_1)$.

Weak convergence to the general g-fold integrated Brownian motion can be dealt with similarly. Define, for a positive integer g,

$$(3.65) \qquad F_g(t) = \int_0^t F_{g-1}(s)ds , \qquad F_0(s) = w(s) ,$$

and construct the I(d) process $\{y_j^{(d)}\}$ generated by

$$(3.66) \qquad (1-L)^d y_j^{(d)} = \varepsilon_j , \qquad (j = 1, \cdots, n) ,$$

with $y_{-(d-1)}^{(d)} = y_{-(d-2)}^{(d)} = \cdots = y_0^{(d)} = 0$ and $\{\varepsilon_j\}$ being i.i.d.$(0, \sigma^2)$. We have

$$(3.67) \qquad y_j^{(d)} = y_{j-1}^{(d)} + y_j^{(d-1)} = y_1^{(d-1)} + \cdots + y_j^{(d-1)} , \qquad y_j^{(0)} = \varepsilon_j$$

and put, for $d \geq 2$,

$$(3.68) \qquad Y_n^{(d)}(t) = \frac{1}{n^{d-\frac{1}{2}}} y_{[nt]}^{(d)} + (nt - [nt]) \frac{1}{n^{d-\frac{1}{2}}} y_{[nt]+1}^{(d-1)}$$
$$= \frac{1}{n} \sum_{i=1}^{j} Y_n^{(d-1)}\left(\frac{i}{n}\right) + n\left(t - \frac{j}{n}\right) \frac{1}{n^{d-\frac{1}{2}}} y_j^{(d-1)}, \quad \left(\frac{j-1}{n} \leq t \leq \frac{j}{n}\right).$$

Define also the integral version of (3.68) by

$$G_{d-1,n}(t) = \int_0^t Y_n^{(d-1)}(s) ds\,.$$

We now prove by induction that $\mathcal{L}(Y_n^{(d)}/\sigma) \to \mathcal{L}(F_{d-1})$ for any $d \geq 2$. The case $d = 2$ was already established. Suppose that $\mathcal{L}(Y_n^{(k-1)}/\sigma) \to \mathcal{L}(F_{k-2})$ holds for some $k \geq 3$ so that $\mathcal{L}(G_{k-1,n}/\sigma) \to \mathcal{L}(F_{k-1})$ by the continuous mapping theorem. Then we have (Problem 8.4), for $(j-1)/n \leq t \leq j/n$,

$$(3.69) \qquad |Y_n^{(k)}(t) - G_{k-1,n}(t)| \leq \left| \sum_{i=1}^{j} \int_{\frac{i-1}{n}}^{\frac{i}{n}} Y_n^{(k-1)}\left(\frac{i}{n}\right) ds - \int_0^t Y_n^{(k-1)}(s) ds \right| + \frac{1}{n^{k-\frac{1}{2}}} \left| y_j^{(k-1)} \right|$$
$$\leq \frac{2}{\sqrt{n}} \max_{1 \leq j \leq n} |\varepsilon_j| + \frac{1}{n} \sup_{0 \leq t \leq 1} \left| Y_n^{(k-1)}(t) \right|.$$

Thus it is seen that $\sup_{0 \leq t \leq 1} |Y_n^{(k)}(t) - G_{k-1,n}(t)|$ converges in probability to 0. Since $\mathcal{L}(G_{k-1,n}/\sigma) \to \mathcal{L}(F_{k-1})$ by assumption, (3.69) yields that $\mathcal{L}(Y_n^{(k)}/\sigma) \to \mathcal{L}(F_{k-1})$.

The above arguments can be easily extended to the case where the innovation sequence $\{\varepsilon_j\}$ follows a linear process. We state an extended result in the following theorem, whose proof is left as Problem 8.5.

Theorem 3.10. *Suppose that the $I(d)$ process $\{y_j^{(d)}\}$ is generated by*

$$(1-L)^d y_j^{(d)} = u_j\,, \qquad (d \geq 2, \quad j = 1, \cdots, n)\,,$$

where $y_{-(d-1)}^{(d)} = y_{-(d-2)}^{(d)} = \cdots = y_0^{(d)} = 0$ and

$$u_j = \sum_{l=0}^{\infty} \alpha_l\, \varepsilon_{j-l}\,, \quad \alpha_0 = 1\,, \quad \sum_{l=0}^{\infty} l\, |\alpha_l| < \infty\,,$$

with $\{\varepsilon_j\}$ being i.i.d.$(0, \sigma^2)$. Define $F_{d-1}(t)$ and $Y_n^{(d)}(t)$ by (3.65) and (3.68), respectively. Then $\mathcal{L}(Y_n^{(d)}/\sigma) \to \mathcal{L}(\alpha F_{d-1})$, where $\alpha = \sum_{l=0}^{\infty} \alpha_l$.

As the first application of Theorem 3.10 we establish the weak convergence of

$$S_T = \frac{1}{T^{2d}} \sum_{j=1}^{T} \left(y_j^{(d)} \right)^2 , \qquad (d \geq 2) . \tag{3.70}$$

For this purpose we put

$$h(x) = \int_0^1 x^2(t) dt , \qquad x \in C . \tag{3.71}$$

Using (3.68) and noting that $Y_T^{(d)}(j/T) = y_j^{(d)}/T^{d-\frac{1}{2}}$ we consider

$$\begin{aligned} S_T - h\left(Y_T^{(d)}\right) &= \frac{1}{T} \sum_{j=1}^{T} \left(Y_T^{(d)}\left(\frac{j}{T}\right)\right)^2 - \int_0^1 \left(Y_T^{(d)}(t)\right)^2 dt \\ &= \sum_{j=1}^{T} \int_{\frac{j-1}{T}}^{\frac{j}{T}} \left[\left(Y_T^{(d)}\left(\frac{j}{T}\right)\right)^2 - \left(Y_T^{(d)}(t)\right)^2 \right] dt , \end{aligned}$$

where the integrand has the following bound:

$$\left| \left(Y_T^{(d)}\left(\frac{j}{T}\right)\right)^2 - \left(Y_T^{(d)}(t)\right)^2 \right| \leq 2 \sup_{0 \leq t \leq 1} \left| Y_T^{(d)}(t) \right| \max_{1 \leq j \leq T} \frac{\left| y_j^{(d-1)} \right|}{T^{d-\frac{1}{2}}} . \tag{3.72}$$

Thus it can be shown (Problem 8.6) that

$$S_T - h\left(Y_T^{(d)}\right) \longrightarrow 0 \qquad \text{in probability.} \tag{3.73}$$

Since $\mathcal{L}\left(Y_T^{(d)}/\sigma\right) \rightarrow \mathcal{L}(\alpha F_{d-1})$ by Theorem 3.10 and $\mathcal{L}\left(h\left(Y_T^{(d)}/\sigma\right)\right) \rightarrow \mathcal{L}(h(\alpha F_{d-1}))$ by the continuous mapping theorem, (3.71) and (3.73) lead us to

$$\mathcal{L}\left(\frac{S_T}{\sigma^2}\right) \longrightarrow \mathcal{L}\left(\alpha^2 \int_0^1 F_{d-1}^2(t) dt\right) . \tag{3.74}$$

