EMPIRICAL PROCESSES OF STATIONARY SEQUENCES

Wei Biao Wu

University of Chicago

Abstract: The paper considers empirical distribution functions of stationary causal processes. Weak convergence of normalized empirical distribution functions to Gaussian processes is established and sample path properties are discussed. The Chibisov-O’Reilly Theorem is generalized to dependent random variables. The proposed dependence structure is related to the sensitivity measure, a quantity appearing in the prediction theory of stochastic processes.

Key words and phrases: Empirical process, Gaussian process, Hardy inequality, linear process, martingale, maximal inequality, nonlinear time series, prediction, short-range dependence, tightness, weak convergence.

1. Introduction

Let \( \varepsilon_k, k \in \mathbb{Z} \), be independent and identically distributed (i.i.d.) random variables and define

\[
X_n = J(\ldots, \varepsilon_{n-1}, \varepsilon_n),
\]

where \( J \) is a measurable function such that \( X_n \) is a proper random variable. The framework (1) is general enough to include many interesting and important examples. Prominent ones are linear processes and nonlinear time series arising from iterated random functions. Given the sample \( X_i, 1 \leq i \leq n \), we are interested in the empirical distribution function

\[
F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{X_i \leq x}.
\]

When the \( X_i \) are i.i.d., the weak convergence of \( F_n \) and its sample path properties have been extensively studied (Shorack and Wellner (1986)). Various generalizations have been made to dependent random variables. It is a challenging problem to develop a weak convergence theory for the associated empirical processes without the independence assumption. One way out is to impose strong mixing conditions to ensure asymptotic independence; see Billingsley (1968), Gastwirth and Rubin (1973), Withers (1973), Mehra and Rao (1975), Doukhan, Massart and Rio (1995), Andrews and Pollard (1994), Shao and Yu (1996) and Rio (2000), among others. Other special processes that have been

We now introduce some notation. Let the triple \((Ω, F, P)\) be the underlying probability space. Let \(F_ε(x|ξ_n) = P(X_{n+1} ≤ x|ξ_n)\) be the conditional distribution function of \(X_{n+1}\) given the sigma algebra generated by \(ξ_n = (\ldots, ε_{n-1}, ε_n)\); let \(F(x) = P(X_1 ≤ x)\) and \(R_n(s) = \sqrt{n}[F_n(s) - F(s)]\). Assume throughout the paper that the conditional density \(f_ε(x|ξ_n) = (\partial/\partial x)F_ε(x|ξ_n)\) exists almost surely. Define the weighted measure \(w_θ(du) = (1 + |u|)^γ(du)\). For a random variable \(Z\) write \(Z ∈ L^q, q > 0\), if \(∥Z∥_q := [E(|Z|^q)]^{1/q} < ∞\). Write the \(L^2\) norm as \(∥Z∥ = ∥Z∥_2\). Define projections \(P_kZ = E(Z|ξ_k) - E(Z|ξ_{k-1}), k ∈ Z\). Denote by \(C_q\) (resp. \(C_γ, C_µ, \text{etc}\)) generic positive constants which only depend on \(q\) (resp. \(γ, μ, \text{etc}\)). Their values may vary from line to line.

The rest of paper is structured as follows. Section 2 concerns sample path properties and weak convergence of \(R_n\). In particular, in Section 2.1, the Chibisov-O’Reilly Theorem, which concerns the weak convergence of weighted empirical processes, is generalized to dependent random variables. Section 2.2 considers the weighted modulus of continuity of \(R_n\). Applications to linear processes and nonlinear time series are made in Section 3. Section 4 presents some useful inequalities that may be of independent interest. The inequalities are applied in Section 5, where proofs of the main results are given.

2. Main results

It is certainly necessary to impose appropriate dependence structures on the process \((X_t)\). We start by proposing a particular dependence condition which is quite different from the classical strong mixing assumption s. Let \(θ\) be a measure on the 1-dimensional Borel space \((R, B)\). For \(θ \in R\), let \(T_n(θ) = \sum_{i=1}^{n} h(θ, ξ_i) - nE[h(θ, ξ_1)]\), where \(h\) is a measurable function such that \(∥h(θ, ξ_1)∥ < ∞\) for almost all \(θ\) (m). Denote by \(h_j(θ, ξ_0) = E[h(θ, ξ_j)|ξ_0]\) the \(j\)-step-ahead predicted mean, \(j ≥ 0\). Let \((ε_i')\) be an i.i.d. copy of \((ε_i, \ldots, ε_{-1}, ε_0, ε_1, \ldots, ε_k)\), \(k ≥ 0\), and define

\[
σ(h, m) = ∑_{j=0}^{∞} \left[ ∫_R ∥h_j(θ, ξ_0) - h_j(θ, ξ_0')∥^2 m(θ)dθ \right]^{1/2}.
\]

(3)

Let \(f_ε(θ)|ξ_k) = (∂/∂θ)f_ε(θ|ξ_k)\). In the case that \(h(θ, ξ_k) = F_ε(θ|ξ_k)\) [resp. \(f_ε(θ|ξ_k)\) or \(f_ε'(θ|ξ_k)\)], we write \(σ(F_ε, m)\) [resp. \(σ(f_ε, m)\) or \(σ(f_ε', m)\)] for \(σ(h, m)\).

Our dependence conditions are expressed as \(σ(F_ε, m) < ∞\), \(σ(f_ε, m) < ∞\) and \(σ(f_ε', m) < ∞\), where \(m(dt) = (1 + |t|)^θ dt\), \(δ ∈ R\). These conditions are interestingly connected with the prediction theory of stochastic processes (Remark 1 and Section 2.3). The quantity \(σ(h, m)\) can be interpreted
as a cumulative weighted prediction measure. It is worthwhile to note that 
\( \| h_j(\theta, \xi_0) - h_j(\theta, \xi_0^*) \| \) has the same order of magnitude as 
\( \| P_0 h(\theta, \xi_j) \| \). To see this, let \( g(\ldots, e_{-1}, e_0) = h(\theta, \ldots, e_{-1}, e_0) \). So \( g(\xi_j^*) = h(\theta, \xi_j^*), g_j(\xi_0^*) = E[g(\xi_j^*)|\xi_0^*] = h_j(\theta, \xi_0^*), \) \( E[g(\xi_j)|\xi_j-1] = E[g(\xi_j^*)|\xi_0] \), and

\[
\| P_0 g(\xi_j) \| = \| E[g_j(\xi_0) - g_j(\xi_0^*)|\xi_0] \| \leq \| g_j(\xi_0) - g_j(\xi_0^*) \| \\
\leq \| g_j(\xi_0) - g_{j+1}(\xi_{-1}) \| + \| g_{j+1}(\xi_{-1}) - g_j(\xi_0^*) \| = 2\| P_0 g(\xi_j) \|. \quad (4)
\]

Note also that \( \| P_0 g(\xi_j) \| \leq \| g(\xi_j) - g(\xi_j^*) \| \). Therefore we have

\[
\frac{1}{2} \sigma(h, m) \leq \tilde{\sigma}(h, m) := \sum_{j=0}^{\infty} \left[ \int_{\mathbb{R}} \| P_0 h(\theta, \xi_j) \|^2 m(d\theta) \right]^{\frac{1}{2}} \leq \sigma(h, m) \\
\leq 2 \sum_{j=0}^{\infty} \left[ \int_{\mathbb{R}} \| h(\theta, \xi_j) - h(\theta, \xi_j^*) \|^2 m(d\theta) \right]^{\frac{1}{2}} := 2\tilde{\sigma}(h, m). \quad (5)
\]

All results in the paper with the condition \( \sigma(h, m) < \infty \) can be re-stated in terms of \( \tilde{\sigma}(h, m) < \infty \), or the stronger version \( \tilde{\sigma}(h, m) < \infty \).

