Nonlinear system theory: Another look at dependence

Wei Biao Wu†

Department of Statistics, University of Chicago, 5734 South University Avenue, Chicago, IL 60637

Communicated by Murray Rosenblatt, University of California at San Diego, La Jolla, CA, August 4, 2005 (received for review April 29, 2005)

Based on the nonlinear system theory, we introduce previously undescribed dependence measures for stationary causal processes. Our physical and predictive dependence measures quantify the degree of dependence of outputs on inputs in physical systems. The proposed dependence measures provide a natural framework for a limit theory for stationary processes. In particular, under conditions with quite simple forms, we present limit theorems for partial sums, empirical processes, and kernel density estimates. The conditions are mild and easily verifiable because they are directly related to the data-generating mechanisms.

nonlinear time series | limit theory | kernel estimation | weak convergence

Let ε_i, $i \in \mathbb{Z}$, be independent and identically distributed (iid) random variables and g be a measurable function such that

$$X_i = g(\ldots, \varepsilon_{i-1}, \varepsilon_i), \qquad [1]$$

is a properly defined random variable. Then (X_i) is a stationary process, and it is causal or nonanticipative in the sense that X_i does not depend on the future innovations ε_j, $j > i$. The causality assumption is quite reasonable in the study of time series. Wiener (1) considered the fundamental coding and decoding problem of representing stationary and ergodic processes in terms of the form Eq. **1**. In particular, Wiener studied the construction of ε_i based on X_k, $k \le i$. The class of processes that Eq. **1** represents is huge and it includes linear processes, Volterra processes, and many time series models. In certain situations, Eq. **1** is also called the nonlinear Wold representation. See refs. 2–4 for other deep contributions of representing stationary and ergodic processes by Eq. **1**. To conduct statistical inference of such processes, it is necessary to consider the asymptotic properties of the partial sum $S_n = \sum_{i=1}^n X_i$ and the empirical distribution function $F_n(x) = n^{-1} \sum_{i=1}^n \mathbf{1}_{X_i \le x}$.

In probability theory, many limit theorems have been established for independent random variables. Those limit theorems play an important role in the related statistical inference. In the study of stochastic processes, however, independence usually does not hold, and the dependence is an intrinsic feature. In an influential paper, Rosenblatt (5) introduced the strong mixing condition. For a stationary process (X_i), let the sigma algebra $\mathcal{A}_m^n = \sigma(X_m, \ldots, X_n)$, $m \le n$, and define the strong mixing coefficients

$$\alpha_n = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{A}_{-\infty}^0, B \in \mathcal{A}_n^\infty\}. \qquad [2]$$

If $\alpha_n \to 0$, then we say that (X_i) is strong mixing. Variants of the strong mixing condition include ρ, ψ, and β-mixing conditions among others (6). A central limit theorem (CLT) based on the strong mixing condition is proved in ref. 5. Since then, as basic assumptions on the dependence structures, the strong mixing condition and its variants have been widely used and various limit theorems have been obtained; see the extensive treatment in ref. 6.

Since the quantity $|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$ in Eq. **2** measures the dependence between events A and B and it is zero if A and B are independent, it is sensible to call α_n and its variants "probabilistic dependence measures." For stationary causal processes, the calculation of probabilistic dependence measures is generally not easy because it involves the complicated manipulation of taking the supremum over two sigma algebras (7–9). Additionally, many well-known processes are not strong mixing. A prominent example is the Bernoulli shift process. Consider the simple AR(1) process $X_n = (X_{n-1} + \varepsilon_n)/2$, where ε_i are iid Bernoulli random variables with success probability 1/2 (see refs. 10 and 11). Then X_n is a causal process with the representation $X_n = \sum_{i=0}^\infty 2^{-i}\varepsilon_{n-i}$ and the innovations ε_n, ε_{n-1}, ..., correspond to the dyadic expansion of X_n. The process X_n is not strong mixing since $\alpha_n \equiv 1/4$ for all n (12). Some alternative ways have been proposed to overcome the disadvantages of strong mixing conditions (8, 9).

Dependence Measures

In this work, we shall provide another look at the fundamental issue of dependence. Our primary goal is to introduce "physical or functional" and "predictive dependence measures" a previously undescribed type of dependence measures that are quite different from strong mixing conditions. In particular, following refs. 1 and 13, we shall interpret Eq. **1** as an input/output system and then introduce dependence coefficients by measuring the degree of dependence of outputs on inputs. Specifically, we view Eq. **1** as a physical system

$$x_i = g(\ldots, e_{i-1}, e_i), \qquad [3]$$

where $e_i, e_{i-1}, \ldots$ are inputs, g is a filter or a transform, and x_i is the output. Then, the process X_i is the output of the physical system **3** with random inputs. It is clearly not a good way to assess the dependence just by taking the partial derivatives $\partial g/\partial e_j$, which may not exist if g is not well-behaved. Nonetheless, because the inputs are random and iid, the dependence of the output on the inputs can be simply measured by applying the idea of coupling. Let (ε'_i) by an iid copy of (ε_i); let the shift process $\xi_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$, $\xi'_i = (\ldots, \varepsilon'_{i-1}, \varepsilon'_i)$. For a set $I \subset \mathbb{Z}$, let $\varepsilon_{j,I} = \varepsilon'_j$ if $j \in I$ and $\varepsilon_{j,I} = \varepsilon_j$ if $j \notin I$; let $\xi_{i,I} = (\ldots, \varepsilon_{i-1,I}, \varepsilon_{i,I})$ and $\xi_i^* = \xi_{i,\{0\}}$. Then $\xi_{i,I}$ is a coupled version of ξ_i with ε_j replaced by ε'_j if $j \in I$. For $p > 0$ write $X \in \mathcal{L}^p$ if $\|X\|_p := [\mathbb{E}(|X|^p)]^{1/p} < \infty$ and $\|X\| = \|X\|_2$.