As the second application we establish the weak convergence of

$$\begin{aligned} U_T &= \frac{1}{T^{2d-1}} \sum_{j=1}^{T} y_{j-1}^{(d)} \left(y_j^{(d)} - y_{j-1}^{(d)} \right) \\ &= \frac{1}{2T^{2d-1}} \left(y_T^{(d)} \right)^2 - \frac{1}{2T^{2d-1}} \sum_{j=1}^{T} \left(y_j^{(d)} - y_{j-1}^{(d)} \right)^2 . \end{aligned} \tag{3.75}$$

Since $\left\{ y_j^{(d)} - y_{j-1}^{(d)} \right\}$ is the I$(d-1)$ process because of (3.67), it is easy to see that the last term on the right side of (3.75) converges in probability to 0 for $d \geq 2$. Then we have, by Theorem 3.10 and the continuous mapping theorem,

$$\mathcal{L}\left(\frac{U_T}{\sigma^2}\right) \rightarrow \mathcal{L}\left(\frac{\alpha^2}{2} F_{d-1}^2(1)\right) = \mathcal{L}\left(\alpha^2 \int_0^1 F_{d-1}(t) dF_{d-1}(t)\right) .$$

The equality above is due to $(d-1)$-times differentiability of $\{F_{d-1}(t)\}$. Note that the situation is completely different from the case $d = 1$, which will be discussed in Section 11.

Problems

8.1 Derive the recursive relations (3.60) and (3.67).

8.2 Prove the inequalities in (3.63).

8.3 Show that (3.64) holds.

8.4 Prove the inequalities in (3.69).

8.5 Prove Theorem 3.10.

8.6 Show that (3.73) holds.

3.9. Weak convergence to the Ornstein-Uhlenbeck process

We introduced in Section 7 of Chapter 2 the Ornstein-Uhlenbeck (O-U) process in C defined by

$$(3.76) \qquad dX(t) = -\beta X(t)dt + dw(t) \Longleftrightarrow X(t) = e^{-\beta t}X(0) + e^{-\beta t}\int_0^t e^{\beta s}dw(s)\,,$$

where β is a constant. We will show in this section that the near random walk process

$$(3.77) \qquad y_j = \left(1 - \frac{\beta}{n}\right) y_{j-1} + \varepsilon_j\,, \qquad (j = 1, \cdots, n),$$

converges weakly to the O-U process in the sense described later, where $\{\varepsilon_j\}$ is assumed to be i.i.d.$(0, \sigma^2)$. Note that (3.77) may be rewritten (Problem 9.1) using Abel's transformation as

$$\begin{aligned}(3.78) \qquad y_j &= \rho_n^j y_0 + \rho_n^{j-1}\varepsilon_1 + \cdots + \rho_n \varepsilon_{j-1} + \varepsilon_j \\ &= \rho_n^j y_0 + \sum_{i=1}^{j} \rho_n^{j-i}(S_i - S_{i-1}) \\ &= \rho_n^j y_0 + \rho_n^{-1} S_j - (1-\rho_n)\sum_{i=1}^{j} \rho_n^{j-i-1} S_i\,,\end{aligned}$$

where $\rho_n = 1 - (\beta/n)$ and

(3.79) $$S_j = \varepsilon_1 + \cdots + \varepsilon_j\,, \qquad S_0 \equiv 0\,.$$

The last expression for y_j in (3.78) is quite useful for subsequent discussions. Note also in (3.78) that we retain the initial value y_0, which we assume, for the time being, to take the following form:

(3.80) $$y_0 = \sqrt{n}\gamma\sigma\,,$$

where γ is a constant. We will consider later the case where y_0 is a random variable of stochastic order $\sqrt{n}$.

The present problem was studied to a large extent by Bobkoski (1983) for the case where y_0 is a constant or a random variable distributed independently of n. In that case y_0 is asymptotically negligible. Let us define, for $(j-1)/n \le t \le j/n$,

(3.81) $$\begin{aligned} X_n(t) &= \frac{1}{\sqrt{n}}\,y_{j-1} + n\left(t - \frac{j-1}{n}\right)\frac{y_j - y_{j-1}}{\sqrt{n}} \\ &= \frac{\rho_n^{j-1}}{\sqrt{n}}\,y_0 + \frac{\rho_n^{-1}}{\sqrt{n}}\,S_{j-1} - \frac{\beta}{n\sqrt{n}}\sum_{i=1}^{j-1}\rho_n^{j-i-2}S_i \\ &\qquad + n\left(t - \frac{j-1}{n}\right)\frac{y_j - y_{j-1}}{\sqrt{n}} \\ &= \rho_n^{j-1}\gamma\sigma + \rho_n^{-1}Y_n\left(\frac{j-1}{n}\right) - \frac{\beta}{n}\sum_{i=1}^{j-1}\rho_n^{j-i-2}\,Y_n\left(\frac{i}{n}\right) \\ &\qquad + n\left(t - \frac{j-1}{n}\right)\frac{y_j - y_{j-1}}{\sqrt{n}}\,, \end{aligned}$$

where $X_n(0) = \gamma\sigma$ and

(3.82) $$Y_n(t) = \frac{1}{\sqrt{n}}\,S_{j-1} + n\left(t - \frac{j-1}{n}\right)\frac{1}{\sqrt{n}}\,\varepsilon_j\,, \qquad \left(\frac{j-1}{n} \le t \le \frac{j}{n}\right).$$

Note that (3.78) has been applied to obtain the second expression in (3.81).

We also consider a function $h(y;\gamma)$ on C, whose value at t denoted as $h_t(y;\gamma)$ is defined by

(3.83) $$h_t(y;\gamma) = e^{-\beta t}\gamma + y(t) - \beta e^{-\beta t}\int_0^t e^{\beta s}y(s)ds\,.$$

It is easy to check (Problem 9.2) that h is a continuous mapping defined on C. We shall show that $\mathcal{L}(X_n/\sigma) \to \mathcal{L}(h(w;\gamma))$ so that $\mathcal{L}(X_n/\sigma) \to \mathcal{L}(X)$ with $X(0) = \gamma$ since it holds (Problem 9.3) that $X(t) = h_t(w;\gamma)$. For this purpose let us consider

$$(3.84) \qquad |X_n(t) - h_t(Y_n;\gamma\sigma)| \le |\gamma|\sigma A_{jn} + B_{jn} + |\beta| C_{jn} + D_{jn}\,,$$

where

$$\begin{aligned}
A_{jn} &= |\rho_n^{j-1} - e^{-\beta t}|\,,\\
B_{jn} &= |\rho_n^{-1} Y_n\left(\frac{j-1}{n}\right) - Y_n(t)|\,,\\
C_{jn} &= \left|\frac{1}{n}\sum_{i=1}^{j-1} \rho_n^{j-i-2}\, Y_n\left(\frac{i}{n}\right) - e^{-\beta t}\int_0^t e^{\beta s}\, Y_n(s) ds\right|\,,\\
D_{jn} &= \frac{1}{\sqrt{n}}\,|y_j - y_{j-1}|\,.
\end{aligned}$$

