Folded standardized time series area variance estimators for simulation

CLAUDIA ANTONINI
Departamento de Matemáticas Puras y Aplicadas,
Universidad Simón Bolívar,
Sartenejas, 1080, Venezuela

CHRISTOS ALEXOPOULOS
H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332, USA

DAVID GOLDSMAN
H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332, USA

JAMES R. WILSON
Edward P. Fitts Department of Industrial and Systems Engineering,
North Carolina State University,
Campus Box 7906,
Raleigh, NC 27695-7906, USA

---

1Member, Institute of Industrial Engineers. E-mail: <cfmantonini@usb.ve>. Telephone: +58-212-906-3294.

2Member, Institute of Industrial Engineers. E-mail: <christos@isye.gatech.edu>. Telephone: (404) 894-2361. Fax: (404) 894-2301.

3Member, Institute of Industrial Engineers, and corresponding author. E-mail: <sman@gatech.edu>. Telephone: (404) 894-2365. Fax: (404) 894-2301.

4Fellow, Institute of Industrial Engineers. E-mail: <jwilson@ncsu.edu>. Telephone: (919) 515-6415. Fax: (919) 515-5281.
Folded standardized time series area variance estimators for simulation

We estimate the variance parameter of a stationary simulation-generated process using “folded” versions of standardized time series area estimators. Asymptotically, different folding levels yield unbiased estimators that are independent scaled chi-squared variates, each with one degree of freedom. We exploit this result to formulate improved variance estimators based on the combination of multiple levels as well as the use of batching; the improved estimators preserve the asymptotic bias properties of their predecessors, but have substantially lower variance. A Monte Carlo example demonstrates the efficacy of the new methodology.

1. Introduction

One of the most important problems in simulation output analysis is the estimation of the mean \( \mu \) of a steady-state (stationary) simulation-generated process \( \{Y_i : i = 1, 2, \ldots \} \). For instance, we may be interested in determining the steady-state mean transit time in a job shop or the long-run expected profit per period arising from a certain inventory policy. Assuming that the simulation is indeed operating in steady state, the estimation of \( \mu \) is not itself a particularly difficult problem—simply use the sample mean of the observations, \( \bar{Y}_n \equiv \frac{1}{n} \sum_{i=1}^{n} Y_i \), as the point estimator.

But point estimation of the mean is usually not enough, since any serious statistical analysis should also include a measure of the variability of the sample mean. One of the most commonly used measures of this variability is the variance parameter, which is defined as the sum of covariances of the process at all lags, and which can often be written as the intuitively pleasing \( \sigma^2 = \lim_{n \to \infty} n \text{Var}(\bar{Y}_n) \). With knowledge of such a measure in hand, we could provide, among other benefits, confidence intervals for \( \mu \)—typically of the form

\[
\mu \in \bar{Y}_n \pm t\sqrt{\hat{\sigma}^2/n},
\]

where \( t \) is a quantile from the appropriate pivot distribution and \( \hat{\sigma}^2 \) is an estimator for \( \sigma^2 \).

Unfortunately, the problem of estimating the variance of the sample mean is not so straightforward. The trouble is caused by the facts that discrete-event simulation data are almost always serially correlated as well as non-normal, e.g., consecutive waiting times in a queueing system. These characteristics render as inappropriate traditional statistical analysis
methods—which may rely on the assumption of independent and identically distributed (i.i.d.) normal observations. This article is concerned with providing underlying theory for estimating the variance parameter $\sigma^2$ of a stationary simulation-generated process.

Over the years, a number of methodologies for estimating $\sigma^2$ have been proposed in the literature (see Law, 2006), e.g., the techniques referred to as nonoverlapping batch means (NBM), overlapping batch means (OBM), standardized time series (STS), spectral analysis, and regeneration. NBM—conceptually the simplest of these methodologies—divides the data $\{Y_i : i = 1, \ldots, n\}$ into nonoverlapping batches, and uses the sample variance of the sample means from the batches (i.e., the batch means) as a foundation to estimate $\sigma^2$. OBM (Meketon and Schmeiser, 1984), on the other hand, effectively re-uses the data by forming overlapping batches, and then invokes an appropriately scaled sample variance of the resulting sample means from the batches to estimate $\sigma^2$. The result is an OBM variance estimator having about the same bias as, but significantly lower variance than, the benchmark NBM estimator employing the same batch and total sample sizes. STS (Schruben, 1983) uses a functional central limit theorem to standardize a stationary time series, such as output from a steady-state discrete-event simulation, into a process that converges to limiting Brownian bridge processes as the batch or total sample sizes become large. Known properties of the Brownian bridge processes are then used to obtain estimators for $\sigma^2$. Similar to OBM, overlapping batched versions of various STS estimators have been shown to have the same bias as, but substantially lower variance than, their nonoverlapping counterparts (Alexopoulos et al., 2007b,c). Additional variance-reducing tricks in which STS re-uses data involve orthonormalizing (Foley and Goldsman, 1999) and linearly combining different STS estimators (Aktaran-Kalayci et al., 2007; Goldsman et al., 2007).

A recurring theme that emerges in the development of new estimators for $\sigma^2$ is that of the re-use of data. In the current article, we study the consequences of a “folding” operation on the original STS process (and its limiting Brownian bridge process). The folding operation produces multiple standardized time series processes, which in turn will ultimately allow us to use the original data to produce multiple estimators for $\sigma^2$—estimators that are often asymptotically independent as the sample size grows. These folded estimators will lead to combined estimators having smaller variance than existing estimators not based on the folding operation.

The article is organized as follows. Section 2 gives some background material on STS. In Section 3, we introduce the notion of folding a Brownian bridge, and we show that each application of folding yields a new Brownian bridge process. We also derive useful expressions for these folded processes in terms of the original Brownian bridge and in terms of the original underlying Brownian motion. Section 4 is concerned with derivations of the
expected values, variances, and covariances of certain functionals related to the area under a folded Brownian bridge. In Section 5, we finally show how to apply these results to the problem of estimating the variance parameter of a steady-state simulation process. The idea is to start with a single STS, form folded versions of that original STS (which converge to corresponding folded versions of a Brownian bridge process), calculate an estimator for $\sigma^2$ from each folded STS, and then combine the estimators into one low-variance estimator. We illustrate the efficacy of the folded estimators via analytical and Monte Carlo examples, and we find that the new estimators indeed reduce estimator variance at little cost in bias. Section 6 presents conclusions, while the technical details of some of the proofs are relegated to the Appendix.

2. Background

This section lays out preliminaries on the STS methodology. We begin with some standard assumptions that we shall invoke whenever needed in the sequel. In plain English, these assumptions will ensure that our upcoming variance estimators to work properly on a wide variety of stationary stochastic processes.

Assumptions A

1. The process $\{Y_i, i \geq 1\}$ is stationary and satisfies the following Functional Central Limit Theorem. For $n = 1, 2, \ldots$ and $t \in [0, 1]$, the process

   $$X_n(t) \equiv \frac{[nt](\bar{Y}_{[nt]} - \mu)}{\sigma \sqrt{n}} \quad (1)$$

   satisfies $X_n \xrightarrow{n \to \infty} W$, where: $\mu$ is the steady-state mean, $\sigma^2$ is the variance parameter, $\lfloor \cdot \rfloor$ denotes the greatest integer function; $W$ is a standard Brownian motion process on $[0, 1]$; and $\xrightarrow{n \to \infty}$ denotes weak convergence in $D[0, 1]$, the space of functions on $[0, 1]$ that are right-continuous with left-hand limits, as $n \to \infty$. See also Billingsley (1968) and Glynn and Iglehart (1990).