Definition 1 (Functional or physical dependence measure): For $p > 0$ and $I \subset \mathbb{Z}$ let $\delta_p(I, n) = \|g(\xi_n) - g(\xi_{n,I})\|_p$ and $\delta_p(n) = \|g(\xi_n) - g(\xi_n^*)\|_p$. Write $\delta(n) = \delta_2(n)$.

Definition 2 (Predictive dependence measure): Let $p \ge 1$ and g_n be a Borel function on $\mathbb{R} \times \mathbb{R} \times \ldots \mapsto \mathbb{R}$ such that $g_n(\xi_0) = \mathbb{E}(X_n|\xi_0)$, $n \ge 0$. Let $\omega_p(I, n) = \|g_n(\xi_0) - g_n(\xi_{0,I})\|_p$ and $\omega_p(n) = \|g_n(\xi_0) - g_n(\xi_0^*)\|_p$. Write $\omega(n) = \omega_2(n)$.

Definition 3 (p-stability): Let $p \ge 1$. The process (X_n) is said to be p-stable if $\Omega_p := \sum_{n=0}^\infty \omega_p(n) < \infty$, and p-strong stable if $\Delta_p := \sum_{n=0}^\infty \delta_p(n) < \infty$. If $\Omega = \Omega_2 < \infty$, we say that (X_n) is stable.

By the causal representation in Eq. **1**, if $\min\{i : i \in I\} > n$, then $\delta_p(I, n) = 0$. Apparently, $\delta_p(I, n)$ quantifies the dependence of $X_n = g(\xi_n)$ on $\{\varepsilon_i, i \in I\}$ by measuring the distance between $g(\xi_n)$ and its coupled version $g(\xi_{n,I})$. In *Definition 2*, $\mathbb{E}(X_n|\xi_0)$ is the n-step ahead predicated mean, and $\omega_p(n)$ measures the contribution of ε_0 in predicting future expected values. In the

Abbreviation: iid, independent and identically distributed.

†E-mail: wbwu@galton.uchicago.edu.

© 2005 by The National Academy of Sciences of the USA

classical prediction theory (14), the conditional expectation of the form $\mathbb{E}(X_n|X_0, X_{-1}, \ldots)$ is studied. The one $\mathbb{E}(X_n|\xi_0)$ used in *Definition 2* has a different form. It turns out that, in studying asymptotic properties and moment inequalities of S_n, it is convenient to use $\mathbb{E}(X_n|\xi_0)$ and predictive dependence measure (*cf. Theorems 2* and *3*), while the other version $\mathbb{E}(X_n|X_0, X_{-1}, \ldots)$ is generally difficult to work with. In the special case in which X_n are martingale differences with respect to the filter $\sigma(\xi_n)$, $g_n = 0$ almost surely and consequently $\omega(n) = 0$, $n \geq 1$.

Roughly speaking, since $g_n(\xi_0) = \mathbb{E}(X_n|\xi_0)$, the p-stability in *Definition 3* indicates that the cumulative contribution of ε_0 in predicting future expected values $\{\mathbb{E}(X_n|\xi_0)\}_{n\geq 0}$ is finite. Interestingly, the stability condition $\Omega_2 < \infty$ implies invariance principles with $\sqrt{n}$-norming in a natural way (*Theorem 3*). By (*i*) of *Theorem 1*, p-strong stability implies p-stability since $\delta_p(n) \geq \omega_p(n)$.

Our dependence measures provide a very convenient and simple way for a large-sample theory for stationary causal processes (see *Theorems 2–5* below). In many applications, functional and predictive dependence measures are easy to use because they are directly related to data-generating mechanisms and because the construction of the coupled process $g(\xi_{n,I})$ is simple and explicit. Additionally, limit theorems with those dependence measures have easily verifiable conditions and are often optimal or nearly optimal. On the other hand, however, our dependence measures rely on the representation **1**, whereas the strong mixing coefficients can be defined in more general situations (6).

Theorem 1. *(i) Let $p \geq 1$ and $n \geq 0$. Then $\delta_p(n) \geq \omega_p(n)$. (ii) Let $p \geq 1$ and the projection operator $\mathcal{P}_k Z = \mathbb{E}(Z|\xi_k) - \mathbb{E}(Z|\xi_{k-1})$, $Z \in \mathcal{L}^p$. Then for $n \geq 0$,*

$$\|\mathcal{P}_0 X_n\|_p \leq \omega_p(n) \leq 2\|\mathcal{P}_0 X_n\|_p \,. \qquad [4]$$

(iii) Let $p > 1$, $C_p = 18p^{3/2}(p-1)^{-1/2}$ if $1 < p < 2$, $C_p = \sqrt{2p}$ if $p \geq 2$; let $I \subset \mathbb{Z}$. Then

$$\delta_p^{p'}(I, n) \leq 2^{p'} C_p^{p'} \sum_{i\in I} \delta_p^{p'}(n-i), \quad \text{where } p' = \min(p, 2). \qquad [5]$$