It can be shown (Problem 9.4) that

$$\begin{aligned}
A_{jn} &\le \sup_{0\le t\le 1} |\rho_n^{[nt]} - e^{-\beta t}| = O\left(\frac{1}{n}\right)\,,\\
B_{jn} &\le \frac{1}{\sqrt{n}} \max_{1\le j\le n} |\varepsilon_j| + |\rho_n^{-1} - 1| \sup_{0\le t\le 1} |Y_n(t)| = o_p(1)\,,\\
C_{jn} &\le \sum_{i=1}^{j-1} \int_{\frac{i-1}{n}}^{\frac{i}{n}} e^{-\beta(t-s)} \left|Y_n\left(\frac{i}{n}\right) - Y_n(s)\right| ds\\
&\qquad + \sum_{i=1}^{j-1} \int_{\frac{i-1}{n}}^{\frac{i}{n}} |\rho_n^{j-i-2} - e^{-\beta(t-s)}| ds \left|Y_n\left(\frac{i}{n}\right)\right|\\
&\qquad + \int_{\frac{[nt]}{n}}^{t} e^{-\beta(t-s)}\, |Y_n(s)|\, ds\\
&\le \frac{O(1)}{\sqrt{n}} \max_{1\le j\le n} |\varepsilon_j| + o(1) \sup_{0\le t\le 1} |Y_n(t)| = o_p(1)\,,\\
D_{jn} &\le \frac{|\beta|}{n\sqrt{n}} \max_{1\le j\le n} |y_j| + \frac{1}{\sqrt{n}} \max_{1\le j\le n} |\varepsilon_j| = o_p(1)\,.
\end{aligned}$$

Thus $\sup_{0\le t\le 1} |X_n(t)/\sigma - h_t(Y_n/\sigma;\gamma)| \to 0$ in probability. Since $\mathcal{L}(Y_n/\sigma) \to \mathcal{L}(w)$, we have $\mathcal{L}(X_n/\sigma) \to \mathcal{L}(h(w;\gamma))$ by the continuous mapping theorem and it follows from Problem 9.3 that $\mathcal{L}(X_n/\sigma) \to \mathcal{L}(X)$.

The above arguments are summarized in the following theorem.

Theorem 3.11. *Let the near random walk* $\{y_j\}$ *be defined by (3.77). On the basis of* $\{y_j\}$ *construct the process* $\{X_n(t)\}$ *in* C *as in (3.81) with* $y_0 = \sqrt{n}\gamma\sigma$. *Then* $\mathcal{L}(X_n/\sigma) \to \mathcal{L}(X)$, *where* $\{X(t)\}$ *is the O-U process defined in (3.76) with* $X(0) = \gamma$.

The case of y_0 being a random variable needs some care. Suppose that

$$y_0 = \sqrt{n}\sigma X(0)\,, \tag{3.85}$$

where $X(0) \sim \mathrm{N}(\gamma, \delta^2)$ and is independent of $\{\varepsilon_j\}$. Then we construct, as in (3.81),

$$\begin{aligned} X_n(t) \;&=\; \rho_n^{j-1}\sigma X(0) + \rho_n^{-1} Y_n\left(\frac{j-1}{n}\right) - \frac{\beta}{n}\sum_{i=1}^{j-1}\rho_n^{j-i-2}\,Y_n\left(\frac{i}{n}\right) \\ &\quad + n\left(t - \frac{j-1}{n}\right)\frac{y_j - y_{j-1}}{\sqrt{n}}\,, \end{aligned} \tag{3.86}$$

which is composed of a random variable $X(0)$ and a stochastic process $\{Y_n(t)\}$ defined in (3.82). Let R be the real line, which is complete and separable under the Euclidean metric. Then the joint weak convergence of $(X(0), Y_n/\sigma)$ on $R \times C$ holds, that is, $\mathcal{L}(X(0), Y_n/\sigma) \to \mathcal{L}(X(0), w)$ (see Billingsley (1968, p.224) and the next section). Defining on $R \times C$

$$h_t(x, y) = e^{-\beta t}x + y(t) - \beta e^{-\beta t}\int_0^t e^{\beta s}y(s)ds\,,$$

we can obtain that $\sup_{0\le t\le 1}|X_n(t)/\sigma - h_t(X(0), Y_n/\sigma)| \to 0$ in probability so that $\mathcal{L}(X_n/\sigma) \to \mathcal{L}(h(X(0), w)) = \mathcal{L}(X)$ with $X(0) \sim \mathrm{N}(\gamma, \delta^2)$.

Theorem 3.12. *Assume the same conditions as in Theorem 3.11 except that* $X_n(t)$ *is defined in (3.86) with* $y_0 = \sqrt{n}\sigma X(0)$ *and* $X(0) \sim N(\gamma, \delta^2)$. *Then* $\mathcal{L}(X_n/\sigma) \to \mathcal{L}(X)$ *with* $\{X(t)\}$ *being the O-U process.*

In this theorem we may put $\delta = 0$. Then $y_0 = \sqrt{n}\sigma\gamma$ so that the theorem reduces to Theorem 3.11. If we assume that $\gamma = 0$ and $\delta^2 = 1/(2\beta)$ with $\beta > 0$, then $\{X(t)\}$ becomes stationary, as was indicated in (2.64).

As an application let us establish the weak convergence of

$$V_T = \frac{1}{T^2}\sum_{j=1}^{T} y_j^2\,, \tag{3.87}$$

where $\{y_j\}$ is the near random walk defined in (3.77). As for y_0 we assume that $y_0 = \sqrt{T}\sigma X(0)$ with $X(0) \sim \mathrm{N}(\gamma, \delta^2)$ and put

$$h(y) = \int_0^1 y^2(t)dt \,.$$

Using $X_T(t)$ defined in (3.86) and noting that $X_T(j/T) = y_j/\sqrt{T}$, it can be easily shown (Problem 9.5) that $V_T - h(X_T)$ converges in probability to 0. Since $h(X_T/\sigma) \to h(X)$ by Theorem 3.12 and the continuous mapping theorem, we have that

$$\mathcal{L}\left(\frac{V_T}{\sigma^2}\right) \longrightarrow \mathcal{L}\left(\int_0^1 X^2(t)dt\right) \,.$$

Extensions to near integrated processes seem straightforward. Consider the near integrated process $\{y_j\}$ defined by

$$y_j = \left(1 - \frac{\beta}{n}\right) y_{j-1} + u_j \,, \tag{3.88}$$

$$u_j = \sum_{l=0}^{\infty} \alpha_l \, \varepsilon_{j-l} \,, \quad \alpha_0 = 1 \,, \quad \sum_{l=0}^{\infty} l \, |\alpha_l| < \infty \,, \tag{3.89}$$

where $\{\varepsilon_j\} \sim$ i.i.d.$(0, \sigma^2)$. The proof of the following theorem is left as Problem 9.6.

Theorem 3.13. *Assume that the near integrated process $\{y_j\}$ is defined by (3.88) and (3.89) with $y_0 = \sqrt{n}\alpha\sigma X(0)$, where $\alpha = \sum_{l=0}^{\infty} \alpha_l$ and $X(0) \sim N(\gamma, \delta^2)$. Define*

$$X_n(t) = \frac{1}{\sqrt{n}} y_{j-1} + n\left(t - \frac{j-1}{n}\right) \frac{y_j - y_{j-1}}{\sqrt{n}} \,, \qquad \left(\frac{j-1}{n} \le t \le \frac{j}{n}\right) \,.$$

Then $\mathcal{L}(X_n/\sigma) \to \mathcal{L}(\alpha X)$, where $\{X(t)\}$ is the O-U process with $X(0) \sim N(\gamma, \delta^2)$.

The extension to the case where $\{\varepsilon_j\}$ is a sequence of square integrable martingale differences is also straightforward. We do not pursue the matter here.