### 2.1. Weak convergence

The classical Donsker theorem asserts that, if \( X_k \) are i.i.d., then \( \{ R_n(s), s \in \mathbb{R} \} \) converges in distribution to an \( F \)-Brownian bridge process. The Donsker theorem has many applications in statistics. To understand the behavior at the two extremes \( s = \pm \infty \), we need to consider the weighted version \( \{ R_n(s)W(s), s \in \mathbb{R} \} \), where \( W(s) \to \infty \) as \( s \to \pm \infty \). Clearly, if \( W \) is bounded, then by the Continuous Mapping Theorem, the weak convergence of the weighted empirical processes follows from that of \( R_n \). By allowing \( W(s) \to \infty \) as \( s \to \pm \infty \), one can have the weak convergence of functionals of empirical processes, \( t(F_n) \), for a wider class of functionals. The Chibisov-O’Reilly Theorem concerns weighted empirical processes of i.i.d. random variables; see Shorack and Wellner (1986, Sec. 11.5). The case of dependent random variables has been far less studied. For strongly mixing processes see Mehra and Rao (1975), Shao and Yu (1996) and Csörgő and Yu (1996).

**Theorem 1.** Let \( \gamma' \geq 0, q > 2, \) and \( \gamma = \gamma'q/2. \) Assume \( \mathbb{E}[|X_1|^\gamma + \log(1+|X_1|)] < \infty \) and

\[
\int_{\mathbb{R}} \mathbb{E}[f_{\xi}^q(u|\xi_0)]w_{\gamma-1+\frac{q}{2}}(du) < \infty. \quad (6)
\]

In addition assume that there exists \( 0 \leq \mu \leq 1 \) and \( \nu > 0 \) such that

\[
\sigma(F_\xi, w_{\gamma-\mu}) + \sigma(f_\xi, w_{\gamma+\mu}) + \sigma(f_\xi', w_{-\nu}) < \infty. \quad (7)
\]
Then (i) $\mathbb{E}[\sup_{s \in \mathbb{R}} |R_n(s)|^2(1+|s|)^{\gamma}] = O(1)$, and (ii) the process $\{R_n(s)(1+|s|)^{\gamma/2}, s \in \mathbb{R}\}$ converges weakly to a tight Gaussian process. In particular, if $X_1 \in L^{\gamma+q/2-1}$ and

$$\sup_{u} f_{\varepsilon}(u|\xi_0) \leq C$$

holds almost surely for some constant $C < \infty$, then (6) holds.

An important issue in applying Theorem 1 is to verify (7), which is basically a short-range dependence condition (cf. Remark 1). For many important models including linear processes and Markov chains, (7) is easily verifiable (Section 3). If $(X_n)$ is a Markov chain, then $\sigma(h,m)$ is related to the sensitivity measure (Fan and Yao (2003, p. 466)) appearing in nonlinear prediction theory (Section 2.3). If $(X_n)$ is a linear process, then (7) reduces to the classical definition of the short-range dependence of linear processes.

Remark 1. If $h(\theta, \xi_k) = f_{\varepsilon}(\theta|\xi_k)$, then $h_k(\theta, \xi_0) = \mathbb{E}[f_{\varepsilon}(\theta|\xi_k)|\xi_0] = f_{k+1}(\theta|\xi_0)$, the conditional density of $X_{k+1}$ at $\theta$ given $\xi_0$. Note that $\xi^*_0 = (\xi_{-1}, \varepsilon_0)$ is a coupled version of $\xi_0$ with $\varepsilon_0$ replaced by $\varepsilon^*_0$. So $h_k(\theta, \xi_0) - h_k(\theta, \xi^*_0) = f_{k+1}(\theta|\xi_0) - f_{k+1}(\theta|\xi^*_0)$ measures the change in the $(k+1)$-step-ahead predictive distribution if $\xi_0$ is changed to its coupled version $\xi^*_0$. In other words, $h_k(\theta, \xi_0) - h_k(\theta, \xi^*_0)$ can be viewed as the contribution of $\varepsilon_0$ in predicting $X_{k+1}$. So, in this sense, the condition $\sigma(h,m) < \infty$ means that the cumulative contribution of $\varepsilon_0$ in predicting future values $X_k, k \geq 1$, is finite. It is then not unnatural to interpret $\sigma(h,m)$ as a cumulative weighted prediction measure. This interpretation seems in line with the connotation of short-range dependence.

We now compare Theorem 1 with the Chibisov-O’Reilly Theorem that concerns the weak convergence of weighted empirical processes for i.i.d. random variables. Note that $(\gamma+q/2-1) \downarrow \gamma'$ as $q \downarrow 2$. The moment condition $X_1 \in L^{\gamma+q/2-1}$ in Theorem 1 is almost necessary in the sense that it cannot be replaced by the weaker

$$\mathbb{E}\{|X_1|^{\gamma'} \log^{-1}(2+|X_1|)\log\log(10+|X_1|)^{-\lambda} \} < \infty$$

for some $\lambda > 0$. To see this let $X_k$ be i.i.d. symmetric random variables with continuous, strictly increasing distribution function $F$; let $F^#$ be the quantile function and $m(u) = [1+|F^#(u)|]^{-\gamma'/2}$. Then we have the distributional equality

$$\{R_n(s)(1+|s|)^{\gamma'}, s \in \mathbb{R}\} \overset{d}{=} \{\frac{R_n(F^#(u))}{m(u)}, u \in (0,1)\}.$$

Assume that $F(s)(1+|s|)^{\gamma'}$ is increasing on $(-\infty, G)$ for some $G < 0$. Then $m(u)/\sqrt{u}$ is decreasing on $(0, F(G))$. By the Chibisov-O’Reilly Theorem,
\{R_n(F^#(u))/m(u), u \in (0, 1)\} is tight if and only if \(\lim_{t \to 0} m(t)/\sqrt{t \log \log (t^{-1})} = \infty\), namely
\[
\lim_{u \to -\infty} F(u)(1 + |u|)^{\gamma'} \log \log |u| = 0. \tag{10}
\]
Condition (10) controls the heaviness of the tail of \(X_1\). Let \(F(u) = |u|^{-\gamma'} (\log \log |u|)^{-1}\) for \(u \leq -10\). Then (9) holds while (10) is violated. It is unclear whether stronger versions of (9), such as \(E(|X_1|^{\gamma'}) < \infty\) or \(E(|X_1|^{\gamma'} \log^{-1}(2 + |X_1|)) < \infty\), are sufficient.

2.2. Modulus of continuity

Theorem 2 concerns the weighted modulus of continuity of \(R_n(\cdot)\). Sample path properties of empirical distribution functions of i.i.d. random variables have also been extensively explored: see for example Csörgö et al. (1986), Shorack and Wellner (1986), and Einmahl and Mason (1988), among others. It is far less studied for the dependent case.