2. $\sum_{i=-\infty}^{\infty} R_i = \sigma^2 \in (0, \infty)$, where $R_i \equiv \text{Cov}(Y_i, Y_{i+1})$, $i = 0, 1, 2, \ldots$

3. $\sum_{i=1}^{\infty} i^2 |R_i| < \infty$.

4. The function $f(\cdot)$, defined on $[0, 1]$, is twice continuously differentiable. Further, $f(t)$ satisfies the normalizing condition $\int_0^1 \int_0^1 f(s)f(t)[\min\{s, t\} - st] \, ds \, dt = 1$. 

iie05.tex  4  March 27, 2007  –  8:55
Assumptions A.1–A.3 are mild conditions that hold for a variety of stochastic processes encountered in practice (see Glynn and Iglehart, 1990). Assumption A.4 gives conditions on the normalized weight function \( f(\cdot) \) that will be used in our estimators for \( \sigma^2 \).

Of fundamental importance to the rest of the paper is the standardized time series of the underlying stochastic process. It is the STS that will form the basis of all of the estimators studied herein.

**Definition 1.** As in Schruben (1983), the (level-0) standardized time series of the process \( \{Y_i\} \) is

\[
T_{0,n}(t) \equiv \frac{\lfloor nt \rfloor (\bar{Y}_n - \bar{Y}_{\lfloor nt \rfloor})}{\sigma \sqrt{n}}
\]  

for \( t \in [0, 1] \). \( \blacktriangle \)

In the next section, we discuss how the STS (2) is related to a Brownian bridge process; and in Section 5.2, we show how to use this process to derive estimators for \( \sigma^2 \).

### 3. Folded Brownian bridges

Our development requires some additional nomenclature. First of all, we define a Brownian bridge, which will turn out to be the limiting process of a standardized time series, much as the standard normal distribution is the limiting distribution of a properly standardized sample mean.

**Definition 2.** Suppose that \( W(\cdot) \) is a standard Brownian motion process. The associated level-0 Brownian bridge process is

\[
B_0(t) \equiv B(t) \equiv W(t) - tW(1)
\]

for \( t \in [0, 1] \). \( \blacktriangle \)

In fact, a Brownian bridge \( \{B(t) : t \in [0, 1]\} \) is a Gaussian process with \( E[B(t)] = 0 \) and \( \text{Cov}(B(s), B(t)) = \min\{s, t\} - st \), for \( s, t \in [0, 1] \). Brownian bridges are important for our purposes because under Assumptions A.1–A.3, Schruben (1983) shows that \( T_{0,n}(\cdot) \xrightarrow{n \to \infty} B(\cdot) \) and that \( \bar{Y}_n \) and \( T_{0,n}(\cdot) \) are asymptotically independent as \( n \to \infty \).

The contribution of the current paper is the development and evaluation of folded estimators for \( \sigma^2 \). We now define precisely what we mean by the folding operation, a map that can be applied either to a STS or a Brownian bridge.
Definition 3. The folding map $\Psi : Y \in D[0,1] \to \Psi_Y \in D[0,1]$ is defined by

$$\Psi_Y(t) \equiv Y(\frac{t}{2}) - Y(1 - \frac{t}{2})$$

for $t \in [0,1]$. Moreover for each nonnegative integer $k$, we define $\Psi^k : Y \in D[0,1] \to \Psi_Y^k \in D[0,1]$, the $k$-fold composition of the folding map $\Psi$, so that for every $t \in [0,1]$,

$$\Psi_Y^k(t) \equiv \begin{cases} Y(t), & \text{if } k = 0, \\ \Psi \circ \Psi_Y^{k-1}(t), & \text{if } k = 1, 2, \ldots \end{cases} <$$

The folding operation can be performed multiple times on the Brownian bridge process, as demonstrated by the following definition.

Definition 4. (see Shorack and Wellner, 1986) For $k = 1, 2, \ldots$, the level-$k$ folded Brownian bridge is

$$B_k(t) \equiv \Psi_{B_{k-1}}(t) = B_{k-1}(\frac{t}{2}) - B_{k-1}(1 - \frac{t}{2}),$$

so that $B_k(t) = \Psi_{B_0}^k(t)$ for $t \in [0,1]$. <

Intuitively speaking, when the folding operator $\Psi$ is applied to a Brownian bridge process $\{B_0(t) : t \in [0,1]\}$, it does the following: (i) $\Psi$ reflects (folds) the portion of the original process defined on the subinterval $[\frac{1}{2}, 1]$ (shown in the upper right-hand portion of Figure 1a) about the vertical line $t = \frac{1}{2}$ (yielding the subprocess shown in the upper left-hand portion of Figure 1a); and (ii) $\Psi$ takes the difference between these two subprocesses defined on $[0, \frac{1}{2}]$ and stretches that difference over the unit interval $[0, 1]$ (yielding the new process shown in Figure 1b).

Lemma 1 shows that as long as we start with a Brownian bridge, folding it will produce another Brownian bridge as well. The proof simply requires that we verify the necessary covariance structure (see Antonini, 2005).

Lemma 1. For $k = 1, 2, \ldots$, the process $\{B_k(t) : t \in [0,1]\}$ is a Brownian bridge.

The next lemma gives an equation relating the level-$k$ Brownian bridge with the original (level-0) Brownian bridge and the initial Brownian motion process. These results will be useful later on when we derive properties of certain functionals of $B_k(t)$.

Lemma 2. For $k = 1, 2, \ldots$,

$$B_k(t) = \sum_{i=1}^{2^{k-1}} [B(\frac{i}{2^{k-1}} + \frac{t}{2^k}) - B(\frac{i}{2^{k-1}} - \frac{t}{2^k})] \tag{3}$$

$$= (1 - t)W(1) + \sum_{i=1}^{2^{k-1}} [W(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}) - W(\frac{i-1}{2^{k-1}} - \frac{t}{2^k})]. \tag{4}$$
4. Some functionals of folded Brownian bridges

The purpose of this section is to highlight results on the weighted areas under successively higher levels of folded Brownian bridges. Such functionals will be used in Section 5 to construct estimators for the variance parameter $\sigma^2$ arising from a stationary stochastic process.

**Definition 5.** For $k = 0, 1, \ldots$, the weighted area under the level-$k$ folded Brownian bridge is

$$N_k(f) \equiv \int_0^1 f(t)B_k(t)\,dt.$$  \hfill $\triangle$

Under simple conditions, Theorem 4.1 shows that $N_k(f)$ has a standard normal distribution; its proof is in the Appendix.

**Theorem 4.1.** For any normalized weight function $f(t)$ and any nonnegative integer $k$, we have $N_k(f) \sim \text{Nor}(0, 1)$.

**Corollary 1.** Under the conditions of Theorem 4.1, we have $A_k(f) \equiv \sigma^2 N_k^2(f) \sim \sigma^2 \chi_1^2$.

Of course, the corollary is an immediate consequence of Theorem 4.1. Besides the distributional result, it follows that $\mathbb{E}[A_k(f)] = \sigma^2$, a finding that we will revisit in Theorem 5.3 when we develop estimators for $\sigma^2$. 

The proof of Lemma 2 is a direct consequence of Definition 4. See Antonini (2005) for the details.
Meanwhile, we proceed with several results concerning the joint distribution of the \( \{N_k(f) : k = 0, 1, \ldots\} \). Our first such result, the proof of which is in the Appendix, gives an explicit expression for the covariance between folded area functionals from different levels. Before stating the theorem, for any weight function \( f(\cdot) \), we define \( F(t) \equiv \int_0^t f(s) \, ds \), \( F \equiv F(1) \), \( \bar{F}(t) \equiv \int_0^t F(s) \, ds \), and \( \bar{F} \equiv \bar{F}(1) \).