Problems

9.1 Establish the relations in (3.78) by showing that

$$\sum_{j=1}^{n} a_j (b_j - b_{j-1}) = a_{n+1} b_n - a_1 b_0 - \sum_{j=1}^{n} (a_{j+1} - a_j) b_j \,.$$

9.2 Prove that $h(y\,;\gamma)$ in (3.83) is a continuous mapping defined on C.

9.3 Show that $X(t) = h_t(w;\gamma)$ with $X(0) = \gamma$, where $X(t)$ and h_t are defined by (3.76) and (3.83), respectively.

9.4 Prove that $\sup_{0\le t\le 1} |X_n(t) - h_t(Y_n;\gamma\sigma)|$ in (3.84) converges in probability to 0.

9.5 Establish the weak convergence of V_T in (3.87).

9.6 Prove Theorem 3.13.

3.10. Weak convergence of vector-valued stochastic processes

Our discussions have so far been concerned with weak convergence of scalar stochastic processes, although an exception is found in Section 9. In practice we need to deal with vector processes, for which we describe here the FCLT's.

3.10.1. Space C^q

Let $(C^q, \mathcal{B}(C^q))$ be a measurable space, where $C^q = C[0,1]\times\cdots\times C[0,1]$ (q copies) and $\mathcal{B}(C^q)$ is the σ-field generated by the subsets of C^q that are open with respect to the metric ρ_q defined by

$$\rho_q(x,y) = \max_{1\le i\le q}\ \sup_{0\le t\le 1} |x_i(t) - y_i(t)| \tag{3.90}$$

for $x = (x_1,\cdots,x_q)'$, $y = (y_1,\cdots,y_q)' \in C^q$. The space C^q is complete and separable under ρ_q. In particular, separability results from C being separable under the uniform metric and it holds that $\mathcal{B}(C^q) = \mathcal{B}(C)\times\cdots\times\mathcal{B}(C)$ (q copies) (Billingsley (1968, p.224)). Separability also implies that, for given probability measures $P^{(i)} (i = 1,\cdots,q)$ on $(C,\mathcal{B}(C))$, the product measure $P^{(1)}\times\cdots\times P^{(q)}$ is a probability measure on $\mathcal{B}(C^q)$ (Billingsley (1968, p.21)).

Let $\{X_n\}$ be a sequence of q-dimensional stochastic processes in C^q, and $\{Q_n\}$ the family of probability measures induced by $\{X_n\}$ as

$$Q_n(A) = P(X_n\in A) = P(X_n^{-1}(A))\,, \qquad A\in\mathcal{B}(C^q)\,.$$

As in the scalar case $\mathcal{L}(X_n) \to \mathcal{L}(X)$ if the finite-dimensional distributions of Q_n converge weakly to those of the probability measure Q induced by X, and if $\{Q_n\}$ is tight. It is usually difficult to prove the joint weak convergence, but separability and independence facilitate its proof. Suppose that the q components of X_n are independent of each other ; so are the q components of X. Suppose further that $\mathcal{L}(X_n(t_1), \cdots, X_n(t_k)) \to \mathcal{L}(X(t_1), \cdots, X(t_k))$ for each finite k and each collection $0 \le t_1 < t_2 < \cdots < t_k \le 1$. Then $\mathcal{L}(X_n) \to \mathcal{L}(X)$ if all the marginal probability measures of $\{Q_n\}$ are tight on the component spaces (Billingsley (1968, p.41)).

In the next subsection we take up an example of $\{X_n\}$ and discuss its weak convergence following the ideas described above.

3.10.2. Basic FCLT for vector processes

Let us now consider a sequence $\{X_n\}$ of stochastic processes in C^q defined by

$$(3.91) \qquad X_n(t) = \Sigma^{-\frac{1}{2}} \left[\frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \varepsilon_j + (nt - [nt]) \frac{1}{\sqrt{n}} \varepsilon_{[nt]+1} \right],$$

where $\{\varepsilon_j\}$ is a sequence of q-dimensional i.i.d.$(0, \Sigma)$ random vectors on $(\Omega, \mathcal{F}, P)$ with $\sum > 0$. We shall show that $\mathcal{L}(X_n) \to \mathcal{L}(\underset{\sim}{w})$, where $\{\underset{\sim}{w}(t)\}$ is the q-dimensional standard Brownian motion.

We first note that, for a single time point t,

$$\|X_n(t) - \frac{\Sigma^{-\frac{1}{2}}}{\sqrt{n}} \sum_{j=1}^{[nt]} \varepsilon_j\| \le \|\Sigma^{-\frac{1}{2}}\| \frac{1}{\sqrt{n}} \|\varepsilon_{[nt]+1}\|,$$

where $\|M\| = [\mathrm{tr}\,(M'M)]^{\frac{1}{2}}$ for any matrix or vector M. Since

$$(3.92) \qquad \frac{1}{\sqrt{n}} \|\varepsilon_{[nt]+1}\| \longrightarrow 0 \qquad \text{in probability}$$

by Chebyshev's inequality, and

$$\mathcal{L}\left(\frac{\Sigma^{-\frac{1}{2}}}{\sqrt{n}} \sum_{j=1}^{[nt]} \varepsilon_j \right) \longrightarrow \mathcal{L}(\underset{\sim}{w}(t))$$

by the multivariate CLT (see, for example, Rao (1973, p.128)), it follows from the vector version of Theorem 3.6 that $\mathcal{L}(X_n(t)) \to \mathcal{L}(\underset{\sim}{w}(t))$.

Consider next two time points s and t with $s < t$. We are to prove $\mathcal{L}(X_n(s), X_n(t)) \to \mathcal{L}(\underset{\sim}{w}(s), \underset{\sim}{w}(t))$, which will follow by the continuous mapping theorem if we prove $\mathcal{L}(X_n(s), X_n(t) - X_n(s)) \to \mathcal{L}(\underset{\sim}{w}(s), \underset{\sim}{w}(t) - \underset{\sim}{w}(s))$. Because of (3.92) it is enough to prove

$$\mathcal{L}\left(\frac{\Sigma^{-\frac{1}{2}}}{\sqrt{n}}\sum_{j=1}^{[nt]}\varepsilon_j\,,\ \ \frac{\Sigma^{-\frac{1}{2}}}{\sqrt{n}}\left(\sum_{j=1}^{[nt]}\varepsilon_j - \sum_{j=1}^{[ns]}\varepsilon_j\right)\right) \longrightarrow \mathcal{L}(\underset{\sim}{w}(s), \underset{\sim}{w}(t) - \underset{\sim}{w}(s))\,.$$

Since the two vectors on the left are independent, this follows by the multivariate CLT for each vector (Billingsley (1968, p.26)). A set of three or more time points can be treated in the same way, and hence the finite-dimensional distributions converge properly.

As was described in the previous subsection tightness of the family $\{Q_n\}$ of probability measures induced by $\{X_n\}$ is ensured if all the marginal probability measures associated with each component $\{X_{in}\}$ of $\{X_n\}$ are tight on C. Since it does hold that $\mathcal{L}(X_{in}) \to \mathcal{L}(w)$, where $\{w(t)\}$ is the one-dimensional standard Brownian motion, the associated marginal probability measures must be relatively compact (Billingsley (1968, p.35)). Thus tightness results from completeness and separability of C under the uniform metric.