**Theorem 2.** Let \(\gamma' > 0, 2 < q < 4\); let \(\gamma = \gamma' q/2\); let \(\delta_n < 1/2\) be a sequence of positive numbers such that \((\log n)^{2q/(q-2)} = O(n\delta_n)\). Assume (8), \(X_1 \in L^{\gamma}\), and \(\sigma(f, w_{\gamma'}) + \sigma(f', w_{\gamma'}) < \infty\).

Then there exists a constant \(0 < C < \infty\), independent of \(n\) and \(\delta_n\), such that for all \(n \geq 1\),
\[
E \left[ \sup_{t \in \mathbb{R}} (1 + |t|)^{\gamma'} \sup_{|s| \leq \delta_n} |R_n(t + s) - R_n(t)|^2 \right] \leq C\delta_n^{1-\frac{2}{q}}. \tag{12}
\]

2.3. Sensitivity measures and dependence

Our basic dependence condition is that \(\sigma(h, m) < \infty\). Here we present its connection with prediction sensitivity measures (Fan and Yao (2003, p.466)) special structure. Assume that \((X_n)\) is a Markov chain expressed in the form of an iterated random function (Elton (1990) and Diaconis and Freedman (1991)):
\[
X_n = M(X_{n-1}, \varepsilon_n), \tag{13}
\]
where \(\varepsilon_k, k \in \mathbb{Z}\), are i.i.d. random variables and \(M(\cdot, \cdot)\) is a bivariate measurable function. For \(k \geq 1\) let \(f_k(\cdot|x)\) be the conditional (transition) density of \(X_k\) given \(X_0 = x\). Then \(f_\varepsilon(\theta|\xi_k) = f_1(\theta|X_k)\) is the conditional density of \(X_{k+1}\) at \(\theta\) given \(X_k\), and, for \(k \geq 0\), \(E[f_\varepsilon(\theta|\xi_k)|\xi_0] = f_{k+1}(\theta|X_0)\). Fan and Yao (2003) argue that
\[
D_k(x, \delta) := \int_{\mathbb{R}} [f_k(\theta|x + \delta) - f_k(\theta|x)]^2 d\theta \tag{14}
\]
is a natural way to measure the deviation of the conditional distribution of $X_k$ given $X_0 = x$. In words, $D_k$ quantifies the sensitivity to initial values and it measures the error in the $k$-step-ahead predictive distribution due to a drift in the initial value. Under certain regularity conditions,

$$
\lim_{\delta \to 0} \frac{D_k(x, \delta)}{\delta^2} = \int_{\mathbb{R}} \left( \frac{\partial f_k(\theta|x)}{\partial x} \right)^2 d\theta =: I_k(x).
$$

Here $I_k$ is called prediction sensitivity measure. It is a useful quantity in the prediction theory of nonlinear dynamical systems. Estimation of $I_k$ is discussed in Fan and Yao (2003, p. 468). Proposition 1 shows the relation between $\sigma(h, m)$ and $I_k$. Since it can be proved in the same way as (i) of Theorem 3, we omit the details of its proof.

**Proposition 1.** Let $k \geq 0$. For the process (13), we have

$$
\int_{\mathbb{R}} \|h_k(\theta, \xi_0) - h_k(\theta, \xi_0^*)\|^2 m(d\theta) \leq 4 \|\tau_k(X_0, X_0^*)\|^2,
$$

where

$$
\tau_k(a, b) = \int_a^b [I_{h_k}(x, m)]^2 dx \quad \text{and} \quad I_{h_k}(x, m) = \int_{\mathbb{R}} \left( \frac{\partial h_k(\theta, x)}{\partial x} \right)^2 m(d\theta).
$$

Consequently, $\sigma(h, m) < \infty$ holds if $\sum_{k=1}^{\infty} \|\tau_k(X_0, X_0^*)\| < \infty$.

In the special case $h(\theta, \xi_k) = f_\epsilon(\theta|\xi_k) = f_1(\theta|X_k)$, $h_k(\theta, \xi_0) = \mathbb{E}[h(\theta, \xi_k)|\xi_0] = f_{k+1}(\theta|X_0)$ and $I_{h_k}(x, m)$ reduces to Fan and Yao’s sensitivity measure $I_{k+1}(x)$ provided $m(d\theta) = d\theta$ is Lebesgue measure. So it is natural to view $I_{h_k}(x, m)$ as a weighted sensitivity measure. Since the $k$-step-ahead conditional density $f_k(\theta|x)$, may have an intractable and complicated form, it is generally not very easy to apply Proposition 1. This is especially so in nonlinear time series where it is often quite difficult to derive explicit forms of $f_k(\theta|x)$. To circumvent such a difficulty, our Theorem 3 provides sufficient conditions which only involve 1-step-ahead conditional densities.

For processes that are not necessarily in the form (13), we assume that there exists an $\sigma(\xi_n)$-measurable random variable $Y_n$ such that

$$
\mathbb{P}(X_{n+1} \leq x|\xi_n) = \mathbb{P}(X_{n+1} \leq x|Y_n) := F(x|Y_n). \tag{15}
$$

Then there exists a similar bound as the one given in Proposition 1. Write $Y_n = I(\xi_n)$, $Y_n^* = I(\xi_n^*)$ and $h(\theta, \xi_n) = h(\theta, Y_n)$. For Markov chains, (15) is satisfied with $Y_n = X_n$. Let $X_n = \sum_{i=0}^{\infty} a_i \varepsilon_{n-i}$ be a linear process. Then (15) is also satisfied with $Y_n = \sum_{i=1}^{\infty} a_i \varepsilon_{n+1-i}$. Let $f(\theta|y)$ be the conditional density of
\(X_{n+1}\) given \(Y_n = y\) and \(f'(\theta | y) = (\partial/\partial \theta) f(\theta | y)\). Then \(\sigma(h,m)\) is also related to a weighted distance between \(Y_n\) and \(Y_n^*\). Define

\[
\rho_{h,m}(a,b) = \int_a^b H_{h,m}(y) dy, \quad \text{where} \quad H_{h,m}(y) = \int_{\mathbb{R}} \left| \frac{\partial}{\partial y} h(\theta, y) \right|^2 m(d\theta). \tag{16}
\]

**Theorem 3.** (i) Let \(\Xi_{h,m} = \sum_{n=0}^{\infty} \|\rho_{h,m}(Y_n, Y_n^*)\|\). Then \(\sigma(h,m) \leq 2\Xi_{h,m}\) and

\[
\int_{\mathbb{R}} \|h_n(\theta, \xi_0) - h_n(\theta, \xi_0^*)\|^2 m(d\theta) \leq 4\rho_{h,m}(Y_n, Y_n^*)^2. \tag{17}
\]

(ii) Assume there exist \(C > 0\) and \(q \in \mathbb{R}\) such that \(H_{h,m}(y) \leq C(1 + |y|)^{q-2}\) holds for all \(y \in \mathbb{R}\). Then \(\|\rho_{h,m}(Y_n, Y_n^*)\| = O(\|Y_n - Y_n^*\|_q^{\min(1, q/2)})\) if \(q > 0\), and \(\|\rho_{h,m}(Y_n, Y_n^*)\| = O(\|Y_n - Y_n^*\|_q^{\min(1, q/2)})\) if \(q < 0\).