**Theorem 4.2.** Let \( f_1(t) \) and \( f_2(t) \) be normalized weight functions. Then for \( \ell = 0, 1, \ldots \) and \( k = 1, 2, \ldots \), we have

\[
\text{Cov}[N_{\ell}(f_1), N_{\ell+k}(f_2)] = \sum_{i=1}^{2^k-1} \int_0^1 f_2(t) \left[ \bar{F}_1\left( \frac{i}{2^k} \right) - \frac{i}{2^k} \right] \, dt - \bar{F}_1 \bar{F}_2.
\] (5)

Lemmas 3–6 give results on the covariance between functionals of Brownian motion from different levels; these will be used later on to establish asymptotic covariances of estimators for \( \sigma^2 \) from different levels. In particular, Lemmas 5 and 6 give simple conditions under which these functionals are uncorrelated.

**Lemma 3.** For \( \ell, k = 0, 1, \ldots \) and \( s, t \in [0, 1] \), \( \text{Cov}[B_\ell(s), B_{\ell+k}(t)] = \text{Cov}[B_0(s), B_k(t)] \).

**Proof.** Follows by induction on \( k \). \( \square \)

**Lemma 4.** For \( \ell, k = 0, 1, \ldots \), \( \text{Cov}[N_{\ell}(f_1), N_{\ell+k}(f_2)] = \text{Cov}[N_0(f_1), N_k(f_2)] \).

**Proof.** By Lemma 3,

\[
\text{Cov}[N_{\ell}(f_1), N_{\ell+k}(f_2)] = \int_0^1 \int_0^1 f_1(s) f_2(t) \text{Cov}[B_\ell(s), B_{\ell+k}(t)] \, ds \, dt
\]

\[
= \int_0^1 \int_0^1 f_1(s) f_2(t) \text{Cov}[B_0(s), B_k(t)] \, ds \, dt. \quad \square
\]

**Lemma 5.** If the normalized weight function \( f(t) \) satisfies \( f(t) = f(1-t) \) for all \( t \in [0, 1] \), then \( \text{Cov}(N_0(f), N_k(f)) = 0 \) for all \( k = 1, 2, \ldots \).

**Proof.** Applying integration by parts to Eq. (5) with \( f_1 = f_2 = f \), we obtain

\[
\text{Cov}[N_0(f), N_k(f)] = \frac{1}{2^k} \sum_{i=1}^{2^k-1} \int_0^1 F(t) \left[ F\left( \frac{i}{2^k} \right) - \frac{i}{2^k} \right] \, dt - \bar{F}^2
\]

\[
= \frac{1}{2^k} \int_0^1 F(t) \sum_{i=1}^{2^k-1} \left[ F\left( 1 - \left( \frac{i-1}{2^k} + \frac{1}{2^k} \right) \right) + F\left( \frac{i-1}{2^k} + \frac{1}{2^k} \right) \right] \, dt - \bar{F}^2
\]

\[
= \frac{FF}{2} - \bar{F}^2,
\]
which follows since \( F(1 - x) = \int_0^{1-x} f(y) \, dy = \int_x^1 f(1 - z) \, dz = \int_x^1 f(z) \, dz \), so that \( F(1 - x) + F(x) = F \) for all \( x \in [0, 1] \). The proof is completed by noting that

\[
\bar{F} = \int_0^1 f(x)(1 - x) \, dx
\]

\[
= \int_0^{1/2} f(x)(1 - x) \, dx + \int_0^{1/2} f(1 - y)y \, dy
\]

\[
= \int_0^{1/2} f(x)(1 - x) \, dx + \int_0^{1/2} f(y)y \, dy
\]

\[
= \int_0^{1/2} f(x) \, dx = \frac{F}{2}. \quad \square
\]

**Lemma 6.** If the normalized weight function satisfies \( f(t) = f(1 - t) \) for all \( t \in [0, 1] \), then for \( \ell = 0, 1, \ldots \) and \( k = 1, 2, \ldots \), \( \text{Cov}[N_{\ell}(f), N_{\ell+k}(f)] = \text{Cov}[N_0(f), N_k(f)] = 0 \).

**Proof.** Immediate from Lemmas 4 and 5. \( \square \)

The following lemma, proven in the Appendix, establishes the multivariate normality of the random vector \( N(f) \equiv [N_0(f), N_1(f), \ldots, N_k(f)] \). It will be used in Theorem 4.3 to obtain the remarkable result that, under relatively simple conditions, the folded functionals \( \{N_k(f) : k = 0, 1, \ldots\} \) are i.i.d. \( \text{Nor}(0, 1) \).

**Lemma 7.** If the normalized weight function satisfies \( f(t) = f(1 - t) \) for all \( t \in [0, 1] \), then for each positive integer \( k \) the random vector \( N(f) \) has a nonsingular multivariate normal distribution.

**Theorem 4.3.** If the normalized weight function \( f(t) \) satisfies \( f(t) = f(1-t) \) for all \( t \in [0, 1] \), then the random variables \( \{N_k(f) : k = 0, 1, \ldots\} \) are i.i.d. \( \text{Nor}(0, 1) \) random variables.

**Proof.** Lemma 6 implies that \( \text{Cov}(N_k(f), N_j(f)) = 0 \) for every \( k \neq j \). Now, since by Lemma 7 the random vector \( N(f) \) has a multivariate normal distribution, we can conclude that the random variables \( N_0(f), N_1(f), \ldots \) are i.i.d. \( \text{Nor}(0, 1) \). \( \square \)

The next corollary, which is immediate from Theorem 4.3, will serve as the basis for our new variance estimators to be derived in Section 5.

**Corollary 2.** Under the conditions of Theorem 4.3, the random variables \( \{A_k(f) : k = 0, 1, \ldots\} \) are i.i.d. \( \sigma^2 \chi^2_1 \).
Example 1. The following weight functions arise in simulation output analysis applications (see Foley and Goldsman, 1999 and Section 5 of the current article): \( f_0(t) \equiv \sqrt{12} \), \( f_2(t) \equiv \sqrt{840(3t^2 - 3t + 1/2)} \), and \( f_{\cos,j}(t) \equiv \sqrt{8\pi j \cos(2\pi j t)} \), \( j = 1, 2, \ldots \), all for \( t \in [0, 1] \). By Theorem 4.3, \( \{N_k(f), k \geq 0\} \) are i.i.d. Nor(0, 1), and by Corollary 2, \( \{A_k(f), k \geq 0\} \) are i.i.d. \( \sigma^2 \chi_1^2 \) for \( f = f_0, f_2, \) or \( f_{\cos,j}, j = 1, 2, \ldots \). ◄

5. Application to variance estimation

We finally show how our work on properties of area functionals of folded Brownian bridges can be used in simulation output analysis. With this application in mind, we apply the folding transformation to Schruben’s level-0 STS (Schruben, 1983) in Section 5.1, thereby obtaining several new versions of the STS. These new series are used in Section 5.2 to produce new estimators for \( \sigma^2 \). Section 5.3 gives obvious methods to improve the estimators, and Section 5.4 presents a simple Monte Carlo example showing that the estimators work as intended.

5.1. Folded standardized time series

Analogous to the level-\( k \) folded Brownian bridge from Definition 4, we define the level-\( k \) folded STS.