Proof: (*i*) Since $\xi_n^* = (\xi_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_n)$,

$$\begin{aligned}&\mathbb{E}[g(\xi_n) - g(\xi_n^*)|\xi_{-1}, \varepsilon_0', \varepsilon_0]\\ &\quad = \mathbb{E}[g(\xi_n)|\xi_{-1}, \varepsilon_0] - \mathbb{E}[g(\xi_n^*)|\xi_{-1}, \varepsilon_0']\\ &\quad = g_n(\xi_0) - g_n(\xi_0^*),\end{aligned}$$

which by Jensen's inequality implies $\delta_p(n) \geq \omega_p(n)$. (*ii*) Since $\mathbb{E}[g(\xi_n)|\xi_{-1}] = \mathbb{E}[g_n(\xi_0)|\xi_{-1}]$ and ε_0' and (ε_i) are independent, we have $\mathbb{E}[g_n(\xi_0)|\xi_{-1}] = \mathbb{E}[g_n(\xi_0^*)|\xi_0]$ and inequality **4** follows from

$$\begin{aligned}\|\mathcal{P}_0 X_n\|_p &= \|\mathbb{E}[g_n(\xi_0) - g_n(\xi_0^*)|\xi_0]\|_p\\ &\leq \|g_n(\xi_0) - g_n(\xi_0^*)\|_p\\ &\leq \|g_n(\xi_0) - \mathbb{E}[g_n(\xi_0)|\xi_{-1}]\|_p\\ &\qquad + \|\mathbb{E}[g_n(\xi_0)|\xi_{-1}] - g_n(\xi_0^*)\|_p\\ &= 2\|\mathcal{P}_0 X_n\|_p.\end{aligned}$$

(*iii*) For presentational clarity, let $I = \{\ldots, -1, 0\}$. For $i \leq 0$ let

$$\begin{aligned}D_i = D_{i,n} &= \mathbb{E}(X_n|\varepsilon_{i+1}, \varepsilon_{i+2}, \ldots, \varepsilon_n) - \mathbb{E}(X_n|\varepsilon_i, \ldots, \varepsilon_n)\\ &= \mathbb{E}[g(\xi_{n,\{i\}}) - g(\xi_n)|\varepsilon_i, \ldots, \varepsilon_n].\end{aligned}$$

Then $D_0, D_{-1}, \ldots$ are martingale differences with respect to the sigma algebras $\sigma(\varepsilon_i, \ldots, \varepsilon_n)$, $i = 0, -1, \ldots$. By Jensen's inequality, $\|D_i\|_p \leq \delta_p(n-i)$. Let $V = \sum_{i=-\infty}^{0} D_i^2$, $M = \sum_{i=-\infty}^{0} D_i$ and $\tilde{X}_n = \mathbb{E}(X_n|\varepsilon_1, \ldots, \varepsilon_n)$. Then $X_n - \tilde{X}_n = -M$ and

$$\delta_p(I, n) \leq \|X_n - \tilde{X}_n\|_p + \|\tilde{X}_n - g(\xi_{n,I})\|_p = 2\|M\|_p \,.$$

To show Eq. **5**, we shall deal with the two cases $1 < p < 2$ and $p \geq 2$ separately. If $1 < p < 2$, then $V^{p/2} \leq \sum_{i=-\infty}^{0} |D_i|^p$. By Burkholder's inequality (15)

$$\|M\|_p^p \leq C_p^p \|V^{1/2}\|_p^p \leq C_p^p \sum_{i=-\infty}^{0} \delta_p^p(n-i).$$

If $p \geq 2$, by proposition 4 in ref. 16, $\|M\|_p^2 \leq 2p \sum_{i=-\infty}^{0} \|D_i\|_p^2$. So Eq. **5** follows.

Inequality **5** suggests the interesting reduction property: the degree of dependence of X_n on $\{\varepsilon_i, i \in I\}$ can be bounded in an element-wise manner, and it suffices to consider the dependence of X_n on individual ε_i. Indeed, our limit theorems and moment inequalities in *Theorems 2–5* involve conditions only on $\delta_p(n)$ and $\omega_p(n)$.

Linear Processes. Let ε_i be iid random variables with $\varepsilon_i \in \mathcal{L}^p$, $p \geq 1$; let (a_i) be real coefficients such that

$$X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}, \qquad [6]$$

is a proper random variable. The existence of X_t can be checked by Kolmogorov's three series theorem. The linear process (X_t) can be viewed as the output from a linear filter and the input $(\ldots, \varepsilon_{t-1}, \varepsilon_t)$ is a series of shocks that drive the system (ref. 17, pp. 8–9). Clearly, $\omega_p(n) = \delta_p(n) = |a_n|c_0$, where $c_0 = \|\varepsilon_0 - \varepsilon_0'\|_p < \infty$. Let $p = 2$. If

$$\sum_{i=0}^{\infty} |a_i| < \infty, \qquad [7]$$

then the filter is said to be stable (17) and the preceding inequality implies short-range dependence since the covariances are absolutely summable. *Definition 3* extends the notion of stability to nonlinear processes.

Volterra Series. Analysis of nonlinear systems is a notoriously difficult problem, and the available tools are very limited (18). Oftentimes it would be unsatisfactory to linearize or approximate nonlinear systems by linear ones. The Volterra representation provides a reasonably simple and general way. The idea is to represent Eq. **3** as a power series of inputs. In particular, suppose that g in Eq. **3** is sufficiently well-behaved so that it has the stationary and causal representation

$$\begin{aligned}&g(\ldots, e_{n-1}, e_n)\\ &\quad = \sum_{k=1}^{\infty} \sum_{u_1, \ldots, u_k=0}^{\infty} g_k(u_1, \ldots, u_k) e_{n-u_1} \ldots e_{n-u_k}, \end{aligned} \qquad [8]$$

where functions g_k are called the Volterra kernel. The right-hand side of Eq. **8** is generically called the Volterra expansion, and it plays an important role in the nonlinear system theory (13, 18–22). There is a continuous-time version of Eq. **8** with summations replaced by integrals. Because the series involved has infinitely many terms, to guarantee the meaningfulness of the representation, there is a convergence issue that is often difficult to deal with, and the imposed conditions can be quite restrictive (18). Fortunately, in our setting, the difficulty can be circumvented because we are dealing with iid random inputs. Indeed, assume that e_t are iid with mean 0, variance 1 and $g_k(u_1, \ldots, u_k)$ is symmetric in $u_1, \ldots, u_k$ and it equals zero if $u_i = u_j$ for some $1 \le i < j \le k$, and

$$\sum_{k=1}^{\infty} \sum_{u_1, \ldots, u_k=0}^{\infty} g_k^2(u_1, \ldots, u_k) < \infty.$$