Let \(h(\theta, Y_n) = f(\theta | Y_n)\). Then \(H_{h,m}(y)\) can be interpreted as a measure of ”local dependence” of \(X_{n+1}\) on \(Y_n\) at \(y\). As does \(D_h(x, \delta)\) in (14), \(H_{h,m}(y)\) measures the distance between the conditional densities of \([X_{n+1} | Y_n = y]\) and \([X_n = y + \delta]\). In many situations \(\|Y_n - Y_n^*\|_q\) is easy to work with since it is directly related to the data generating mechanisms.

**3. Applications**

**3.1. Iterated random functions**

Consider the process (13). The existence of stationary distributions has been widely studied and there are many versions of sufficient conditions; see Diaconis and Freedman [1999], Meyn and Tweedie [1993], Jarner and Tweedie [2001], Wu and Shao [2004], among others. Here we adopt the conditions given by Diaconis and Freedman [1999]. The recursion (13) has a unique stationary distribution if there exist \(\alpha > 0\) and \(x_0\) such that

\[
L_{\varepsilon_0} + |M(x_0, \varepsilon_0)| \in L^\alpha \text{ and } \mathbb{E}[\log(L_{\varepsilon_0})] < 0, \quad \text{where} \quad L_\varepsilon = \sup_{x \neq x'} \frac{|M(x, \varepsilon) - M(x', \varepsilon)|}{|x - x'|}. \tag{18}
\]

Condition (13) also implies the geometric-moment contraction (GMC(\(\beta\), see Wu and Shao [2004])) property: there exist \(\beta > 0\), \(r \in (0, 1)\), and \(C < \infty\) such that

\[
\mathbb{E} \left[ |J(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n) - J(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n)|^{\beta} \right] \leq C r^n \tag{19}
\]

holds for all \(n \in \mathbb{N}\), where \((\varepsilon_k)\) is an i.i.d. copy of \((\varepsilon_k)\). Hsing and Wu [2004] argued that (19) is a convenient condition for limit theorems. Doukhan [2003] gave a brief review of weak convergence of \(R_n\) under various dependence structures including strong mixing conditions. Meyn and Tweedie...
\[3.1.1. \text{ARCH models}\]

Consider the model
\[X_n = \varepsilon_n \sqrt{a^2 + b^2 X_{n-1}^2}, \]
where \(a\) and \(b\) are real parameters for which \(ab \neq 0\) and \(\varepsilon_i\) are i.i.d. innovations with density \(f\). Let \(M(x, \varepsilon) = \varepsilon \sqrt{a^2 + b^2 x^2}\). Then \(L_x = \sup_x |\partial M(x, \varepsilon)/\partial x| \leq |\varepsilon|\). Assume that there exists \(\beta > 0\) such that \(r_0 := \mathbb{E}(|\varepsilon|^\beta) < 1\). Then \(m\) holds with \(\alpha = \beta\) and \(|M(x, \varepsilon_n) - M(x', \varepsilon_n)|^\beta \leq r_0 |x - x'|^\beta\). Iterations of the latter inequality imply \((16)\) with \(r = r_0\). For more details see Wu and Shao (2004).

Corollary 1. Assume \(\mathbb{E}(|\varepsilon|^\beta) < 1, \beta > 0\). Let \(\phi \in (0, \beta)\) and \(\nu > 0\) satisfy
\[\int_{\mathbb{R}} |f'(u)|^2 w_{1+\phi}(du) + \int_{\mathbb{R}} |f''(u)|^2 w_{3+\phi}(du) + \int_{\mathbb{R}} |f'''(u)|^2 w_{2-\nu}(du) < \infty. \]

Then \(\{R_n(s)(1 + |s|)^{\phi/2}, s \in \mathbb{R}\}\) converges weakly to a tight Gaussian process.

Proof. Let \(t = t_0 = \sqrt{a^2 + b^2 \gamma^2}\). Then \(F(\theta|y) = F_x(\theta/t), f(\theta|y) = F_x(\theta/t)/t,\) and \(f'(\theta|y) = f_x'(\theta/t)/t^2\). By (21) of Lemma 1 and (21), we have \(\sup_x F_x(x) < \infty\) and (8). Choose \(q > 2\) such that \(\tau = \phi q/2 + q/2 - 1 < \beta\). Then \(X_1 \in \mathcal{L}^\tau\). By Theorem 1, it remains to verify (19). Let \(\mu = 1\). Recall (19) for \(H_{h,m}(y)\).

By considering the cases \(\phi \geq 1\) and \(\phi < 1\) separately, we have, by (21) and the inequality \(|ut| \leq 1 + |ut| \leq (1 + |u|)(1 + |t|),\) that
\[\int_{\mathbb{R}} |\partial F_x(\theta/t)/\partial t|^2 w_{\phi-1}(d\theta) = t^{-1} \int_{\mathbb{R}} f_x^2(u) u^2 (1 + |ut|)^{\phi-2} du \leq C t^{\phi-2}\]
holds for some constant $C \in (0, \infty)$. Since $|dt/du| \leq |b|$, $\int_{\mathbb{R}} |\partial F(\theta|y)/\partial y|^2 w_{\phi-1}(d\theta) \leq C(1+|y|)^{\phi-2}$. Similar but lengthy calculations show that $\int_{\mathbb{R}} |\partial f(\theta|y)/\partial y|^2 w_{\phi+1}(d\theta) \leq C(1+|y|)^{\phi-2}$ and $\int_{\mathbb{R}} |\partial f'(\theta|y)/\partial y|^2 w_{-2-\nu}(d\theta) \leq C(1+|y|)^{-\nu-2}$. By (19), $\|Y_n - Y_n^\ast\|_\phi = O(r^n)$ for some $r \in (0, 1)$. By Theorem 3, simple calculations show that (7) holds.

Corollary 1 allows heavy-tailed ARCH processes. [Tsay (2005)] argued that in certain applications it is more appropriate to assume that $\varepsilon_k$ has heavy tails. Let $\varepsilon_k$ have a standard Student-$t$ distribution with degrees of freedom $\nu$, with density $f_\nu(u) = (1 + u^2/\nu)^{-(1+\nu)/2}c_\nu$, where $c_\nu = \Gamma((\nu+1)/2)/[\Gamma(\nu/2)\sqrt{\nu\pi}]$. Then (21) holds if $\phi < 2\nu$. Note that $\varepsilon_k \in L^\phi$ if $\phi < \nu$, and consequently $X_k \in L^\phi$ if $E(|\varepsilon_0|^{\phi}) < 1$.

### 3.2. Linear processes

Let $X_t = \sum_{i=0}^{\infty} a_i \varepsilon_{t-i}$, where the $\varepsilon_k$ are i.i.d. random variables with mean 0 and finite and positive variance, and the coefficients $a_i$ satisfy $\sum_{i=0}^{\infty} a_i^2 < \infty$. Assume without loss of generality that $a_0 = 1$. Let $F_\varepsilon$ and $f_\varepsilon = F'_\varepsilon$ be the distribution and density functions of $\varepsilon_k$. Then the conditional density of $X_{n+1}$ given $\xi_n$ is $f_\varepsilon(x - Y_n)$, where $Y_n = X_{n+1} - \varepsilon_{n+1}$ (cf. (15)).