**Definition 6.** For \( k = 1, 2, \ldots \), the level-\( k \) folded STS is

\[
T_{k,n}(t) \equiv \Psi_{T_{k-1,n}}(t) = T_{k-1,n}(\frac{t}{2}) - T_{k-1,n}(1 - \frac{t}{2})
\]

so that \( T_{k,n}(t) = \Psi_{T_{0,n}}^k(t) \) for \( t \in [0, 1] \). ◄

The next goal is to examine the convergence of the level-\( k \) folded STS to the analogous level-\( k \) folded Brownian bridge process. The following result is an immediate consequence of the almost-sure continuity of \( \Psi^k \) on \( D[0, 1] \) for \( k = 0, 1, \ldots \), and the Continuous Mapping Theorem (CMT) (Billingsley, 1968).

**Theorem 5.1.** If Assumptions A.1–A.3 hold, then for any fixed nonnegative integer \( k \), we have

\[
[T_{0,n}(\cdot), \ldots, T_{k,n}(\cdot)] \xrightarrow{n \to \infty} [B_0(\cdot), \ldots, B_k(\cdot)].
\]

Moreover, \( \sqrt{n}(\bar{Y}_n - \mu) \) and \( [T_{0,n}(\cdot), \ldots, T_{k,n}(\cdot)] \) are asymptotically independent as \( n \to \infty \).
5.2. Folded area estimators

We introduce folded versions of the STS area estimator for $\sigma^2$, along with their asymptotic distributions, expected values, and variances. To begin, we define our new estimators, along with their limiting Brownian bridge functionals.

**Definition 7.** For each nonnegative integer $k$, the STS level-$k$ folded area estimator for $\sigma^2$ is

$$A_k(f; n) \equiv N_k^2(f; n),$$

where

$$N_k(f; n) \equiv \frac{1}{n} \sum_{j=1}^{n} f\left(\frac{j}{n}\right) \sigma T_{k,n}\left(\frac{j}{n}\right)$$

and $f(\cdot)$ is a normalized weight function (satisfying Assumption A.4). The case $k = 0$, $f = f_0$ corresponds to Schruben’s original area estimator (Schruben, 1983).

**Definition 8.** Let $A(f; n) \equiv [A_0(f; n), A_1(f; n), \ldots, A_k(f; n)]$ and $A(f) \equiv [A_0(f), A_1(f), \ldots, A_k(f)]$.

The following definitions provide the necessary set-up to establish in Theorem 5.2 below the asymptotic distribution of the random vector $A(f; n)$ as $n \to \infty$.

**Definition 9.** Let $\Lambda$ denote the class of strictly increasing, continuous mappings of $[0, 1]$ onto itself such that for every $\lambda \in \Lambda$, we have $\lambda(0) = 0$ and $\lambda(1) = 1$. If $X,Y \in D[0, 1]$, then the Skorohod metric $\rho(X,Y)$ defining the “distance” between $X$ and $Y$ in $D[0, 1]$ is the infimum of those positive $\xi$ for which there exists a $\lambda \in \Lambda$ such that $\sup_{t \in [0, 1]} |\lambda(t) - t| \leq \xi$ and $\sup_{t \in [0, 1]} |X(t) - Y[\lambda(t)]| \leq \xi$. (See Billingsley, 1968 for further details.)

**Definition 10.** For each positive integer $n$, let $\Omega^n : Y \in D[0, 1] \to \Omega^n Y \in D[0, 1]$ be the approximate (discrete) STS map

$$\Omega^n_Y(t) \equiv \frac{\left\lfloor nt \right\rfloor Y(1)}{n} - Y(t)$$

for $t \in [0, 1]$. Moreover, let $\Omega : Y \in D[0, 1] \to \Omega Y \in D[0, 1]$ denote the corresponding asymptotic STS map

$$\Omega_Y(t) \equiv \lim_{n \to \infty} \Omega^n_Y(t) = tY(1) - Y(t)$$

for $t \in [0, 1]$.

Note that $\Omega^n$ maps the process (1) into the corresponding standardized time series (2) so that we have $\Omega^n\sigma_n(t) = T_{0,n}(t)$ for $t \in [0, 1]$ and $n = 1, 2, \ldots$; moreover, $\Omega$ maps a standard Brownian motion process into a standard Brownian bridge process, $\Omega W(t) = tW(1) - W(t) \sim \mathcal{B}_0(t)$ for $t \in [0, 1]$. 
Definition 11. For a given normalized weight function, for every nonnegative integer \( k \), and for every positive integer \( n \), the approximate (discrete) folded area map \( \Theta^k_n : Y \in D[0,1] \rightarrow \Theta^k_n(Y) \in \mathbb{R} \) is defined by

\[
\Theta^k_n(Y) \equiv \left[ \frac{1}{n} \sum_{i=1}^{n} f\left(\frac{i}{n}\right) \Psi^k \circ \Omega^n_Y \left(\frac{i}{n}\right) \right]^2.
\]

Moreover, the corresponding asymptotic folded area map \( \Theta^k : Y \in D[0,1] \rightarrow \Theta^k(Y) \in \mathbb{R} \) is defined by

\[
\Theta^k(Y) \equiv \left[ \int_0^1 f(t) \Psi^k \circ \Omega_Y(t) \, dt \right]^2.
\]

In terms of Eq. (1), the definition of \( A_k(f) \) from Corollary 1, and the Definitions 8–11, we see that \( \Theta^k_n(X_n) = A_k(f;n) \) and \( \Theta^k(W) = A_k(f) \) for every nonnegative integer \( k \). We are now ready to proceed with the main convergence theorem, which shows that the folded area estimators converge jointly to their asymptotic counterparts.

**Theorem 5.2.** If Assumptions A hold, then

\[
A(f,n) \xrightarrow{n \to \infty} A(f).
\] (6)

**Sketch of Proof.** Although the proof of Theorem 5.2 is detailed in the Appendix, it can be summarized as follows. Our goal is to apply the generalized CMT—that is, Theorem 5.5 of Billingsley (1968)—to prove that the \((k + 1) \times 1\) random vector with \( j \)th element \( \Theta^j_n(x_n) \) converges in distribution to the \((k + 1) \times 1\) random vector with \( j \)th element \( \Theta^j(W) \) for \( j = 0, 1, \ldots, k \). To establish the hypotheses of the generalized CMT, we show that if \( \{x_n\} \subset D[0,1] \) is any sequence of functions converging to a realization \( W \) of a standard Brownian motion process in the Skorohod metric on \( D[0,1] \), then the real-valued sequence \( \Theta^j_n(x_n) \) converges to \( \Theta^j(W) \) almost surely. First we exploit the almost-sure continuity of \( W(u) \) at every \( u \in [0,1] \) and the convergence of \( \{x_n\} \) to \( W \) in \( D[0,1] \) to show that for every nonnegative integer \( j \), with probability one we have \( |\Psi^j \circ \Omega^n_{x_n}(t) - \Psi^j \circ \Omega^n_{W}(t)| \to 0 \) uniformly for \( t \in [0,1] \) as \( n \to \infty \); and it follows that

\[
\lim_{n \to \infty} |\Theta^j_n(x_n) - \Theta^j_n(W)| = 0 \text{ with probability one.} \quad (7)
\]

Next we exploit the almost-sure convergence \( |\Psi^j \circ \Omega^n_{W}(t) - \Psi^j \circ \Omega^n_{W}(t)| \to 0 \) for all \( t \in [0,1] \) as \( n \to \infty \) together with the almost-sure continuity and Riemann integrability of \( f(t)\Psi^j \circ \Omega^n_{W}(t) \) for \( t \in [0,1] \) to show that

\[
\lim_{n \to \infty} |\Theta^j_n(W) - \Theta^j(W)| = 0 \text{ with probability one.} \quad (8)
\]
Combining (7) and (8) and applying the triangle inequality, we see that the corresponding vector-valued sequence \( \{[\Theta_n^0(x_n), \ldots, \Theta_n^k(x_n)] : n = 1, 2, \ldots \} \) converges to \([\Theta^0(W), \ldots, \Theta^k(W)]\) in \(\mathbb{R}^{k+1}\) with probability one; and thus the desired result follows directly from the generalized CMT.

\[ \square \]

**Remark 1.** Under Assumptions A, and for the weight functions in Example 1, Theorem 5.2 and Corollary 1 imply that \(A_0(f; n), \ldots, A_k(f; n)\) are asymptotically (as \(n \to \infty\)) i.i.d. \(\sigma^2 \chi_1^2\) random variables. \(\triangleright\)

Under relatively modest conditions, Theorem 5.3 gives asymptotic expressions for the expected values and variances of the level-\(k\) area estimators.