Then X_n exists and is in $\mathcal{L}^2$. Simple calculations show that

$$\frac{\omega^2(n)}{2} = \sum_{k=1}^{\infty} \sum_{\min(u_1, \ldots u_k)=n} g_k^2(u_1, \ldots, u_k)$$

$$= \sum_{k=1}^{\infty} k \sum_{u_2, \ldots u_k=n+1}^{\infty} g_k^2(n, u_2, \ldots, u_k),$$

and

$$\frac{\delta^2(n)}{2} = \sum_{k=1}^{\infty} k \sum_{u_2, \ldots u_k=0}^{\infty} g_k^2(n, u_2, \ldots, u_k).$$

The Volterra process is stable if $\sum_{i=1}^{\infty}\omega(i) < \infty$.

Nonlinear Transforms of Linear Processes. Let (X_t) be the linear process defined in Eq. **6** and consider the transformed process $Y_t = K(X_t)$, where K is a possibly nonlinear filter. Let $\omega(n, Y)$ be the predictive dependence measure of (Y_t). Assume that ε_i have mean 0 and finite variance. Under mild conditions on K, we have $\|\mathcal{P}_0 Y_n\| = O(|a_n|)$ (*cf.* theorem 2 in ref. 23). By *Theorem 1*, $\omega(n, Y) = O(|a_n|)$. In this case, if (X_t) is stable, namely Eq. **7** holds, then (Y_t) is also stable.

Quite interesting phenomena happen if (X_n) is unstable. Under appropriate conditions on K, (Y_n) could possibly be stable. With a nonlinear transform, the dependence structure of (Y_t) can be quite different from that of (X_n) (24–27). The asymptotic problem of $S_n(K) = \sum_{t=1}^{n} K(X_t)$ has a long history (see refs. 23 and 27 and references therein). Let $K_\infty(w) = \mathbb{E}[K(w + X_t)]$ and assume $K_\infty \in \mathcal{C}^\tau(\mathbb{R})$ for some $\tau \in \mathbb{N}$. Consider the remainder of the τ-th order Volterra expansion of Y_n

$$L^{(\tau)}(\xi_n) = Y_n - \sum_{r=0}^{\tau} \kappa_r U_{n,r}, \qquad [9]$$

where $\kappa_r = K_\infty^{(r)}(0)$, $r = 0, \ldots, \tau$, and

$$U_{n,r} = \sum_{0 \le j_1 < \ldots < j_r < \infty} \prod_{s=1}^{r} a_{j_s}\varepsilon_{n-j_s}.$$

Let $\theta_n = |a_{n-1}|[|a_{n-1}| + A_n^{1/2}(4) + A_n^{\tau/2}(2)]$ and $A_n(j) = \sum_{t=n}^{\infty} |a_t|^j$. Under mild regularity conditions on K and ε_n, by theorem 5 in ref. 23, $\|\mathcal{P}_0 L^{(\tau)}(\xi_n)\| = O(\theta_{n+1})$. By *Theorem 1*, the predictive dependence measure $\omega^{(\tau)}(n)$ of the remainder $L^{(\tau)}(\xi_n)$ satisfies

$$\omega^{(\tau)}(n) = O(\theta_{n+1}). \qquad [10]$$

It is possible that $\sum_{n=1}^{\infty} \theta_n < \infty$ while $\sum_{n=1}^{\infty} |a_n| = \infty$. Consider the special case $a_n = n^{-\beta}l(n)$, where $1/2 < \beta < 1$ and l is a slowly varying function, namely, for any $c > 0$. $l(cn)/l(n) \to 1$ as $n \to \infty$. By Karamata's Theorem (28) for $j \ge 2$, $A_n(j) = O[n^{1-\beta j}l^j(n)]$. If $\tau > (2\beta - 1)^{-1} - 1$, then $\theta_n = O[n^{\tau(1/2-\beta)}l^\tau(n)]$ is summable. Therefore, if the function K satisfies $\kappa_r = 0$ for $r = 0, \ldots, \tau$ and $(\tau + 1)(2\beta - 1) > 1$, then $Y_t = K(X_t)$ is stable even though X_t is not. Appell polynomials (29) satisfy such conditions. For example, let $K(x) = x^2 - \mathbb{E}(X_n^2)$, then $K_\infty(w) = w^2$ and $\kappa_1 = 0$, $\kappa_2 = 2$. If $\beta \in (3/4, 1)$, then the process $X_t^2 - \mathbb{E}(X_t^2)$ is stable. If $1/2 < \beta < 3/4$, then $S_n(K)/\|S_n(K)\|$ converges to the Rosenblatt distribution.

Uniform Volterra expansions for $F_n(x)$ over $x \in \mathbb{R}$ are established in refs. 30 and 31. Wu (32) considered nonlinear transforms of linear processes with infinite variance innovations.