**Corollary 2.** Let $\gamma \geq 0$. Assume $\varepsilon_k \in L^{2+\gamma}$, sup$_u f_\varepsilon(u) < \infty$, and

$$\sum_{n=1}^{\infty} |a_n| < \infty,$$  \hspace{1cm} (22)

$$\int_{\mathbb{R}} |f'_\varepsilon(u)|^2 w_\gamma(du) + \int_{\mathbb{R}} |f''_\varepsilon(u)|^2 w_{-\gamma}(du) < \infty.$$  \hspace{1cm} (23)

Then $\{R_n(s)(1 + |s|)^{\gamma/2}, \ s \in \mathbb{R}\}$ converges weakly to a tight Gaussian process.

**Proof.** Let $q = 2 + \gamma$. Since $\varepsilon_k \in L^q$, we have $Y_1 \in L^q$ and $\|Y_n - Y_n^\ast\|_q = \|a_n(\varepsilon_0 - \varepsilon_0')\|_q = O(|a_n|)$. Note that $f(x|y) = f_\varepsilon(x - y)$. Since $\int_{\mathbb{R}} f_\varepsilon(\theta|w_\gamma(\theta) = E(1 + |\varepsilon_1|)\gamma/2 < \infty$ and sup$_u f_\varepsilon(u) < \infty$, we have $\int_{\mathbb{R}} f_\varepsilon(u)^2 w_\gamma(du) < \infty$. Since $1 + |v + y| \leq (1 + |v|)(1 + |y|)$,

$$\int_{\mathbb{R}} |f_\varepsilon(\theta - y)|^2 + |f'_\varepsilon(\theta - y)|^2 + |f''_\varepsilon(\theta - y)|^2 w_\gamma(\theta) \leq (1 + |y|)^\gamma \int_{\mathbb{R}} |f_\varepsilon(\theta)|^2 + |f'_\varepsilon(\theta)|^2 + |f''_\varepsilon(\theta)|^2 w_\gamma(\theta) \leq C(1 + |y|)^\gamma.$$

On the other hand,

$$\int_{\mathbb{R}} |f''_\varepsilon(\theta - y)|^2(1 + |\theta|)^{-\gamma}d\theta \leq \int_{\mathbb{R}} |f''_\varepsilon(\theta - y)|^2(1 + |\theta - y|)^{-\gamma}(1 + |y|)^\gamma d\theta = (1 + |y|)^\gamma \int_{\mathbb{R}} |f''_\varepsilon(u)|^2 w_{-\gamma}(du).$$

By Theorems 1 and 3, the corollary follows.
Condition (22), together with \( \varepsilon_1 \in L^2 \), implies that the covariances of \( (X_i) \) are absolutely summable, a well-known condition for a linear process being short-range dependent. If \( \varepsilon_1 \) is violated, then we say that the linear process \( (X_n) \) is long-range dependent. In this case properties of \( F_n \) have been discussed by Ho and Hsing (1996) and Wu (2003).

**Remark 2.** Strong mixing properties of linear processes have been discussed by Withers (1981) and Pham and Tran (1983), among others. It seems that for linear processes strong mixing conditions do not lead to results with best possible conditions. Let \( \varepsilon_i \in L^2 \) have a density and \( A_n = \sum_{i=n}^{\infty} a_i^2 \). Withers (1981) showed that \( (X_i) \) is strongly mixing if \( \sum_{i=1}^{\infty} \max(A_i^{1/\beta}, \sqrt{\delta A_i |\log A_i|}) < \infty \). In the case that \( a_n = n^{-\delta} \), the latter condition requires \( \delta > 2 \). In comparison, our condition (22) only requires that \( \delta > 1 \). More stringent conditions are required for \( \beta \)-mixing.

**Remark 3.** Let \( \phi(t) = \mathbb{E}\exp(\sqrt{-1}t^\varepsilon_0) \); let \( f_k \) (resp. \( \phi_k \)) be the density (resp. characteristic) function of \( \sum_{i=0}^{k} \varepsilon_i \). Doukhan (2003) mentioned the following result (see also Doukhan and Surgailis (1998)): \( R_n \) converges weakly if for some \( 0 < \alpha \leq 1 \) and \( s, C, \delta > 0 \) such that \( s\delta > 2\alpha \), \( \varepsilon_k \in L^s \), \( \sum_{i=0}^{\infty} |a_i|^{\alpha} < \infty \) and \( |\phi(t)| \leq C/(1 + |u|^{\beta}) \). (Doukhan noted that the proof of the latter result is unpublished.) These results allow heavy-tailed \( \varepsilon_k \). Our Corollary 2 deals with weighted empirical processes, so there are different ranges of applications. Consider the special case of Corollary 2 in which \( \gamma = 0 \). A careful check of the proofs of Theorems 1 and 3 suggests that Corollary 2 still holds if \( f_k \) in (23) is replaced by \( f_k \) for some fixed \( k \) (see also Wu (2003)). Note that, by Parseval’s identity, if \( \gamma = 0 \), then \( 2\pi \int_{\mathbb{R}} |f_k''(u)|^2 du = \int_{\mathbb{R}} |t\phi_k(t)|^2 dt \), which is finite if \( |\phi_k(t)| \leq C/(1 + t^2) \). The latter condition holds if \( |\phi(t)| \leq C/(1 + |u|^{\beta}) \) and \( \# \{i : a_i \neq 0\} = \infty \).

**Remark 4.** Let \( g \) be a Lipschitz continuous function such that, for some \( r < 1 \), \( |g(x) - g(x')| \leq r|x - x'| \) holds for all \( x, x' \in \mathbb{R} \). Consider the nonlinear model \( X_{n+1} = g(X_n) + \varepsilon_{n+1} \), where \( \varepsilon_n \) satisfy conditions in Corollary 2. Then \( X_n \) is GMC(\( q \)) and consequently \( Y_n = g(X_n) \) is also GMC(\( q \)), \( q = 2 + \gamma \). It is easily seen that the argument in Corollary 2 also applies to this model and that the same conclusion holds. Special cases include the threshold autoregressive model (Tong (1990)) \( X_{n+1} = a \max(X_n, 0) + b \min(X_n, 0) + \varepsilon_{n+1} \) with \( \max(|a|, |b|) < 1 \), and the exponential autoregressive model (Haggan and Ozaki (1981)) \( X_{n+1} = [a + b \exp(-cX_n^2)]X_n + \varepsilon_{n+1} \), where \( |a| + |b| < 1 \) and \( c > 0 \).

4. **Inequalities**

The inequalities presented in this section are of independent interest and...
may have wider applicability. They are used in the proofs of the results in other sections.

**Lemma 1.** Let $H$ be absolutely continuous. (i) If $\mu \leq 1$ and $\gamma \in \mathbb{R}$, then

$$
\sup_{x \geq M} H^2(x)(1 + |x|)^{\gamma} \leq C_{\gamma, \mu} \int_{M}^{\infty} H^2(u) w_{\gamma-\mu}(du) + C_{\gamma, \mu} \int_{M}^{\infty} [H'(u)]^2 w_{\gamma+\mu}(du) 
$$

(24)

holds for all $M \geq 0$, where $C_{\gamma, \mu}$ is a positive constant. The above inequality also holds if $M = -\infty$ and $\sup_{x \geq M}$ is replaced by $\sup_{x \in \mathbb{R}}$. (ii) If $\gamma > 0$, $\mu = 1$, and $H(0) = 0$, then

$$
\int_{\mathbb{R}} H^2(u) w_{\gamma-1}(du) \leq \frac{4}{\gamma^2} \int_{\mathbb{R}} [H'(u)]^2 w_{\gamma+1}(du). 
$$

(26)

(iii) If $\gamma > 0$ and $H(\pm \infty) = 0$, then $\sup_{x \in \mathbb{R}} [H^2(x)(1 + |x|)^{\gamma}] \leq \gamma^{-1} \int_{\mathbb{R}} [H'(u)]^2 w_{\gamma+1}(du)$ and $\int_{\mathbb{R}} H^2(u) w_{\gamma-1}(du) \leq 4 \gamma^{-2} \int_{\mathbb{R}} [H'(u)]^2 w_{\gamma+1}(du)$.