**Theorem 5.3.** Suppose that Assumptions A hold. Further, for fixed \(k \geq 0\), suppose that the family of random variables \(\{A_k^2(f; n) : n \geq 1\}\) is uniformly integrable (see Billingsley, 1968 for a definition and sufficient conditions). Then we have

\[
E[A_k(f; n)] \rightarrow_{n \to \infty} E[A_k(f)] = \sigma^2,
\]

and

\[
\text{Var}[A_k(f; n)] \rightarrow_{n \to \infty} \text{Var}[A_k(f)] = 2\sigma^4.
\]

**Remark 2.** One can obtain finer-tuned results for \(E[A_0(f; n)]\) and \(E[A_1(f; n)]\). In particular, under Assumptions A, Foley and Goldsman (1999) and Goldsman et al. (1990) show that

\[
E[A_0(f; n)] = \sigma^2 + \frac{[(F - \bar{F})^2 + \bar{F}^2] \gamma}{2n} + o(1/n).
\]

In a companion paper, Alexopoulos et al. (2007a), we find that if Assumptions A hold and \(n\) is even, then

\[
E[A_1(f; n)] = \sigma^2 + \frac{\bar{F}^2 \gamma}{n} + o(1/n).
\]

### 5.3. Enhanced estimators

The individual estimators whose properties are given in Theorem 5.3 are all based on a single long run of \(n\) observations, and all involve a single level \(k\) of folding. This section discusses some obvious extensions of the estimators that have improved asymptotic properties—batching and combining levels.

**Batching:** In actual applications, we often organize the data by breaking the \(n\) observations into \(b\) contiguous, nonoverlapping batches, each of size \(m\) so that \(n = bm\); and then we
can compute the folded variance estimators from each batch separately. As the batch size $m \to \infty$, the variance estimators computed from different batches are asymptotically independent under broadly applicable conditions on the original (unbatched) process $\{Y_i, i \geq 1\}$; and thus more stable (i.e., more accurate) variance estimators can be obtained by combining the folded variance estimators computed from all available batches.

In view of this motivation, suppose that the $i$th batch of size $m$ consists of the observations $Y_{(i-1)m+1}, Y_{(i-1)m+2}, \ldots, Y_{im}$, for $i = 1, 2, \ldots, b$. Using the obvious minor changes to the appropriate definitions, one can construct the level-$k$ STS from the $i$th batch of observations, say $T_{k,m,i}(t)$; and from there, one can obtain the resulting level-$k$ area estimator from the $i$th batch, say $A_{k,i}(f; m)$. Finally, we define the level-$k$ batched folded area estimator for $\sigma^2$ by

$$\tilde{A}_k(f; b, m) \equiv \frac{1}{b} \sum_{i=1}^{b} A_{k,i}(f; m).$$

Under the conditions of Theorem 5.3, we have $\lim_{m \to \infty} \mathbb{E}[\tilde{A}_k(f; b, m)] = \sigma^2$ and $\lim_{m \to \infty} \text{Var}[\tilde{A}_k(f; b, m)] = 2\sigma^4/b$, where the latter result follows from the fact that the $A_{k,i}(f; m)$’s, $i = 1, 2, \ldots, b$, are asymptotically independent as $m \to \infty$. Thus, we obtain batched estimators with approximately the same expected value $\sigma^2$ as a single folded estimator arising from one long run, yet with substantially smaller variance.

**Combining levels of folding:** Theorem 5.3 shows that, for a particular weight function $f(\cdot)$, all of the estimators from different levels of folding behave about the same asymptotically in terms of their expected value and variance. We can improve upon these individual estimators by combining the different levels. To this end, denote the average of the folded area estimators from levels 0, 1, \ldots, $k$ by

$$\bar{A}_k(f; n) \equiv \frac{1}{k+1} \sum_{j=0}^{k} A_j(f; n).$$

Under the conditions of Remark 1 and Theorem 5.3, we have $\lim_{n \to \infty} \mathbb{E}[\bar{A}_k(f; n)] = \sigma^2$, and $\lim_{n \to \infty} \text{Var}[\bar{A}_k(f; n)] = 2\sigma^4/(k+1)$. Thus, we obtain combined estimators with approximately the same expected value $\sigma^2$ as a single folded estimator arising from one level, yet with significantly smaller variance.

### 5.4. Example

We illustrate the performance of the new folded estimator with a simple Monte Carlo experiment involving a stationary first-order autoregressive [AR(1)] process. An AR(1) process is constructed by setting $Y_i = \phi Y_{i-1} + \epsilon_i, i = 1, 2, \ldots$, where the $\epsilon_i$’s are i.i.d. $\text{Nor}(0,1-\phi^2)$, $Y_0$
is a standard normal random variable that is independent of the \(\epsilon_i\)'s, and \(-1 < \phi < 1\) (to preserve stationarity). It is well known that, for the AR(1) process, \(R_k = \phi^k, k = 0, 1, 2 \ldots\), and \(\sigma^2 = (1 + \phi)/(1 - \phi)\).

In the current example,

- We set the parameter \(\phi = 0.9\) (so that \(\sigma^2 = 19\)).
- We used \(b = 32\) batches of observations, various batch sizes \(m\), and weight function \(f_0(\cdot)\).
- We carried out 4096 independent realizations of the level-0 and level-1 batched folded area estimators, denoted by \(\tilde{A}_0(f_0;32,m)\) and \(\tilde{A}_1(f_0;32,m)\).
- We calculated the estimated mean and variance of the estimators, averaged over the 4096 realizations.

The idea is to demonstrate the convergence of the expected values of the estimators to the variance parameter \(\sigma^2\) as the batch size \(m\) increases. The details are given in columns 2–5 of Table 1. In the table display, a hat denotes the estimate of the analogous performance metric; in addition, we suppress the estimator notation \("(f_0;32,m)"\) for succinctness.

We also checked the performance of the batched estimators when we combine levels. In particular, we used the realizations from the individual levels 0 and 1 to calculate realizations of

\[
A_1(f_0;32,m) \equiv \frac{1}{2}[\tilde{A}_0(f_0;32,m) + \tilde{A}_1(f_0;32,m)],
\]

and we obtained the estimated performance characteristics for this combined estimator, shown in the last two columns of Table 1.

We see that the estimators \(\tilde{A}_0(f_0;32,m)\) and \(\tilde{A}_1(f_0;32,m)\) perform as anticipated by the theory—for large batch size \(m\), the estimated expected values are nearly \(\sigma^2\) (\(\approx 19\) in this case) and the variances are reasonably close to \(2\sigma^4/b\) (\(\approx 22.563\) for the AR(1) process we chose). We also observe that reasonable expected values are maintained for the combined estimator \(A_1(f_0;32,m)\) as \(m\) becomes large, with the bonus of substantially (\(\approx 50\%\)) reduced variance.