Nonlinear Time Series. Let ε_t be iid random variables and consider the recursion

$$X_t = R(X_{t-1}, \varepsilon_t), \qquad [11]$$

where R is a measurable function. The framework **11** is quite general, and it includes many popular nonlinear time series models, such as threshold autoregressive models (33), exponential autoregressive models (34), bilinear autoregressive models, autoregressive models with conditional heteroscedasticity (35), among others. If there exists $\alpha > 0$ and x_0 such that

$$\mathbb{E}(\log L_\varepsilon) < 0 \quad \text{and} \quad L_{\varepsilon_0} + |R(x_0, \varepsilon_0)| \in \mathcal{L}^\alpha, \qquad [12]$$

where

$$L_\varepsilon = \sup_{x \ne x'} \frac{|R(x, \varepsilon) - R(x', \varepsilon)|}{|x - x'|},$$

then Eq. **11** admits a unique stationary distribution (36), and iterations of Eq. **11** give rise to Eq. **1**. By theorem 2 in ref. 37, Eq. **12** implies that there exists $p > 0$ and $r \in (0, 1)$ such that

$$\|X_n - g(\xi_{n,I})\|_p = O(r^n), \qquad [13]$$

where $I = \{\ldots, -1, 0\}$. Recall $\xi_n^* = \xi_{n,\{0\}}$. By stationarity, $\|g(\xi_n^*) - g(\xi_{n,I})\|_p = \|g(\xi_{n+1}) - g(\xi_{n+1,I})\|_p$. So Eq. **13** implies $\delta_p(n) = \|g(\xi_n^*) - X_n\|_p = O(r^n)$. On the other hand, by *Theorem 1 (iii)*, if $\delta_p(n) = O(r^n)$ holds for some $p > 1$ and for some $r \in (0,1)$, then Eq. **13** also holds. So they are equivalent if $p > 1$. In refs. 37 and 38, the property **13** is called geometric-moment contraction, and it is very useful in studying asymptotic properties of nonlinear time series.

Inequalities and Limit Theorems

For (X_i) defined in Eq. **1**, let $S_u = S_n + (u - n)X_{n+1}$, $n \le u \le n + 1$, $n = 0, 1, \ldots,$ be the partial sum process. Let $R_n(s) = \sqrt{n}[F_n(s) - F(s)]$, where $F(s) = \mathbb{P}(X_0 \le s)$ is the distribution function of X_0. Primary goals in the limit theory of stationary processes include obtaining asymptotic properties of $\{S_u, 0 \le u \le n\}$ and $\{R_n(s), s \in \mathbb{R}\}$. Such results are needed in the related statistical inference. The physical and predictive dependence measures provide a natural vehicle for an asymptotic theory for S_n and R_n.

Partial Sums. Let $S_n^* = \max_{i \le n} |S_i|$, $Z_n = S_n^*/\sqrt{n}$ and $B_p = p\sqrt{2p}/(p - 1)$, $p > 1$. Recall $\Omega_p = \sum_{k=0}^{\infty}\omega_p(k)$ and let

$$\Theta_p = \sum_{k=0}^{\infty} \|\mathcal{P}_0 X_k\|_p.$$

By *Theorem 1*, $\Theta_p \le \Omega_p \le 2\Theta_p$. Moment inequalities and limit theorems of S_n are given in *Theorems 2* and *3*, respectively. Denote by *IB* the standard Brownian motion. An interesting feature in the large deviation result in *Theorem 2(ii)* is that Ω_p and X_k do not need to be bounded.

Theorem 2. *Let $p \geq 2$. (i) We have $\|Z_n\|_p \leq B_p\Theta_p \leq B_p\Omega_p$. (ii) Let $0 < \alpha \leq 2$ and assume*

$$\gamma := \limsup_{p\to\infty} p^{1/2-1/\alpha}\Omega_p < \infty. \qquad [14]$$

Then $m(t) := \sup_{n\in\mathbb{N}}\mathbb{E}[\exp(tZ_n^\alpha)] < \infty$ for $0 \leq t < t_0$, where $t_0 = (e\alpha\gamma^\alpha)^{-1}2^{-\alpha/2}$. Consequently, for $u > 0$, $\mathbb{P}(Z_n > u) \leq \exp(-tu^\alpha)m(t)$.

Proof: (*i*) It follows from W.B.W. (unpublished results) and theorem 2.5 in ref. 39. For completeness we present the proof here. Let $M_{k,j} = \sum_{i=1}^j \mathcal{P}_{i-k}X_i$, $k, j \geq 0$ and $M^*_{k,n} = \max_{j\leq n}|M_{k,j}|$. Then $S_n = \sum_{k=0}^\infty M_{k,n}$. By Doob's maximal inequality and theorem 2.5 in ref. 39 (or proposition 4 in ref. 16),

$$\|M^*_{k,n}\|_p \leq p(p-1)^{-1}\|M_{k,n}\|_p \leq B_p\sqrt{n}\|M_{k,1}\|_p.$$

Since $S^*_n \leq \sum_{k=0}^\infty M^*_{k,n}$, (*i*) follows. (*ii*) Let $Z = Z_n$ and $p_0 = [2/\alpha] + 1$. By Stirling's formula and Eq. **14**

$$\limsup_{p\to\infty}\frac{tB^\alpha_{\alpha p}\Omega^\alpha_{\alpha p}}{(p!)^{1/p}} = \limsup_{p\to\infty}\frac{tB^\alpha_{\alpha p}\Omega^\alpha_{\alpha p}}{(2\pi p)^{1/(2p)}p/e} = te\alpha\gamma^\alpha 2^{\alpha/2} < 1.$$

By (*i*), since $e^\nu = \sum_{p=0}^\infty \nu^p/(p!)$, (*ii*) follows from

$$\sum_{p=p_0}^\infty \frac{\mathbb{E}[(tZ^\alpha)^p]}{p!} \leq \sum_{p=p_0}^\infty \frac{t^p(B_{\alpha p}\Omega_{\alpha p})^{\alpha p}}{p!} < \infty.$$