**Proof.** (i) By Lemma 4 in [Wu (2003)], for $t \in \mathbb{R}$ and $\delta > 0$ we have

$$
\sup_{t \leq s \leq t + \delta} H^2(s) \leq \frac{2}{\delta} \int_{t}^{t+\delta} H^2(u)du + 2 \delta \int_{t}^{t+\delta} |H'(u)|^2 du. 
$$

(27)

We first consider the case $\mu < 1$. Let $\alpha = 1/(1 - \mu)$. In (24) let $t = t_n = n^\alpha$ and $\delta_n = (n + 1)^{\alpha} - n^\alpha$, $n \in \mathbb{N}$, and $I_n = [t_n, t_{n+1}]$. Since $\lim_{n \to \infty} \delta_n/ (n^{\alpha-1}) = 1$,

$$
\sup_{x \in I_n} [H^2(x)(1 + |x|)^{\gamma}] \leq 2 \sup_{x \in I_n} (1 + x)^{\gamma} \left[ \delta_n^{-1} \int_{I_n} H^2(u)du + \delta_n \int_{I_n} |H'(u)|^2 du \right] 
$$

$$
\leq C \int_{I_n} H^2(u) w_{\gamma-\mu}(du) + C \int_{I_n} |H'(u)|^2 w_{\gamma+\mu}(du). 
$$

(28)

It is easily seen by (24) that (28) also holds for $n = 0$ by choosing a suitable $C$. By summing (28) over $n = 0, 1, \ldots$, we obtain (24) with $M = 0$. The case $M > 0$ can be similarly dealt with by letting $t_n = n^\alpha + M$.

If $\mu = 1$, we let $t_n = 2^n$, $\delta_n = t_{n+1} - t_n = t_n$ and $I_n = [t_n, t_{n+1}]$, $n = 0, 1, \ldots$. The argument above yields the desired inequality.

(ii) Let $s \geq 0$. Since $H(s) = \int_{0}^{s} H'(u)du$, by the Cauchy-Schwarz Inequality,

$$
H^2(s) \leq \int_{0}^{s} |H'(u)|^2 (1 + u)^{1-\gamma} du \times \int_{0}^{s} (1 + u)^{\gamma-1} du 
$$

$$
\leq \int_{\mathbb{R}} |H'(u)|^2 w_{\gamma-1}(du) \times \frac{(1 + s)^{\gamma - 1}}{\gamma}. 
$$
So (25) follows. Applying Theorem 1.14 in Opic and Kufner (1990, p. 13) with \( p = q = 2 \), the Hardy-type inequality (26) easily follows. The proof of (iii) is similar to that of (ii).

**Lemma 2.** Let \( m \) be a measure on \( \mathbb{R} \), \( A \subset \mathbb{R} \) be a measurable set, and \( T_n(\theta) = \sum_{i=1}^{n} h(\theta, \xi_i) \), where \( h \) is a measurable function. Then

\[
\sqrt{\int_{A} \|T_n(\theta) - \mathbb{E}[T_n(\theta)]\|^2 m(d\theta)} \leq \sqrt{n} \sum_{j=0}^{\infty} \sqrt{\int_{A} \|P_{j}h(\theta, \xi_j)\|^2 m(d\theta)}.
\]

(29)

**Proof.** For \( j = 0, 1, \ldots \) let \( T_{n,j}(\theta) = \sum_{i=1}^{n} \mathbb{E}[h(\theta, \xi_i)|\xi_{i-j}] \) and \( \lambda_j^2 = \int_{A} \|P_{j}h(\theta, \xi_j)\|^2 m(d\theta) \), \( \lambda_j \geq 0 \). By the orthogonality of \( \mathbb{E}[h(\theta, \xi_i)|\xi_{i-j}] - \mathbb{E}[h(\theta, \xi_i)|\xi_{i-j-1}] \), \( i = 1, 2, \ldots, n \),

\[
\int_{A} \|T_{n,j}(\theta) - T_{n,j-1}(\theta)\|^2 m(d\theta) = n \int_{A} \|\mathbb{E}[h(\theta, \xi_1)|\xi_{j-1}] - \mathbb{E}[h(\theta, \xi_1)|\xi_{j-2}]\|^2 m(d\theta)
\]

\[= n \int_{A} \|P_{j-1}h(\theta, \xi_1)\|^2 m(d\theta) = n \lambda_j^2.
\]

Note that \( T_n(\theta) = T_{n,0}(\theta) \). Let \( \Delta = \sum_{j=0}^{\infty} \lambda_j \). By the Cauchy-Schwarz Inequality,

\[
\int_{A} \mathbb{E}[T_n(\theta) - \mathbb{E}[T_n(\theta)]]^2 m(d\theta) = \int_{A} \mathbb{E}\left\{ \sum_{j=0}^{\infty} |T_{n,j}(\theta) - T_{n,j+1}(\theta)| \right\}^2 m(d\theta)
\]

\[\leq \Delta \int_{A} \mathbb{E}\left\{ \sum_{j=0}^{\infty} \frac{|T_{n,j}(\theta) - T_{n,j+1}(\theta)|^2}{\lambda_j} \right\} m(d\theta) = n \Delta^2,
\]

and (29) follows.

**Lemma 3.** Let \( D_i \) be \( L^q \) \( (q > 1) \) martingale differences and \( C_q = 18q^{3/2}(q - 1)^{-1/2} \). Then

\[
\|D_1 + \cdots + D_n\|_q^r \leq C_q^r \sum_{i=1}^{n} \|D_i\|_q^r, \quad \text{where} \quad r = \min(q, 2).
\]

(30)

**Proof.** Let \( M = \sum_{i=1}^{n} D_i^2 \). By Burkholder’s inequality, \( \| \sum_{i=1}^{n} D_i \|_q \leq C_q \|M\|_{q/2} \). Then (30) easily follows by considering the cases \( q > 2 \) and \( q \leq 2 \) separately.

**Lemma 4.** (Wang (2003)) Let \( q > 1 \) and \( Z_i, 1 \leq i \leq 2^d \), be random variables in \( L^q \), where \( d \) is a positive integer. Let \( S_n = Z_1 + \cdots + Z_n \) and \( S_n^* = \max_{1 \leq \ell \leq n} |S_\ell| \). Then

\[
\|S_{2^d}\|_q \leq \sum_{r=0}^{d} \sum_{m=1}^{2^d-1} \|S_{2^r m} - S_{2^r (m-1)}\|_{q^*}^{1/r},
\]

(31)
5. Proofs of Theorems 1 and 2

To illustrate the idea behind our approach, let $\hat{F}_n(x) = n^{-1} \sum_{i=1}^{n} F_\varepsilon(x|\xi_{i-1})$ be the conditional empirical distribution function and write

$$F_n(x) - F(x) = [F_n(x) - \hat{F}_n(x)] + [\hat{F}_n(x) - F(x)].$$

The decomposition (32) has two important and useful properties. First, $F_n(x) - \hat{F}_n(x)$ is a martingale with stationary, ergodic and bounded martingale differences. Second, $\hat{F}_n - F$ is differentiable with derivative $\hat{f}_n(x) - f(x)$, where $\hat{f}_n(x) = (\partial/\partial x)\hat{F}_n(x) = n^{-1} \sum_{i=1}^{n} f_\varepsilon(x|\xi_i)$. The property of differentiability is useful in establishing tightness. The idea was applied in [Wu and Mićnićžuk 2002].