### 6. Conclusions

The main purpose of this article was to introduce “folded” versions of the standardized time series area estimator for the variance parameter arising from a stationary simulation process.
Table 1. Empirical performance of the enhanced folded area estimators

<table>
<thead>
<tr>
<th>m</th>
<th>$\bar{E}[\tilde{A}_0]$</th>
<th>$\bar{Var}[\tilde{A}_0]$</th>
<th>$\bar{E}[\tilde{A}_1]$</th>
<th>$\bar{Var}[\tilde{A}_1]$</th>
<th>$\bar{E}[A_1]$</th>
<th>$\bar{Var}[A_1]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>15.02</td>
<td>14.53</td>
<td>13.43</td>
<td>10.96</td>
<td>14.23</td>
<td>6.37</td>
</tr>
<tr>
<td>256</td>
<td>16.91</td>
<td>18.16</td>
<td>16.52</td>
<td>16.99</td>
<td>16.71</td>
<td>8.75</td>
</tr>
<tr>
<td>512</td>
<td>17.89</td>
<td>21.05</td>
<td>17.73</td>
<td>19.62</td>
<td>17.81</td>
<td>10.41</td>
</tr>
<tr>
<td>1024</td>
<td>18.50</td>
<td>21.52</td>
<td>18.34</td>
<td>21.38</td>
<td>18.42</td>
<td>10.80</td>
</tr>
<tr>
<td>2048</td>
<td>18.82</td>
<td>22.99</td>
<td>18.66</td>
<td>21.87</td>
<td>18.74</td>
<td>11.02</td>
</tr>
<tr>
<td>4096</td>
<td>18.78</td>
<td>23.33</td>
<td>18.85</td>
<td>22.61</td>
<td>18.82</td>
<td>11.40</td>
</tr>
<tr>
<td>8192</td>
<td>18.98</td>
<td>22.68</td>
<td>18.82</td>
<td>22.57</td>
<td>18.90</td>
<td>11.21</td>
</tr>
<tr>
<td>16384</td>
<td>18.90</td>
<td>22.24</td>
<td>18.98</td>
<td>22.84</td>
<td>18.94</td>
<td>11.32</td>
</tr>
</tbody>
</table>

We provided theoretical results showing that the folded estimators converge to appropriate functionals of Brownian motion; and these convergence results allow us to produce asymptotically unbiased and low-variance estimators using multiple folding levels in conjunction with standard batching techniques. At each folding level, and for each weight function in Example 1, the proposed estimators can be computed in $O(n)$ time; the detailed computations will be listed in a forthcoming article.

Ongoing work includes the following. As in Remark 2, we can derive precise expressions for the expected values of the folded estimators—expressions that show just how quickly any estimator bias dies off as the batch size increases. We can also produce analogous folding results for other “primitive” STS variance estimators, e.g., for Cramér–von Mises estimators, as described in Antonini (2005). In addition, whatever type of primitive estimator we choose to use, there is interest in finding the best ways to combine batching and multiple folding levels in order to produce even-better estimators for $\sigma^2$, and subsequently, good confidence intervals for the underlying steady-state mean $\mu$. In any case, we have also planned a massive Monte Carlo analysis to examine estimator performance over a variety of benchmark processes. Future work includes the development of overlapping versions of the folded estimators, as in Alexopoulos et al. (2007b,c) and in Meketon and Schmeiser (1984).

Acknowledgements

Partial support for our research was provided by National Science Foundation Grant DMI-0400260.
References


Appendix

Proof of Theorem 4.1

Since \( N_k(f) \) is the integral of a continuous function over the closed interval \([0, 1]\), its Riemann sum satisfies (see p. 229 of Bartle, 1976)

\[
N_{k,m}(f) \equiv \frac{1}{m} \sum_{i=1}^{m} f\left(\frac{i}{m}\right) B_k\left(\frac{i}{m}\right) \xrightarrow{a.s.} \lim_{m \to \infty} N_k(f), \quad (A1)
\]

where “\( \xrightarrow{a.s.} \) \( m \to \infty \)” denotes almost-sure convergence as \( m \to \infty \). For fixed \( k \) and \( m \), \( N_{k,m}(f) \) is normal since it is a finite linear combination of jointly normal random variables. Furthermore, \( N_{k,m}(f) \) has expected value 0 and variance

\[
\text{Var}[N_{k,m}(f)] = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} f\left(\frac{i}{m}\right) f\left(\frac{j}{m}\right) \text{Cov} [B_k\left(\frac{i}{m}\right), B_k\left(\frac{j}{m}\right)]
\]

\[
= \int_0^1 \int_0^1 f(s)f(t) \left[ \min\{s,t\} - st \right] ds\, dt + O(1/m)
\]

\[
= 1 + O(1/m).
\]
Eq. (A1) implies that the characteristic function of $N_{k,m}(f)$ converges to the characteristic function of $N_k(f)$ as $m \to \infty$ (see p. 172 of Grimmett and Stirzaker, 1992). Since $N_{k,m}(f)$ is normal with mean 0 and variance $1 + O(1/m)$, its characteristic function is given by

$$\varphi_m(t) = \mathbb{E}\{\exp[\sqrt{-1} t N_{k,m}(f)]\} = \exp\left\{-\frac{t^2}{2} [1 + O(1/m)]\right\}.$$  

It follows immediately that $\lim_{m \to \infty} \varphi_m(t) = \exp(-t^2/2)$, the characteristic function of the standard normal distribution; and thus $N_k(f) \sim \text{Nor}(0, 1)$.  

**Proof of Theorem 4.2**

By Lemma 4 and Eq. (3),

$$\text{Cov}[N_t(f_1), N_{t+k}(f_2)] = \text{Cov}[N_0(f_1), N_k(f_2)]$$

$$= \int_0^1 \int_0^1 f_1(s) f_2(t) \text{Cov}[\mathcal{B}_0(s), \mathcal{B}_k(t)] \, ds \, dt$$

$$= \sum_{i=1}^{2^{k-1}} \int_0^1 \int_0^1 f_1(s) f_2(t) \text{Cov}[\mathcal{B}(s), \mathcal{B}\left(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}\right) - \mathcal{B}\left(\frac{i}{2^{k-1}} - \frac{t}{2^k}\right)] \, ds \, dt$$

$$= \sum_{i=1}^{2^{k-1}} \int_0^1 \int_0^1 f_1(s) f_2(t) \left[ \min\{s, \frac{i-1}{2^{k-1}} + \frac{t}{2^k}\} - \min\{s, \frac{i}{2^{k-1}} - \frac{t}{2^k}\} + \frac{s(1-t)}{2^{k-1}} \right] \, ds \, dt$$

$$= \sum_{i=1}^{2^{k-1}} \int_0^1 f_2(t) \left[ \int_0^{\frac{i}{2^{k-1}} + \frac{t}{2^k}} f_1(s) \, ds + \int_{\frac{i}{2^{k-1}} + \frac{t}{2^k}}^{1} f_1(s) \left(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}\right) \, ds \right. $$

$$- \int_0^{\frac{i}{2^{k-1}} - \frac{t}{2^k}} f_1(s) \, ds - \int_{\frac{i}{2^{k-1}} - \frac{t}{2^k}}^{1} f_1(s) \left(\frac{i-1}{2^{k-1}} - \frac{t}{2^k}\right) \, ds \right] \, dt + (F_1 - F_1) \bar{F}_2$$

$$= \sum_{i=1}^{2^{k-1}} \int_0^1 f_2(t) \left[ - \int_{\frac{i}{2^{k-1}} + \frac{t}{2^k}}^{\frac{i-1}{2^{k-1}} + \frac{t}{2^k}} f_1(s) \, ds + \left(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}\right) \int_{\frac{i}{2^{k-1}} + \frac{t}{2^k}}^{1} f_1(s) \, ds \right. $$