It is an immediate consequence of the above discussions and the continuous mapping theorem to obtain

$$\mathcal{L}\left(\frac{1}{T^2}\sum_{j=1}^{T} y_j y_j'\right) \longrightarrow \mathcal{L}\left(\Sigma^{\frac{1}{2}}\int_0^1 \underset{\sim}{w}(t)\underset{\sim}{w}'(t)dt\,\Sigma^{\frac{1}{2}}\right)\,, \tag{3.93}$$

where $y_j = y_{j-1} + \varepsilon_j$, $y_0 = 0$ and $\{\varepsilon_j\} \sim$ i.i.d.$(0, \Sigma)$. It can also be shown (Problem 10.1) that

$$\mathcal{L}\left(\frac{1}{T^2}\sum_{j=1}^{T} y_j' H' H y_j\right) \longrightarrow \mathcal{L}\left(\int_0^1 \underset{\sim}{w}'(t)\Sigma^{\frac{1}{2}}H'H\Sigma^{\frac{1}{2}}\underset{\sim}{w}(t)dt\right) \tag{3.94}$$

for any $q \times q$ constant matrix H. In particular suppose that any two components $\{y_{kj}\}$ and $\{y_{lj}\}$ $(k \neq l)$ of $\{y_j\}$ are independent and $\Sigma = I_q$. It then follows from (3.94) that

$$\mathcal{L}\left(\frac{1}{T^2}\sum_{j=1}^{T} y_{kj} y_{lj}\right) \longrightarrow \mathcal{L}\left(\int_0^1 w_k(t)w_l(t)dt\right)\,.$$

A special case of this was dealt with in Section 4 of Chapter 1 together with the c.f. of the limiting distribution.

3.10.3. FCLT for vector-valued linear processes

Similar arguments can be applied to stochastic processes defined by

$$Y_n(t) = A^{-1}\left[\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} u_j + (nt-[nt])\frac{1}{\sqrt{n}}\, u_{[nt]+1}\right], \tag{3.95}$$

$$u_j = \sum_{l=0}^{\infty} A_l \varepsilon_{j-l}\,, \qquad \sum_{l=0}^{\infty} l\,\|A_l\| < \infty\,, \tag{3.96}$$

$$A = \sum_{l=0}^{\infty} A_l\,, \tag{3.97}$$

where $\{\varepsilon_j\} \sim$ i.i.d.$(0, I_q)$, $\|A_l\| = [\mathrm{tr}\,(A_l'A_l)]^{\frac{1}{2}}$ and A is nonsingular. Note that $V(\varepsilon_j) = I_q$ and we do not assume $A_0 = I_q$, but do assume A_0 to be nonsingular and block lower triangular.

The BN decomposition used before is also applied to decompose the vector u_j into

$$u_j = A\varepsilon_j + \tilde{\varepsilon}_{j-1} - \tilde{\varepsilon}_j\,, \qquad \tilde{\varepsilon}_j = \sum_{l=0}^{\infty} \tilde{A}_l \varepsilon_{j-l}\,, \qquad \tilde{A}_l = \sum_{k=l+1}^{\infty} A_k\,. \tag{3.98}$$

The sequence $\{\tilde{\varepsilon}_j\}$ also becomes stationary (Problem 10.2) with $0 < \|A\| < \infty$ and $0 < \sum_{l=0}^{\infty}\|\tilde{A}_l\| < \infty$. We now have

$$Y_n(t) = X_n(t) + R_n(t)\,,$$

where

$$\begin{aligned} X_n(t) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} \varepsilon_j + (nt-[nt])\frac{1}{\sqrt{n}}\,\varepsilon_{[nt]+1}\,, \\ R_n(t) &= A^{-1}\left[\frac{1}{\sqrt{n}}\,(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_{[nt]}) + (nt-[nt])\frac{1}{\sqrt{n}}\,(\tilde{\varepsilon}_{[nt]} - \tilde{\varepsilon}_{[nt]+1})\right]. \end{aligned}$$

It can be shown (Problem 10.3) that

$$|R_{in}(t)| \le \|A^{-1}\|\left[\frac{1}{\sqrt{n}}\,\|\tilde{\varepsilon}_0\| + \frac{3}{\sqrt{n}}\max_{0\le j\le n}\|\tilde{\varepsilon}_j\|\right] \longrightarrow 0 \tag{3.99}$$

in probability so that $\rho_q(Y_n, X_n) \to 0$ in probability, where $R_{in}(t)$ is the i-th component of $R_n(t)$. Since $\mathcal{L}(X_n) \to \mathcal{L}(\underset{\sim}{w})$, we establish the following theorem using the vector version of Theorem 3.6.

Theorem 3.14. *Let the sequence $\{Y_n\}$ of q-dimensional stochastic processes be defined by (3.95), (3.96) and (3.97), where $\{\varepsilon_j\} \sim i.i.d.(0, I_q)$. Then $\mathcal{L}(Y_n) \to \mathcal{L}(\underset{\sim}{w})$.*

As an application we consider the weak convergence of

$$V_T = \frac{1}{T^2} \sum_{j=1}^{T} y_j y_j' , \tag{3.100}$$

where $y_j = y_{j-1} + u_j$, $y_0 = 0$ with $\{u_j\}$ being the linear process generated by (3.96) with $\{\varepsilon_j\} \sim$ i.i.d.$(0, I_q)$. Let us put

$$h(x) = \int_0^1 x(t)x'(t)dt , \qquad x \in C^q , \tag{3.101}$$

which is a continuous function of x (Problem 10.4). Using $Y_T(t)$ defined in (3.95) and noting that $Y_T(j/T) = A^{-1}y_j/\sqrt{T}$ we have

$$\begin{aligned} A^{-1}V_T(A^{-1})' - h(Y_T) &= \frac{1}{T^2}\sum_{j=1}^{T}(A^{-1}y_j)(A^{-1}y_j)' - \int_0^1 Y_T(t)Y_T'(t)dt \\ &= \sum_{j=1}^{T} \int_{\frac{j-1}{T}}^{\frac{j}{T}} \left[Y_T\left(\frac{j}{T}\right) Y_T'\left(\frac{j}{T}\right) - Y_T(t)Y_T'(t)\right] dt . \end{aligned}$$

The (k,l)-element of the integrand has the following bound (Problem 10.5) :

$$\begin{aligned} \left| Y_{kT}\left(\frac{j}{T}\right) Y_{lT}\left(\frac{j}{T}\right) - Y_{kT}(t)Y_{lT}(t)\right| &\leq \left| \left(Y_{kT}\left(\frac{j}{T}\right) - Y_{kT}(t)\right) Y_{lT}\left(\frac{j}{T}\right)\right| \\ &\quad + \left| \left(Y_{lT}\left(\frac{j}{T}\right) - Y_{lT}(t)\right) Y_{kT}(t)\right| \\ &\leq 2\|A^{-1}\| \sup_{0\leq t\leq 1} \|Y(t)\| \max_{1\leq j\leq T} \frac{\|u_j\|}{\sqrt{T}} , \end{aligned} \tag{3.102}$$

which converges in probability to 0. Since $\mathcal{L}(h(Y_T)) \to \mathcal{L}(h(\underset{\sim}{w}))$ by the continuous mapping theorem, $\mathcal{L}(A^{-1}V_T(A^{-1})') \to \mathcal{L}(h(\underset{\sim}{w}))$ and, using again the continuous mapping theorem, we establish that

$$\begin{aligned} \mathcal{L}(V_T) &\longrightarrow \mathcal{L}(Ah(\underset{\sim}{w})A') \\ &= \mathcal{L}(A\int_0^1 \underset{\sim}{w}(t)\underset{\sim}{w}'(t)dtA') . \end{aligned} \tag{3.103}$$