Following (32), let $G_n(s) = n^{1/2}[F_n(x) - \hat{F}_n(x)]$ and $Q_n(s) = n^{1/2}[\hat{F}_n(x) - F(x)]$. Then $R_n(s) = G_n(s) + Q_n(s)$. Sections 5.1 and 5.2 deal with $G_n$ and $Q_n$, respectively. Theorems 1 and 2 are proved in Sections 5.3 and 5.4.

5.1. Analysis of $G_n$

The main result is this section is Lemma 7 which concerns the weak convergence of $G_n$.

**Lemma 5.** Let $q > 2$ and $\alpha = \max(1, q/4) - q/2$. Then there is a constant $C_q$ such that

$$\|G_n(y) - G_n(x)\|_q^q \leq C_q n^\alpha |F(y) - F(x)| + C_q (y - x)^{\frac{q}{2} - 1} \int_x^y \mathbb{E}[f_\varepsilon^2(u|\xi_0)] du$$

holds for all $n \in \mathbb{N}$ and all $x < y$, and

$$\|G_n(x)\|_q^q \leq C_q \min\{n \mathbb{E}[F(x), 1 - F(x)].$$

**Proof.** Let $q' = q/2$ and $p' = q'/(q' - 1)$; let $d_i(s) = 1_{X_i \leq s} - \mathbb{E}(1_{X_i \leq s}|\xi_{i-1})$, $d_i = d_i(y) - d_i(x)$, $D_i = d_i^2 - \mathbb{E}(d_i^2|\xi_{i-1})$, $K_n = \sum_{i=1}^{n} D_i$, and $L_n = \sum_{i=1}^{n} \mathbb{E}(d_i^2|\xi_{i-1})$. Then both $(D_i)$ and $(d_i)$ are martingale differences. By Burkholder’s inequality ([Chow and Teicher 1988]),

$$\|G_n(y) - G_n(x)\|_q^q = n^{-\frac{q}{2}} \mathbb{E}(|d_1 + \ldots + d_n|)$$

$$\leq \frac{C_q}{n^{\frac{q}{2}}} \mathbb{E}[(d_1^2 + \ldots + d_n^2)^{\frac{q}{2}}] \leq \frac{C_q}{n^{\frac{q}{2}}} (\|K_n\|_{q'}^{q'} + \|L_n\|_{q'}^{q'}).$$

By Lemma 3,

$$\frac{\|K_n\|_{q'}^{q'}}{n^{\max(1, \frac{q}{2})}} \leq C_q \|D_1\|_{q'}^{q'} \leq C_q 2^{q' - 1} [\|d_1^2\|_{q'}^{q'} + \|\mathbb{E}(d_1^2|\xi_0)\|_{q'}^{q'}] \leq C_q 2^{q'} [d_1^2]\|_{q'}^{q'},$$

(36)
where we have applied Jensen’s inequality in \( \| E(d_1^2 | \xi_0) \|_{q'} \leq \| d_1^2 \|_{q'} \). Notice that \( |d_1| \leq 1 \),
\[
\| d_1^2 \|_{q'} \leq \| d_1 \|_{q'} \leq 2^{q'-1} (\| 1_{x \leq X_1 \leq y} \|_q + \| E(1_{x \leq X_1 \leq y} | \xi_0) \|_{q'}) \leq 2^{q'} [F(y) - F(x)].
\]
Since \( E(d_1^2 | \xi_0) \leq E(1_{x \leq X_1 \leq y} | \xi_0) \) and \( 1/p' + 1/q' = 1 \), we have by Hölder’s Inequality that
\[
\| L_n \|_{q'} \leq n^{q'} \| E(d_1^2 | \xi_0) \|_{q'} \leq n^{q'} \mathbb{E} \left\{ \left[ \int_x^y f_x(u | \xi_0) du \right]^{q'} \right\}
\]
\[
\leq n^{q'} \mathbb{E} \left[ (y-x)^{q'/p'} \int_x^y f_x(u | \xi_0) du \right].
\]
Combining (35), (36), (37) and (38), we have (33).

To show (34), in (35) we let
\[
\sum_{r=m}^{n} q \leq \sup_{1 \leq t \leq b} \mathbb{E} \left[ (y-x)^{q'/p'} \int_x^y f_x(u | \xi_0) du \right].
\]
Completes the proof.

**Lemma 6.** Let \( q > 2 \) and \( \alpha = \max(1, q/4) - q/2 \). Then there exists a constant \( C_q < \infty \) such that, for all \( b > 0, a \in \mathbb{R} \) and \( n, d \in \mathbb{N} \),
\[
\mathbb{E} \left[ \sup_{0 \leq s < b} | G_n(a+s) - G_n(a) |^q \right] \leq C_q n^{\alpha} [F(a+b) - F(a)] + C_q b^{\frac{2}{q} - 1} [1 + n^{\frac{1}{2}} 2^{d(1 - \frac{2}{q})}] \int_a^{a+b} \mathbb{E} [f_x \theta (u | \xi_0)] du. \quad (39)
\]
In particular, for \( d = 1 + [(\log n) / [(1 - 2/q) \log 2]] \), we have
\[
\mathbb{E} \left[ \sup_{0 \leq s < b} | G_n(a+s) - G_n(a) |^q \right] \leq C_q (\log n)^q n^{\alpha} [F(a+b) - F(a)] + C_q b^{\frac{2}{q} - 1} \int_a^{a+b} \mathbb{E} [f_x \theta (u | \xi_0)] du. \quad (40)
\]

**Proof.** Let \( h = b^{2-d} \), \( S_j = G_n(a+jh) - G_n(a) \) and \( Z_j = S_j - S_{j-1} \). By Lemma 5,
\[
\sum_{m=1}^{2^{d-1}} \| S_{2^m} - S_{2^m-1} \|_q^q \leq \sum_{m=1}^{2^{d-1}} C_q n^{\alpha} [F(a+2^mh) - F(a+2^m(m-1)h)]
\]
\[
+ \sum_{m=1}^{2^{d-1}} C_q (2^m h)^{\frac{2}{q} - 1} \int_a^{a+2^m(h-1)} \mathbb{E} [f_x \theta (u | \xi_0)] du
\]
\[
= C_q n^{\alpha} [F(a+b) - F(a)] + C_q (2^m h)^{\frac{2}{q} - 1} V,
\]
where \( V \) is a norm that satisfies some properties.
where \( V = \int_a^b E[f_i^{3/2}(u|\xi_0)] \, du \). By Lemma 4,

\[
\| S_{2d}^e \|_q \leq \sum_{r=0}^d \{ C_q n^r [F(a + b) - F(a)] \}^{\frac{1}{q}} + \sum_{r=0}^d \{ C_q (2^r h)^{q/2 - 1} V \}^{\frac{1}{q}}
\]

\[
\leq d \{ C_q n^r [F(a + b) - F(a)] \}^{\frac{1}{q}} + \{ C_q (2^d h)^{q/2 - 1} V \}^{\frac{1}{q}}.
\]