$$- \left(\frac{i}{2^{k-1}} - \frac{t}{2^k}\right) \int_{\frac{i}{2^{k-1}} - \frac{t}{2^k}}^{\frac{i-1}{2^{k-1}} + \frac{t}{2^k}} f_1(s) \, ds \right] \, dt + (F_1 - F_1) \bar{F}_2$$

$$= \sum_{i=1}^{2^{k-1}} \int_0^1 f_2(t) \left[ - \left(\frac{i}{2^{k-1}} - \frac{t}{2^k}\right) F_1\left(\frac{i-1}{2^{k-1}} - \frac{t}{2^k}\right) + \left(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}\right) F_1\left(\frac{i}{2^{k-1}} + \frac{t}{2^k}\right) \right.$$

$$+ F_1\left(\frac{i}{2^{k-1}} - \frac{t}{2^k}\right) - F_1\left(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}\right) + \left(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}\right) F_1(1) - F_1\left(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}\right) \right] \, dt + (F_1 - F_1) \bar{F}_2$$

(since $\int_0^b s f_1(s) \, ds = b F_1(b) - a F_1(a) - F_1(b) + F_1(a)$)

$$= \sum_{i=1}^{2^{k-1}} \int_0^1 f_2(t) \left\{ \bar{F}_1\left(\frac{i}{2^{k-1}} - \frac{t}{2^k}\right) - \bar{F}_1\left(\frac{i-1}{2^{k-1}} + \frac{t}{2^k}\right) - \frac{(1-t) F_1}{2^{k-1}} \right\} \, dt + (F_1 - F_1) \bar{F}_2,$$

from which the result follows.  

$\square$
Proof of Lemma 7

First, we show that every linear combination $\sum_{j=0}^{k} a_j N_j(f)$ has a normal distribution, and hence, $N(f)$ has a multivariate normal distribution by virtue of Theorem 2.6.2 of Anderson (1984). Indeed,

$$\sum_{j=0}^{k} a_j N_j(f) = \int_{0}^{1} \left[ \sum_{j=0}^{k} a_j f(t) B_j(t) \right] dt.$$ 

Further, by Definition 2 and Eq. (4), for each $B_j(t)$,

$$Z(t) = \sum_{j=0}^{k} a_j f(t) B_j(t)$$

$$= a_0 f(t) (W(t) - tW(1)) + \sum_{j=1}^{k} a_j \sum_{i=1}^{2^{j-1}} f(t) \left[ W(i - \frac{1}{2^{j-1}} + \frac{t}{2^j}) - W(i - \frac{1}{2^{j-1}} - \frac{t}{2^j}) \right]$$

$$+ \left( \sum_{j=1}^{k} a_j \right) (1 - t) f(t) W(1).$$

Now, let $c_1, c_2, \ldots, c_m$ be real constants and $0 \leq t_1 < \cdots < t_m \leq 1$. Then

$$\sum_{\ell=1}^{m} c_\ell Z(t_\ell) = \sum_{\ell=1}^{m} c_\ell a_0 f(t_\ell) (W(t_\ell) - t_\ell W(1))$$

$$+ \sum_{\ell=1}^{m} c_\ell \sum_{j=1}^{k} a_j \sum_{i=1}^{2^{j-1}} f(t_\ell) \left[ W(i - \frac{1}{2^{j-1}} + \frac{t_\ell}{2^j}) - W(i - \frac{1}{2^{j-1}} - \frac{t_\ell}{2^j}) \right]$$

$$+ \sum_{\ell=1}^{m} c_\ell \left( \sum_{j=1}^{k} a_j \right) (1 - t_\ell) f(t_\ell) W(1).$$

Let $\mathcal{T}$ be the set of all times of the form $\frac{i-1}{2^{j-1}} + \frac{t_\ell}{2^j}$ or $\frac{i-1}{2^{j-1}} - \frac{t_\ell}{2^j}$, for some $\ell = 1, \ldots, m$, $j = 1, \ldots, k$, and $i = 1, \ldots, 2^{j-1}$. Let $\{\tau_1, \ldots, \tau_L\}$ be an increasing ordering of $\mathcal{T} \cup \{1\}$. Clearly, we can write $\sum_{\ell=1}^{m} c_\ell Z(t_\ell)$ as $\sum_{\ell=1}^{L} d_\ell W(\tau_\ell)$, for some real constants $d_1, \ldots, d_L$ since the function $f(\cdot)$ is deterministic. Since $W$ is a Gaussian process, the latter summation is Gaussian and thus, $Z$ is a Gaussian process. Notice also that $Z$ has continuous paths because $W$ has continuous paths. Finally recall that $\sum_{j=0}^{k} a_j N_j(f) = \int_{0}^{1} Z(t) dt$; and the same methodology used in Theorem 4.1 can be used to show that the latter integral is a normal random variable.

To prove that $N(f) = [N_0(f), \ldots, N_k(f)]$ has a nonsingular multivariate normal distribution, we show that the variance-covariance matrix $\Sigma_{N(f)}$ is positive definite. This follows
immediately from Lemma 6 since
\[ a \Sigma_N(f) a^T = \text{Var} \left[ \sum_{j=0}^{k} a_j N_j(f) \right] = \sum_{j=0}^{k} a_j^2 > 0, \]
for all \( a = (a_0, \ldots, a_k) \in \mathbb{R}^{k+1} - \{0\}. \)

\begin{proof}[Proof of Theorem 5.2]
In terms of the definition (1) and the Definitions 8–11, we see that
\[ \Theta^k_n(X_n) = A_k(f; n) \quad \text{for} \quad k = 0, 1, \ldots; \]
and we seek to apply the generalized CMT—that is, Theorem 5.5 of Billingsley (1968)—to prove that the \((k+1) \times 1\) random vector with \(j\)th element \(\Theta^j_n(X_n), \ j = 0, 1, \ldots, k,\) converges in distribution to the \((k+1) \times 1\) random vector with \(j\)th element \(\Theta^j(W).\) To verify the hypotheses of the generalized CMT, we establish the following result. In terms of the set of discontinuities
\[ D_j \equiv \{ x \in D[0, 1] : \text{for some sequence} \ \{x_n\} \subset D[0, 1] \text{converging to} \ x, \]
the sequence \(\{\Theta^j_n(x_n)\} \text{does not converge to} \ \Theta^j(x) \} \quad \text{(A2)}\]
for \(j = 0, \ldots, k,\) we will show that
\[ \Pr\left\{ W(\cdot) \in D[0, 1] - \bigcup_{j=0}^{k} D_j \right\} = 1. \quad \text{(A3)} \]

To prove (A3), we will exploit the almost-sure continuity of sample paths of \(W(\cdot)\):

With probability 1, the function \(W(t)\) is continuous at every \(t \geq 0; \quad \text{(A4)}\)
see §41.3.A of Loève (1978) or p. 64 of Billingsley (1968). Thus we may assume without loss of generality that we are restricting our attention to an event \(\mathcal{H} \subset D[0, 1]\) for which (A4) holds so that
\[ \Pr\{W \in \mathcal{H}\} = 1. \quad \text{(A5)} \]
Suppose \(\{x_n\} \subset D[0, 1]\) converges to \(W \in \mathcal{H}\) and that \(j \in \{0, 1, \ldots, k\}\) is a fixed integer. Next we seek to prove the key intermediate result,