3.10.4. FCLT for the vector-valued integrated Brownian motion

In this subsection we extend the scalar I(d) ($d \geq 2$) processes discussed in Section 8 to vector processes, for which we establish the FCLT. Let us define

$$(3.104) \qquad (1-L)^d y_j^{(d)} = u_j\,, \qquad y_{-(d-1)}^{(d)} = y_{-(d-2)}^{(d)} = \cdots = y_0^{(d)} = 0\,,$$

$$(3.105) \qquad u_j = \sum_{l=0}^{\infty} A_l\, \varepsilon_{j-l}\,, \qquad \sum_{l=0}^{\infty} l\|A_l\| < \infty\,,$$

$$(3.106) \qquad A = \sum_{l=0}^{\infty} A_l\,,$$

where $\{\varepsilon_j\}$ is i.i.d.$(0, I_q)$. Note that $y_j^{(d)} = y_{j-1}^{(d)} + y_j^{(d-1)}$ with $y_j^{(0)} = u_j$.

We now consider, as in Section 8, a sequence $\{Y_n^{(d)}\}$ of q-dimensional stochastic processes defined by

$$\begin{aligned} Y_n^{(d)}(t) &= \frac{1}{n^{d-\frac{1}{2}}}\, y_{[nt]}^{(d)} + (nt-[nt])\frac{1}{n^{d-\frac{1}{2}}}\, y_{[nt]+1}^{(d-1)} \\ &= \frac{1}{n}\sum_{j=1}^{[nt]} Y_n^{(d-1)}\left(\frac{j}{n}\right) + (nt-[nt])\frac{1}{n^{d-\frac{1}{2}}}\, y_{[nt]+1}^{(d-1)}\,. \end{aligned}$$

It can be shown almost in the same way as in the scalar case (Problem 10.6) that

$$(3.107) \qquad \mathcal{L}(Y_n^{(2)}) \longrightarrow \mathcal{L}(AF_1)\,,$$

where $\{F_g(t)\}$ is the q-dimensional g-fold integrated Brownian motion defined by

$$F_g(t) = \int_0^t F_{g-1}(t)dt\,, \qquad F_0(t) = \underset{\sim}{w}(t)\,.$$

For general d (≥ 3) we can prove (Problem 10.7) by induction that

$$(3.108) \qquad \mathcal{L}(Y_n^{(d)}) \longrightarrow \mathcal{L}(AF_{d-1})\,.$$

As the first application let us consider the weak convergence of

$$\begin{aligned} S_T^{(d)} &= \frac{1}{T^{2d}}\sum_{j=1}^{T} y_j^{(d)}(y_j^{(d)})' \\ &= \frac{1}{T}\sum_{j=1}^{T} Y_T^{(d)}\left(\frac{j}{T}\right)\left(Y_T^{(d)}\left(\frac{j}{T}\right)\right)'\,. \end{aligned}$$

For this purpose put

$$h(x) = \int_0^1 x(t)x'(t)dt\,, \qquad x \in C^q\,.$$

Then we have (Problem 10.8)

$$\begin{aligned}
(3.109) \qquad \|S_T^{(d)} \;&-\; h(Y_T^{(d)})\| \\
&\le \sum_{j=1}^{T} \int_{\frac{j-1}{T}}^{\frac{j}{T}} \|Y_T^{(d)}\left(\frac{j}{T}\right)\left(Y_T^{(d)}\left(\frac{j}{T}\right)\right)' - Y_T^{(d)}(t)\left(Y_T^{(d)}(t)\right)'\|dt \\
&\le \frac{1}{T^{2d-1}} \max_{1\le j\le T} \|y_j^{(d-1)}\|^2 + 2\sup_{0\le t\le 1}\|Y_T^{(d)}(t)\|\,\frac{1}{T^{d-\frac{1}{2}}}\max_{1\le j\le T}\|y_j^{(d-1)}\|\,,
\end{aligned}$$

which evidently converges in probability to 0. Therefore we obtain, by the continuous mapping theorem,

$$\begin{aligned}
\mathcal{L}(S_T^{(d)}) \;&\longrightarrow\; \mathcal{L}(h(AF_{d-1})) \\
&=\; \mathcal{L}\left(A\int_0^1 F_{d-1}(t)F'_{d-1}(t)dtA'\right)\,.
\end{aligned}$$

As the second application we establish the weak convergence of

$$\begin{aligned}
U_T^{(d)} \;&=\; \frac{1}{T^{2d-1}}\sum_{j=1}^{T} y_{j-1}^{(d)}\left(y_j^{(d)} - y_{j-1}^{(d)}\right)' \\
&=\; \sum_{j=1}^{T} Y_T^{(d)}\left(\frac{j-1}{T}\right)\left[Y_T^{(d)}\left(\frac{j}{T}\right) - Y_T^{(d)}\left(\frac{j-1}{T}\right)\right]'\,.
\end{aligned}$$

It can be shown (Problem 10.9) that

$$(3.110) \qquad \int_0^1 Y_T^{(d)}(t)\left(dY_T^{(d)}(t)\right)' = U_T^{(d)} + \frac{1}{2T^{2d-1}}\sum_{j=1}^{T} y_j^{(d-1)}\left(y_j^{(d-1)}\right)'\,.$$

Here the last term on the right side converges in probability to the null matrix. Since $\mathcal{L}(Y_T^{(d)}) \to \mathcal{L}(AF_{d-1})$ and the limiting random vector is $(d-1)$ -times continuously differentiable, we obtain, for $d \ge 2$,

$$\begin{aligned}
\mathcal{L}(U_T^{(d)}) \;&\longrightarrow\; \mathcal{L}\left(A\int_0^1 F_{d-1}(t)(dF_{d-1}(t))'A'\right) \\
&=\; \mathcal{L}\left(A\int_0^1 F_{d-1}(t)F'_{d-2}(t)dtA'\right)\,.
\end{aligned}$$

Because of the nonsymmetric nature of $U_T^{(d)}$ we cannot reduce the final expression above to a simple form. We, however, obtain (Problem 10.10)

$$\mathcal{L}\left(U_T^{(d)} + (U_T^{(d)})'\right) \longrightarrow \mathcal{L}\left(AF_{d-1}(1)\, F'_{d-1}(1)A'\right) \tag{3.111}$$

for $d \geq 2$.

Note that the above results do not hold for $d = 1$, which we discuss in the next section.

Problems

10.1 Establish the weak convergence in (3.94).

10.2 Prove that $\{\tilde{\varepsilon}_j\}$ defined in (3.98) is second-order stationary.

10.3 Prove the inequality in (3.99).

10.4 Prove that the function $h(x)$ defined in (3.101) is a continuous function of x.

10.5 Establish the weak convergence in (3.103) by proving that the right side of (3.102) converges in probability to 0.

10.6 Establish the weak convergence in (3.107).

10.7 Establish the weak convergence in (3.108) for $d \geq 3$.

10.8 Prove the inequalities in (3.109).

10.9 Show that the relation in (3.110) holds.

10.10 Establish the weak convergence in (3.111).

3.11. Weak convergence to the Ito integral

As the final topic in this chapter we discuss weak convergence to the Ito integral introduced in Sections 5 and 6 of Chapter 2. The difficulty consists in the fact that we cannot use the continuous mapping theorem because of the unbounded variation property of the Brownian motion.