Example 1: For the linear process **6**, assume that

$$\#\{i: |a_i| > \eta\} = O(\eta^{-1/2}) \quad \text{as} \quad \eta \downarrow 0, \qquad [15]$$

and $A := \mathbb{E}(e^{|\varepsilon_0|}) < \infty$. We now apply (*ii*) of *Theorem 2* to the sum $n[F_n(u) - F(u)] = \sum_{i=1}^n \tilde{g}(\xi_i)$, where $\tilde{g}(\xi_i) = \mathbf{1}_{X_i\leq u} - F(u)$. To this end, we need to calculate the predictive dependence measure $\omega_p(n, \tilde{g})$ (say) of the process $\tilde{g}(\xi_n)$. Without loss of generality let $a_0 = 1$. Let F_ε and f_ε be the distribution and density functions of ε_0 and assume $c := \sup_u f_\varepsilon(u) < \infty$. Then Eq. **14** holds with $\alpha = 1$. To see this, let $Y_{n-1} = X_n - \varepsilon_n$, $Z_{n-1} = Y_{n-1} - a_n\varepsilon_0$ and $Y^*_{n-1} = Z_{n-1} + a_n\varepsilon'_0$. Let $n \geq 1$. Then $\mathbb{E}(\mathbf{1}_{X_n\leq u}|\xi_0) = \mathbb{E}[F_\varepsilon(u - Y_{n-1})|\xi_0]$ and $\mathbb{E}[F_\varepsilon(u - Z_{n-1})|\xi^*_0] = \mathbb{E}[F_\varepsilon(u - Z_{n-1})|\xi_0]$. By the triangle inequality,

$$\begin{aligned} Q_n &:= |\mathbb{E}[F_\varepsilon(u - Y_{n-1})|\xi_0] - \mathbb{E}[F_\varepsilon(u - Y^*_{n-1})|\xi^*_0]| \\ &\leq |\mathbb{E}[F_\varepsilon(u - Y_{n-1})|\xi_0] - \mathbb{E}[F_\varepsilon(u - Z_{n-1})|\xi_0]| \\ &\quad + |\mathbb{E}[F_\varepsilon(u - Z_{n-1})|\xi^*_0] - \mathbb{E}[F_\varepsilon(u - Y^*_{n-1})|\xi^*_0]| \\ &\leq \mathbb{E}[c|Y_{n-1} - Z_{n-1}||\xi_0] + \mathbb{E}[c|Z_{n-1} - Y^*_{n-1})|\xi^*_0] \\ &= c|a_n|(|\varepsilon_0| + |\varepsilon'_0|). \end{aligned}$$

Hence, $\omega_p(n, \tilde{g}) = \|Q_n\|_p \leq 2c|a_n|\|\varepsilon_0\|_p$. Since $A = \mathbb{E}(e^{|\varepsilon_0|})$, we have $\mathbb{E}(|\varepsilon_0|^p) \leq p!A$, $\|\varepsilon_0\|_p \leq pA^{1/p}$. Clearly, $0 \leq Q_n \leq 1$. So $\omega_p(n, \tilde{g}) \leq \min(1, C|a_n|p)$, where $C = 2cA$. For $\eta > 0$ let the set $J(\eta) = \{i \geq 0 : \eta/2 \leq |a_i| < \eta\}$. By Eq. **15**

$$\begin{aligned} \Omega_p &\leq \sum_{i=0}^\infty \min(1, C|a_i|p) \\ &= \sum_{i:|a_i|\geq p^{-1}} \min(1, C|a_i|p) \\ &\quad + \sum_{k=0}^\infty \sum_{i\in J((p2^k)^{-1})} \min(1, C|a_i|p) \\ &= O(\sqrt{p}) + \sum_{k=0}^\infty O[(p2^{k+1})^{1/2}(p2^k)^{-1}Cp] \\ &= O(\sqrt{p}). \end{aligned}$$

Condition **15** holds if $a_i = O(i^{-2})$.

Theorem 3. (*i*) *Assume that $\Omega_2 < \infty$. Then*

$$\{S_{nt}/\sqrt{n}, 0 \leq t \leq 1\} \Rightarrow \{\sigma IB(t), 0 \leq t \leq 1\}, \qquad [16]$$

where $\sigma = \|\sum_{i=0}^\infty \mathcal{P}_0X_i\| \leq \Omega_2$. (ii) Let $2 < p \leq 4$ and assume that $\sum_{i=0}^\infty i\delta_p(i) < \infty$. Then on a possibly richer probability space, there exists a Brownian motion IB such that

$$\sup_{u\in[0,n]}|S_u - \sigma IB(u)| = O[n^{1/p}l(n)] \text{ almost surely}, \qquad [17]$$

where $l(n) = (\log n)^{1/2+1/p}(\log\log n)^{2/p}$.

The proof of the strong invariance principle (*ii*) is given by W.B.W. (unpublished results). *Theorem 3*(*i*) follows from corollary 3 in ref. 40, and the expression $\sigma = \|\sum_{i=0}^\infty \mathcal{P}_0X_i\|$ is a consequence of the martingale approximation: let $D_k = \sum_{i=k}^\infty \mathcal{P}_kX_i$ and $M_n = D_1 + \ldots + D_n$, then $\|S_n - M_n\| = o(\sqrt{n})$ and $\|S_n\|/\sqrt{n} = \sigma + o(1)$ (see theorem 6 in ref. 41). *Theorem 3*(*i*) also can be proved by using the argument in ref. 42. The invariance principle in the latter paper has a slightly different form. We omit the details. See refs. 43 and 44 for some related works.