For each \(W \in \mathcal{H}\) and for each sequence \(\{x_n\} \subset D[0, 1]\) converging to \(W,\) we have
\[ \lim_{n \to \infty} \left| \Psi^j \circ \Omega^n_{x_n}(t) - \Psi^j \circ \Omega^n_{W}(t) \right| = 0 \text{ uniformly for} \ t \in [0, 1]. \quad \text{(A6)} \]
We prove (A6) by induction on \( j \), starting with \( j = 0 \). Choose \( \varepsilon > 0 \) arbitrarily. Throughout the following discussion, \( \mathcal{W} \in \mathcal{H} \) and \( \{x_n\} \) are fixed; and thus virtually all the quantities introduced in the rest of the proof depend on \( \mathcal{W} \) and \( \{x_n\} \). The sample-path continuity property (A4) and Theorem 4.47 of Apostol (1974) imply that \( \mathcal{W}(t) \) is uniformly continuous on \([0,1] \); and thus we can find \( \zeta > 0 \) such that

For all \( t, t' \in [0,1] \) with \( |t - t'| < \zeta \), we have \( |\mathcal{W}(t) - \mathcal{W}(t')| < \varepsilon/4 \). \hfill (A7)

Because \( \{x_n\} \) converges to \( \mathcal{W} \) in \( D[0,1] \), there is a sufficiently large integer \( N \) such that for each \( n \geq N \), there exists \( \lambda_n(\cdot) \in \Lambda \) satisfying

\[
\sup_{t \in [0,1]} |\lambda_n(t) - t| < \min\{\zeta, \varepsilon/4\} \hfill (A8)
\]

and

\[
\sup_{t \in [0,1]} |x_n(t) - \mathcal{W}[\lambda_n(t)]| < \min\{\zeta, \varepsilon/4\}. \hfill (A9)
\]

When \( j = 0 \), the map \( \Psi^j \) is the identity; and in this case for each \( n \geq N \) we have

\[
\left| \Psi^j \circ \Omega^n_{x_n}(t) - \Omega^n_{\mathcal{W}}(t) \right| = \left| \Omega^n_{x_n}(t) - \Omega^n_{\mathcal{W}}(t) \right|
\]

\[
= \left| \left[ \frac{nt}{n} x_n(1) - x_n(t) \right] - \left[ \frac{nt}{n} \mathcal{W}(1) - \mathcal{W}(t) \right] \right|
\]

\[
\leq \frac{nt}{n} |x_n(1) - \mathcal{W}(1)| + |x_n(t) - \mathcal{W}(t)| \hfill (A10)
\]

\[
\leq |x_n(1) - \mathcal{W}[\lambda_n(1)]| + |\mathcal{W}[\lambda_n(1)] - \mathcal{W}(1)| + |x_n(t) - \mathcal{W}[\lambda_n(t)]|
\]

\[
+ |\mathcal{W}[\lambda_n(t)] - \mathcal{W}(t)| \hfill (A11)
\]

\[
\leq \varepsilon/4 + \varepsilon/4 + \varepsilon/4 + \varepsilon/4 = \varepsilon \quad \text{for} \quad t \in [0,1], \hfill (A12)
\]

where (A10) and (A11) follow from the triangle inequality and (A12) follows from (A7), (A8) and (A9). This establishes (A6) for \( j = 0 \).

Now suppose that (A6) holds for some \( j \geq 0 \). Again we choose \( \varepsilon > 0 \) arbitrarily. The induction hypothesis ensures that there exists \( N' \) sufficiently large such that for each \( n \geq N' \), we have

\[
\left| \Psi^j \circ \Omega^n_{x_n}(t) - \Psi^j \circ \Omega^n_{\mathcal{W}}(t) \right| < \varepsilon/2 \quad \text{for all} \quad t \in [0,1]. \hfill (A13)
\]

We have

\[
\left| \Psi^{j+1} \circ \Omega^n_{x_n}(t) - \Psi^{j+1} \circ \Omega^n_{\mathcal{W}}(t) \right|
\]

\[
= \left| \left[ \Psi^j \circ \Omega^n_{x_n}(\frac{t}{2}) - \Psi^j \circ \Omega^n_{x_n}(1 - \frac{t}{2}) \right] - \left[ \Psi^j \circ \Omega^n_{\mathcal{W}}(\frac{t}{2}) - \Psi^j \circ \Omega^n_{\mathcal{W}}(1 - \frac{t}{2}) \right] \right| \hfill (A14)
\]

\[
\leq \left| \Psi^j \circ \Omega^n_{x_n}(\frac{t}{2}) - \Psi^j \circ \Omega^n_{\mathcal{W}}(\frac{t}{2}) \right| + \left| \Psi^j \circ \Omega^n_{x_n}(1 - \frac{t}{2}) - \Psi^j \circ \Omega^n_{\mathcal{W}}(1 - \frac{t}{2}) \right| \hfill (A15)
\]

\[
< \varepsilon/2 + \varepsilon/2 = \varepsilon \quad \text{for} \quad t \in [0,1] \text{ and } n \geq N', \hfill (A16)
\]
where: (A14) follows from Definition 3 of the folding map; (A15) follows from the triangle inequality; and (A16) follows from (A13). This establishes (A6) for $j = 0, 1, \ldots$.

Since $f(\cdot)$ is continuous on $[0, 1]$ by Assumption A.4, we have

$$f^* \equiv \max_{t \in [0, 1]} |f(t)| < \infty;$$

thus (A5), (A6), and (A17) imply that

$$\lim_{n \to \infty} |\Theta_n^j(x_n) - \Theta_n^j(W)| \leq f^* \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} |\Psi^j \circ \Omega_{x_n}^n(\frac{i}{n}) - \Psi^j \circ \Omega_{W}^n(\frac{i}{n})| = 0 \text{ with probability 1.}$$

(A18)

An argument similar to that justifying (A6) proves that

With probability 1, $\lim_{n \to \infty} |\Psi^j \circ \Omega_{W}^n(t) - \Psi^j \circ \Omega_{W}^n(t)| = 0.$

(A19)

In view of the almost-sure continuity of sample paths of $W(\cdot)$ and the continuity of $f(\cdot)$, it is straightforward to show that

With probability 1, the function $f(t)\Psi^j \circ \Omega_W(t)$ is continuous at every $t \in [0, 1]$; (A20)

and thus it follows that $f(t)\Psi^j \circ \Omega_W(t)$ is Riemann integrable with probability 1 and that

$$\lim_{n \to \infty} |\Theta_n^j(W) - \Theta^j(W)| = \lim_{n \to \infty} \left| \frac{1}{n} \sum_{i=1}^{n} f(\frac{i}{n}) \Psi^j \circ \Omega_W^n(\frac{i}{n}) - \int_{0}^{1} f(t)\Psi^j \circ \Omega_W(t) \, dt \right| = 0 \text{ with probability 1.}$$

(A21)

Combining (A18) and (A21) and applying the triangle inequality, we see that for each $j \in \{0, 1, \ldots, k\}$,

$$\lim_{n \to \infty} |\Theta_n^j(x_n) - \Theta^j(W)| \leq \lim_{n \to \infty} |\Theta_n^j(x_n) - \Theta_n^j(W)| + \lim_{n \to \infty} |\Theta_n^j(W) - \Theta^j(W)| = 0 \text{ with probability 1.}$$

(A22)

It follows that the corresponding vector-valued sequence $\{[\Theta_n^0(x_n), \ldots, \Theta_n^k(x_n)] : n = 1, 2, \ldots\}$ converges to $[\Theta^0(W), \ldots, \Theta^k(W)]$ in $\mathbb{R}^{k+1}$ with probability one; and thus the desired result (6) follows directly from the generalized CMT. \qed