Let us first deal with the scalar case and consider

$$S_T = \frac{1}{T}\sum_{j=1}^{T} x_{j-1}(x_j - x_{j-1}),$$

where $\{x_j\}$ is the near random walk defined by

$$x_j = \left(1 - \frac{\beta}{T}\right) x_{j-1} + \varepsilon_j, \qquad (j = 1, \cdots, T),$$

with $\{\varepsilon_j\} \sim$ i.i.d.$(0,1)$. Suppose that $x_0 = \sqrt{T}X(0)$ with $X(0) \sim \mathrm{N}(\gamma, \delta^2)$ which is independent of $\{\varepsilon_j\}$. Then we shall show

$$\mathcal{L}(S_T) \longrightarrow \mathcal{L}\left(\int_0^1 X(t)dX(t)\right) = \mathcal{L}\left(\frac{1}{2}(X^2(1) - X^2(0) - 1)\right), \tag{3.112}$$

where $\{X(t)\}$ is the O-U process defined by

$$dX(t) = -\beta X(t)dt + dw(t) \Longleftrightarrow X(t) = e^{-\beta t}X(0) + e^{-\beta t}\int_0^t e^{\beta s}dw(s).$$

To establish (3.112) it is convenient to rewrite S_T as

$$\begin{aligned} S_T &= -\frac{1}{2T}\left[\sum_{j=1}^{T}(x_j - x_{j-1})^2 - \sum_{j=1}^{T} x_j^2 + \sum_{j=1}^{T} x_{j-1}^2\right] \\ &= \frac{1}{2T}(x_T^2 - x_0^2) - \frac{1}{2T}\sum_{j=1}^{T}\left(-\frac{\beta}{T}x_{j-1} + \varepsilon_j\right)^2, \end{aligned}$$

from which and Theorem 3.12 the latter expression on the right side of (3.112) evidently follows. The equivalence to the former is because of the Ito calculus described in (2.54).

If we follow arguments in Section 9, we are led to construct

$$X_T(t) = \frac{1}{\sqrt{T}}x_{j-1} + T\left(t - \frac{j-1}{T}\right)\frac{x_j - x_{j-1}}{\sqrt{T}}, \quad \left(\frac{j-1}{T} \le t \le \frac{j}{T}\right).$$

Then we have

$$S_T = \sum_{j=1}^{T} X_T\left(\frac{j-1}{T}\right)\left[X_T\left(\frac{j}{T}\right) - X_T\left(\frac{j-1}{T}\right)\right].$$

It can be shown (Problem 11.1) that

$$\int_0^1 X_T(t)dX_T(t) = S_T + \frac{1}{2T}\sum_{j=1}^{T}\left(-\frac{\beta}{T}x_{j-1} + \varepsilon_j\right)^2. \tag{3.113}$$

Here the last term on the right side converges in probability to $\frac{1}{2}$ so that, although $\mathcal{L}(X_T) \to \mathcal{L}(X)$,

$$\mathcal{L}\left(\int_0^1 X_T(t)dX_T(t)\right) \not\to \mathcal{L}\left(\int_0^1 X(t)dX(t)\right). \tag{3.114}$$

In fact it holds that

$$\begin{aligned}\mathcal{L}\left(\int_0^1 X_T(t)dX_T(t)\right) &= \mathcal{L}\left(\frac{1}{2}(X_T^2(1) - X_T^2(0))\right)\\ &\to \mathcal{L}\left(\frac{1}{2}(X^2(1) - X^2(0))\right)\\ &= \mathcal{L}\left(\int_0^1 X(t)dX(t) + \frac{1}{2}\right).\end{aligned}$$

The fact described in (3.114) is a consequence of the unbounded variation property of the Brownian motion. Note that $dX(t) = dw(t)$ when $\beta = 0$. The situation is certainly different if we consider the I(d) process with $d \geq 2$, as was discussed in Section 8 for the scalar case and in Section 10 for the vector case.

It is now easy to establish (Problem 11.2) that

$$\mathcal{L}\left(\frac{1}{T}\sum_{j=1}^{T} x_{j-1}\varepsilon_j\right) \to \mathcal{L}\left(\int_0^1 X(t)dw(t)\right). \tag{3.115}$$

We can easily extend the above result to the case where $x_j = \left(1 - \frac{\beta}{T}\right) x_{j-1} + u_j$ with $\{u_j\}$ being generated by

$$u_j = \sum_{l=0}^{\infty} \alpha_l \varepsilon_{j-l}\,, \quad \sum_{l=0}^{\infty} l|\alpha_l| < \infty\,, \quad \{\varepsilon_j\} \sim \text{i.i.d.}(0,1)\,. \tag{3.116}$$

The following theorem can be proved (Problem 11.3) by using Theorem 3.13 and the weak law of large numbers.

Theorem 3.15. *Suppose that* $x_j = \left(1 - \frac{\beta}{T}\right) x_{j-1} + u_j$, *where* $x_0 = \sqrt{T}\alpha X(0)$, $\alpha = \sum_{l=0}^{\infty}\alpha_l$ *and* $X(0) \sim N(\gamma, \delta^2)$ *with* $X(0)$ *being independent of* $\{u_j\}$ *defined by (3.116). Then we have*

$$\begin{aligned}\mathcal{L}\left(\frac{1}{T}\sum_{j=1}^{T} x_{j-1}(x_j - x_{j-1})\right) &\to \mathcal{L}\left(\frac{1}{2}\left(\alpha^2(X^2(1) - X^2(0)) - \sum_{l=0}^{\infty}\alpha_l^2\right)\right)\\ &= \mathcal{L}\left(\alpha^2 \int_0^1 X(t)dX(t) + \frac{1}{2}\left(\alpha^2 - \sum_{l=0}^{\infty}\alpha_l^2\right)\right).\end{aligned}$$

It is now an easy matter to establish (Problem 11.4) that

$$\mathcal{L}\left(\frac{1}{T}\sum_{j=1}^{T} x_{j-1}u_j\right) \longrightarrow \mathcal{L}\left(\alpha^2\int_0^1 X(t)dw(t) + \frac{1}{2}\left(\alpha^2 - \sum_{l=0}^{\infty}\alpha_l^2\right)\right). \tag{3.117}$$

Comparing with (3.115) it is seen that the effect of the stationarity assumption on $\{u_j\}$ is not much simple. This is because of the presence of the Ito integral.

We next discuss the weak convergence of the random matrix defined by

$$V_T = \frac{1}{T}\sum_{j=1}^{T} y_{j-1}(y_j - y_{j-1})', \tag{3.118}$$

where we assume, only for simplicity,

$$y_j = y_{j-1} + \varepsilon_j, \quad y_0 = 0,$$

$$\{\varepsilon_j\} \sim \text{i.i.d.}(0, I_q).$$

The present problem was solved by Chan and Wei (1988). We have

$$\mathcal{L}(V_T) \longrightarrow \mathcal{L}\left(\int_0^1 \underset{\sim}{w}(t)d\underset{\sim}{w}'(t)\right), \tag{3.119}$$

where $\{w(t)\}$ is the q-dimensional standard Brownian motion. Using this result it can be shown (Problem 11.5) that

$$\mathcal{L}\left(\frac{1}{T}\sum_{j=1}^{T} y_j' H\varepsilon_j\right) \longrightarrow \mathcal{L}\left(\int_0^1 \underset{\sim}{w}'(t)Hd\underset{\sim}{w}(t) + \text{tr}\,(H)\right) \tag{3.120}$$

for any $q \times q$ constant matrix H. A few special cases of (3.120) were discussed in Section 4 of Chapter 1, among which is Lévy's stochastic area.