Empirical Distribution Functions. Let $H_i(u|\xi_0) = \mathbb{P}(X_i \leq u|\xi_0)$, $u \in \mathbb{R}$, be the conditional distribution function of X_i given ξ_0. By *Definition 2*, the predictive dependence measure for $\tilde{g}(\xi_i) = \mathbf{1}_{X_i\leq u} - F(u)$, at a fixed u, is $\|H_i(u|\xi_0) - H_i(u|\xi^*_0)\|_p$. To study the asymptotic properties of R_n, it is certainly necessary to consider the whole range $u \in (-\infty, \infty)$. To this end, we introduce the integrated predictive dependence measure

$$\phi_p^{(j)}(i) = \left[\int_{\mathbb{R}} \|H_i^{(j)}(u|\xi_0) - H_i^{(j)}(u|\xi^*_0)\|_p^p du\right]^{1/p}, \qquad [18]$$

and the uniform predictive dependence measure

$$\varphi_p^{(j)}(i) = \sup_u \|H_i^{(j)}(u|\xi_0) - H_i^{(j)}(u|\xi^*_0)\|_p, \qquad [19]$$

where $H_i^{(j)}(u|\xi_0) = \partial^j H_i(u|\xi_0)/\partial u^j$, $j = 0, 1, \ldots, i \geq 1$. Let $h_i(t|\xi_0) = H_i^{(1)}(t|\xi_0)$. *Theorem 4* below concerns the weak convergence of R_n based on $\phi_2^{(j)}(i)$. It follows from corollary 1 by W.B.W. (unpublished results).

Theorem 4. *Assume that $X_1 \in \mathcal{L}^\tau$ and $\sup_u h_1(u|\xi_0) \leq c_0$ for some positive constants τ, $c_0 < \infty$. Further assume that*

$$\sum_{i=1}^\infty [\phi_2^{(0)}(i) + \phi_2^{(1)}(i) + \phi_2^{(2)}(i)] < \infty. \qquad [20]$$

Then $R_n \Rightarrow \{W(s), s \in \mathbb{R}\}$, where W is a centered Gaussian process.

Kernel Density Estimation. An important problem in nonparametric inference of stochastic processes is to estimate the marginal

APPLIED MATHEMATICS

density function f (say) given the data $X_1, \ldots, X_n$. A popular method is the kernel density estimation (45,46). Let K be a bounded kernel function for which $\int_{\mathbb{R}} K(u)du = 1$ and $b_n > 1$ be a sequence of bandwidths satisfying

$$b_n \to 0 \quad \text{and} \quad nb_n \to \infty. \qquad [21]$$

Let $K_b(x) = K(x/b)$. Then f can be estimated by

$$f_n(x) = \frac{1}{nb_n} \sum_{i=1}^{n} K_{b_n}(x - X_i). \qquad [22]$$

If X_i are iid, Parzen (46) proved a central limit theorem for $f_n(x) - \mathbb{E}[f_n(x)]$ under the natural condition **21**. There has been a substantial literature on generalizing Parzen's result to time series (47,48). Wu and Mielniczuk (49) solved the open problem that, for short-range dependent linear processes, Parzen's central limit theorem holds under Eq. **21**. See references therein for historical developments. Here, we shall generalize the result in ref. 49 to nonlinear processes. To this end, we shall adopt the uniform predictive dependence measure **19**. The asymptotic normality of f_n requires a summability condition of $\varphi^{(1)}(k) = \sup_t \|h_k(t|\xi_0) - h_k(t|\xi_0^*)\|$.

Theorem 5. *Assume that* $\sup_u h_1(u|\xi_0) \le c_0$ *for some constant* $c_0 < \infty$ *and that* $f = F'$ *is continuous. Let* $\kappa := \int_{\mathbb{R}} K^2(u)du < \infty$. *Then under Eq.* **21** *and*

$$\sum_{k=1}^{\infty} \varphi^{(1)}(k) < \infty, \qquad [23]$$

we have $\sqrt{nb_n}\{f_n(x) - \mathbb{E}[f_n(x)]\} \Rightarrow N[0, f(x)\kappa]$ *for every* $x \in \mathbb{R}$.

Proof: Let m be a nonnegative integer. By the identity $\mathbb{E}[\mathbb{P}(X_{m+1} \le u|\xi_m)|\xi_0] = \mathbb{P}(X_{m+1} \le u|\xi_0)$ and the Lebesgue dominated convergence theorem, we have $\mathbb{E}[h_1(u|\xi_m)|\xi_0| = h_{m+1}(u|\xi_0)$ and h_{m+1} is also bounded by c_0. By *Theorem 1(ii)*, $\|\mathcal{P}_0 h_1(u|\xi_m)\| \le \varphi^{(1)}(m+1)$. Let $A_n(u) = \sum_{i=1}^{n} h_1(u|\xi_{i-1}) - nf(u)$. By *Theorem 2(i)* and Eq. **23**

$$\frac{\sup_u \|A_n(u)\|}{B_2\sqrt{n}} \le \sup_u \sum_{m=0}^{\infty} \|\mathcal{P}_0 h_1(u|\xi_m)\| < \infty.$$

Let $M_n = \sum_{i=1}^{n} \mathcal{P}_i K_{b_n}(x - X_i)$ and $N_n = \int_{\mathbb{R}} K(v) A_n(x - b_n v)dv$. Observe that

$$\mathbb{E}[K_{b_n}(x - X_i)|\xi_{i-1}] = b_n \int_{\mathbb{R}} K(v) h_1(x - b_n v|\xi_{i-1})dv.$$

Then $nb_n\{f_n(x) - \mathbb{E}[f_n(x)]\} = M_n + b_n N_n$. Following the argument of lemma 2 in ref. 49, $M_n/\sqrt{nb_n} \Rightarrow N[0, f(x)\kappa]$, which finishes the proof since $\mathbb{E}|N_n| = O(n^{1/2})$ and $b_n \to 0$.