The above situation was extended by Phillips (1988) to the case where

$$y_j = y_{j-1} + u_j, \qquad y_0 = 0, \tag{3.121}$$

$$u_j = \sum_{l=0}^{\infty} A_l\,\varepsilon_{j-l}, \qquad \sum_{l=0}^{\infty} l\,\|A_l\| < \infty, \tag{3.122}$$

$$A = \sum_{l=0}^{\infty} A_l, \tag{3.123}$$

with $\{\varepsilon_j\} \sim$ i.i.d.$(0, I_q)$. The following is a simplified proof of Phillips (1988) for the weak convergence of V_T in (3.118). Using the BN decomposition it can be shown (Problem 11.6) that

$$
\begin{aligned}
(3.124) \qquad V_T &= \frac{1}{T}\sum_{j=1}^{T}[Az_{j-1} + \tilde{\varepsilon}_0 - \tilde{\varepsilon}_{j-1}][A\varepsilon_j + \tilde{\varepsilon}_{j-1} - \tilde{\varepsilon}_j]' \\
&= \frac{1}{T}\left[A\sum_{j=1}^{T} z_{j-1}\varepsilon_j' A' + A\sum_{j=1}^{T}\varepsilon_j\tilde{\varepsilon}_j' - \sum_{j=1}^{T}\tilde{\varepsilon}_{j-1}u_j'\right] + o_p(1)\,,
\end{aligned}
$$

where the term $o_p(1)$ is the matrix quantity which converges in probability to the null matrix, while

$$
z_j = z_{j-1} + \varepsilon_j\,, \qquad z_0 = 0\,,
$$

$$
\tilde{\varepsilon}_j = \sum_{l=0}^{\infty}\tilde{A}_l\varepsilon_{j-l}\,, \qquad \tilde{A}_l = \sum_{k=l+1}^{\infty} A_k\,.
$$

It holds (Problem 11.7) that

$$
(3.125) \qquad \frac{1}{T}A\sum_{j=1}^{T}\varepsilon_j\tilde{\varepsilon}_j' \longrightarrow A(A - A_0)' \text{ in probability,}
$$

$$
(3.126) \qquad \frac{1}{T}\sum_{j=1}^{T}\tilde{\varepsilon}_{j-1}u_j' \longrightarrow \sum_{l=0}^{\infty}\left(\sum_{k=l+1}^{\infty} A_k\right)A_{l+1}' \text{ in probability,}
$$

and the difference of these converges in probability to

$$
\begin{aligned}
(3.127) \qquad A(A - A_0)' - \sum_{l=0}^{\infty}\left(\sum_{k=l+1}^{\infty} A_k\right)A_{l+1}' &= \sum_{l=0}^{\infty}\sum_{m=l+1}^{\infty} A_l A_m' \\
&= \sum_{l=0}^{\infty}\sum_{k=1}^{\infty} A_l A_{k+l}' \\
&= \sum_{k=1}^{\infty} E(u_0 u_k')\,.
\end{aligned}
$$

Applying (3.119) to the first term on the right side of (3.124) and using (3.127) we can establish the following theorem.

Theorem 3.16. *Suppose that the q-dimensional I(1) process $\{y_j\}$ is generated by (3.121) with (3.122) and (3.123). Then we have*

$$\mathcal{L}\left(\frac{1}{T}\sum_{j=1}^{T} y_{j-1}(y_j - y_{j-1})'\right) \longrightarrow \mathcal{L}\left(A\int_0^1 \underset{\sim}{w}(t)d\underset{\sim}{w}'(t)A' + \Lambda\right), \tag{3.128}$$

where

$$\Lambda = \sum_{k=1}^{\infty} E(u_0 u_k') = \sum_{k=1}^{\infty}\sum_{l=0}^{\infty} A_l A_{k+l}'\,.$$

It can be checked (Problem 11.8) that the present theorem reduces to Theorem 3.15 with $X(t) = w(t)$ when quantities are scalar.

Finally we indicate how to derive weak convergence results associated with statistics in the form of matrix quotients. For this purpose let us consider

$$\hat{B} = \sum_{j=2}^{T} y_j y_{j-1}' \left(\sum_{j=2}^{T} y_{j-1}y_{j-1}'\right)^{-1}, \tag{3.129}$$

where $\{y_j\}$ is the q-dimensional I(1) process defined by (3.121) together with (3.122) and (3.123). We have $T(\hat{B} - I_q) = U_T V_T^{-1}$, where

$$U_T = \frac{1}{T}\sum_{j=2}^{T} u_j y_{j-1}'\,, \qquad V_T = \frac{1}{T^2}\sum_{j=2}^{T} y_{j-1}y_{j-1}'\,.$$

Then it holds that

$$\mathcal{L}(\theta_1 U_T + \theta_2 V_T) \longrightarrow \mathcal{L}(\theta_1 h_1(\underset{\sim}{w}) + \theta_2 h_2(\underset{\sim}{w})\,)$$

for any θ_1 and θ_2, where

$$\begin{aligned} h_1(\underset{\sim}{w}) &= \left(A\int_0^1 \underset{\sim}{w}(t)d\underset{\sim}{w}'(t)A' + \Lambda\right)', \\ h_2(\underset{\sim}{w}) &= A\int_0^1 \underset{\sim}{w}(t)\underset{\sim}{w}'(t)dtA'\,. \end{aligned}$$

Therefore we can deduce that $\mathcal{L}(U_T, V_T) \longrightarrow \mathcal{L}(h_1(\underset{\sim}{w}), h_2(\underset{\sim}{w}))$. Since $P(h_2(\underset{\sim}{w}) > 0) = 1$ if A is nonsingular, we finally obtain, by the continuous mapping theorem,

$$\begin{aligned} \mathcal{L}(U_T V_T^{-1}) &= \mathcal{L}(T(\hat{B} - I_q)) \\ &\to \mathcal{L}(h_1(\underset{\sim}{w})h_2^{-1}(\underset{\sim}{w})) \\ &= \mathcal{L}\left(\left(A\int_0^1 \underset{\sim}{w}(t)d\underset{\sim}{w}'(t)A' + \Lambda\right)'\left(A\int_0^1 \underset{\sim}{w}(t)\underset{\sim}{w}'(t)dtA'\right)^{-1}\right). \end{aligned} \tag{3.130}$$

A similar procedure can be used to derive weak convergence results for other kinds of matrix-valued statistics discussed in later chapters. A scalar, simple case corresponding to (3.130) was earlier discussed in Section 3 of Chapter 1.

Problems

11.1 Establish the relation in (3.113).

11.2 Establish the weak convergence in (3.115).

11.3 Prove Theorem 3.15.

11.4 Establish the weak convergence in (3.117).

11.5 Establish the weak convergence in (3.120).

11.6 Derive the expressions for V_T in (3.124).

11.7 Prove the convergence results (3.125) and (3.126).

11.8 Show that Theorem 3.16 reduces to Theorem 3.15 with $X(t) = w(t)$ when quantities in the former are scalar.