I thank J. Mielniczuk, M. Pourahmadi, and X. Shao for useful comments. I am very grateful for the extremely helpful suggestions of two reviewers. This work was supported by National Science Foundation Grant DMS-0448704.

1. Wiener, N. (1958) *Nonlinear Problems in Random Theory* (MIT Press, Cambridge, MA).
2. Rosenblatt, M. (1959) *J. Math. Mech.* **8,** 665–681.
3. Rosenblatt, M. (1971) *Markov Processes. Structure and Asymptotic Behavior* (Springer, New York).
4. Kallianpur, G. (1981) in *Norbert Wiener, Collected Works with Commentaries* eds. Wiener, N. & Masani, P. (MIT Press, Cambridge, MA) pp. 402–424.
5. Rosenblatt, M. (1956) *Proc. Natl. Acad. Sci. USA* **42,** 43–47.
6. Bradley, R. C. (2005) *Introduction to Strong Mixing Conditions* (Indiana Univ. Press, Bloomington, IN).
7. Blum, J. R. & Rosenblatt, M. (1956) *Proc. Natl. Acad. Sci. USA* **42,** 412–413.
8. Doukhan, P. & Louhichi, S. (1999) *Stochastic Process. Appl.* **84,** 313–342.
9. Dedecker, J. & Prieur, C. (2005) *Probab. Theor. Rel. Fields* **132,** 203–236.
10. Rosenblatt, M. (1964) *J. Res. Natl. Bureau Standards Sect. D* **68D,** 933–936.
11. Rosenblatt, M. (1980) *J. Appl. Prob.* **17,** 265–270.
12. Andrews, D. W. K. (1984) *J. Appl. Prob.* **21,** 930–934.
13. Priestley, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis* (Academic, London).
14. Pourahmadi, M. (2001) *Foundations of Time Series Analysis and Prediction Theory* (Wiley, New York).
15. Chow, Y. S. & Teicher, H. (1988) *Probability Theory* (Springer, New York).
16. Dedecker, J. & Doukhan, P. (2003) *Stochastic Process. Appl.* **106,** 63–80.
17. Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994) *Time Series Analysis: Forecasting and Control* (Prentice–Hall, Englewood Cliffs, NJ).
18. Rugh, W. J. (1981) *Nonlinear System Theory: The Volterra/Wiener Approach* (Johns Hopkins Univ. Press, Baltimore).
19. Schetzen, M. (1980) *The Volterra and Wiener Theories of Nonlinear Systems* (Wiley, New York).
20. Casti, J. L. (1985) *Nonlinear System Theory* (Academic, Orlando, FL).
21. Bendat, J. S. (1990) *Nonlinear System Analysis and Identification from Random Data* (Wiley, New York).
22. Mathews, V. J. & Sicuranza, G. L. (2000) *Polynomial Signal Processing* (Wiley, New York).
23. Wu, W. B. (2006) *Econometric Theory* **22,** in press.
24. Dittmann, I. & Granger, C. W. J. (2002) *J. Econometrics* **110,** 113–133.
25. Sun, T. C. (1963) *J. Math. Mech.* **12,** 945–978.
26. Taqqu, M. S. (1975) *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **31,** 287–302.
27. Ho, H. C. & Hsing, T. (1997) *Ann. Prob.* **25,** 1636–1669.
28. Feller, W. (1971) *An Introduction to Probability Theory and its Applications* (Wiley, New York) Vol. II.
29. Avram, F. & Taqqu, M. (1987) *Ann. Prob.* **15,** 767–775.
30. Ho, H. C. & Hsing, T. (1996) *Ann. Stat.* **24,** 992–1024.
31. Wu, W. B. (2003) *Bernoulli* **9,** 809–831.
32. Wu, W. B. (2003) *Statistica Sinica* **13,** 1259–1267.
33. Tong, H. (1990) *Non-linear Time Series: A Dynamical System Approach* (Oxford Univ. Press, Oxford).
34. Haggan, V. & Ozaki, T. (1981) *Biometrika* **68,** 189–196.
35. Engle, R. F. (1982) *Econometrica* **50,** 987–1007.
36. Diaconis, P. & Freedman, D. (1999) *SIAM Rev.* **41,** 41–76.
37. Wu, W. B. & Shao, X. (2004) *J. Appl. Prob.* **41,** 425–436.
38. Hsing, T. & Wu, W. B. (2004) *Ann. Prob.* **32,** 1600–1631.
39. Rio, E. (2000) *Theorie Asymptotique des Processus Aleatoires Faiblement Dependants* (Springer, Berlin).
40. Dedecker, J. & Merlevède, P. (2003) *Stochastic Process. Appl.* **108,** 229–262.
41. Volný, D. (1993) *Stochastic Process. Appl.* **44,** 41–74.
42. Hannan, E. J. (1979) *Stochastic Process. Appl.* **9,** 281–289.
43. Hannan, E. J. (1973) *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **26,** 157–170.
44. Heyde, C. C. (1974) *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **30,** 315–320.
45. Rosenblatt, M. (1956) *Ann. Math. Stat.* **27,** 832–837.
46. Parzen, E. (1962) *Ann. Math. Stat.* **33,** 1965–1976.
47. Robinson, P. M. (1983) *J. Time Ser. Anal.* **4,** 185–207.
48. Bosq, D. (1996) *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction* (Springer, New York).
49. Wu, W. B. & Mielniczuk, J. (2002) *Ann. Stat.* **30,** 1441–1459.