(41)

Let \( B_j = \sqrt{n} [\bar{F}_n(a + jh) - \bar{F}_n(a + (j-1)h)] \). Recall \( \bar{F}_n(x) = n^{-1} \sum_{i=1}^n F(x|\xi_i - 1) \) and \( \bar{f}_n(x) = \bar{F}_n'(x) \). Since \( q' = q/2 > 1 \), \( \| \bar{f}_n(x) \|_{q'} \leq \| f_e(x|\xi_0) \|_{q'} \). Note that

\[
E \left[ \max_{j \leq 2d} B_j^q \right] \leq \sum_{j=1}^{2d} E(B_j^q) = \sum_{j=1}^{2d} n^{q'j} \| \bar{F}_n(a + jh) - \bar{F}_n(a + (j-1)h) \|_{q'}^q
\]

\[
\leq \sum_{j=1}^{2d} n^{q'j} \| \bar{F}_n(a + jh) - \bar{F}_n(a + (j-1)h) \|_{q'}^q
\]

\[
\leq \sum_{j=1}^{2d} n^{q'j} h^{q'-1} \int_{a+(j-1)h}^{a+jh} E[|f_e(x)|_{q'}] \, du \leq n^{q'j} h^{q'-1} V.
\]

(42)

Observe that

\[
G_n(a + h \{ s/h \}) - \max_{j \leq 2d} B_j \leq G_n(a + s) \leq G_n(a + h \{ s/h + 1 \}) + \max_{j \leq 2d} B_j.
\]

Hence (39) follows from (11), (12) and, since \( S_j = G_n(a + jh) - G_n(a) \),

\[
\sup_{0 \leq s < b} |G_n(a + s) - G_n(a)| \leq \sup_{0 \leq s < b} |G_n(a + h \{ s/h + 1 \}) - G_n(a)|
\]

\[
+ \sup_{0 \leq s < b} |G_n(a + h \{ s/h \}) - G_n(a)| + 2 \max_{j \leq 2d} B_j
\]

\[
\leq 2 S_{2d}^e + 2 \max_{j \leq 2d} B_j
\]

by noticing that \( h = 2^{-d} b \). For \( d = 1 + [(\log n)/(1 - 2/q) \log 2] \), we have \( n^q/2d(1-q/2) \leq 1 \), hence (10) is an easy consequence of (39).

**Lemma 7.** Let \( \gamma \geq 0 \) and \( q > 2 \). Assume \( E[|X_1|^\gamma + \log(1 + |X_1|)] < \infty \), and (10). Then (i) \( E[\sup_{s \in R} |G_n(s)\|^q(1 + |s|)^\gamma] = \mathcal{O}(1) \), and (ii) the process \( \{ G_n(s)(1 + |s|)^\gamma, s \in R \} \) is tight.

**Remark 5.** In Lemma 7, the term \( \log(1 + |X_1|) \) is not needed if \( \gamma > 0 \).

**Proof.** (i) Without loss of generality we show that \( E[\sup_{s \geq 0} |G_n(s)|^q(1 + |s|)^\gamma] = \mathcal{O}(1) \), since the case of \( s < 0 \) follows similarly. Let \( \alpha_n = (\log n)^q n^{-\alpha} \).
By (6) and (10) of Lemma 6, with $a = b = 2^k$,

$$\sum_{k=1}^{\infty} 2^{k\gamma} \mathbb{E}\left[ \sup_{2^k \leq s < 2^{k+1}} |G_n(s) - G_n(2^k)|^q \right]$$

$$\leq C_\gamma \sum_{k=1}^{\infty} 2^{k\gamma} \alpha_n [F'(2^{k+1}) - F(2^k)] + C_\gamma \sum_{k=1}^{\infty} 2^{k\gamma} (2^k)^{\frac{2}{1}} - 1 \int_{2^k}^{2^{k+1}} \mathbb{E}[f_{\gamma}^2(u|\xi_0)]du$$

$$\leq C_{\gamma,q} \alpha_n \int_{2}^{\infty} f(u)(1 + u)^\gamma du + C_{\gamma,q} \int_{2}^{\infty} (1 + u)^\gamma u^{\frac{2}{1}} - 1 \mathbb{E}[f_{\gamma}^2(u|\xi_0)]du$$

$$\leq C_{\gamma,q} \alpha_n + C_{\gamma,q} = \mathcal{O}(1). \quad (43)$$

Observe that the function $\ell(x) = \sum_{k=1}^{\infty} 2^{k\gamma} 1_{2^k \leq x}$, $x > 0$, is bounded by $C_\gamma [x^\gamma + \log(1 + x)]$, where $C_\gamma$ is a constant. Then by (6) of Lemma 5, we have

$$\sum_{k=1}^{\infty} 2^{k\gamma} \|G_n(2^k)\|^q \leq \sum_{k=1}^{\infty} 2^{k\gamma} \mathbb{E}[\mathbf{1}_{2^k \leq X_1}] \leq C \mathbb{E}[|X_1|^\gamma + \log(1 + |X_1|)] < \infty. \quad (44)$$

Simple calculations show that (i) follows from (6), (6) and (10), with $a = 0$ and $b = 2$.

(ii) It is easily seen that the argument in (i) entails

$$\lim_{r \to \infty} \limsup_{n \to \infty} \mathbb{E}\left[ \sup_{|s| > r} |G_n(s)|^q (1 + |s|)^\gamma \right] = 0. \quad (45)$$

Let $\delta \in (0, 1)$. For $s, t \in [-r, r]$ with $0 \leq s - t \leq \delta$, we have

$$|G_n(s)(1 + |s|)^{\frac{2}{1}} - G_n(t)(1 + |t|)^{\frac{2}{1}}|$$

$$\leq |(1 + |s|)^{\frac{2}{1}} [G_n(s) - G_n(t)]| + |G_n(t)[(1 + |s|)^{\frac{2}{1}} - (1 + |t|)^{\frac{2}{1}}]|$$

$$\leq (1 + r)^{\frac{2}{1}} |G_n(s) - G_n(t)| + C_{r,\gamma,q} \delta \sup_{u \in [-r, r]} |G_n(u)|. \quad (46)$$

By (i), $\| \sup_{u \in \mathbb{R}} |G_n(u)| \| = \mathcal{O}(1)$. Let $I_k = I_k(\delta) = [k\delta, (k + 1)\delta]$. By Lemma 6,

$$\sum_{k=\left\lfloor \frac{r}{\delta} \right\rfloor - 1}^{\left\lfloor \frac{r}{\delta} \right\rfloor + 1} \mathbb{P}\left[ \sup_{s \in I_k} |G_n(s) - G_n(k\delta)| > \epsilon \right]$$

$$\leq \epsilon^{-q} \sum_{k=\left\lfloor \frac{r}{\delta} \right\rfloor - 1}^{\left\lfloor \frac{r}{\delta} \right\rfloor + 1} \left\{ C_q \alpha_n \mathbb{P}(X_1 \in I_k) + C_q \delta^{\frac{2}{1}} - 1 \int_{I_k} \mathbb{E}[f_{\gamma}^2(u|\xi_0)]du \right\}$$

$$\leq \epsilon^{-q} C_q \alpha_n + \epsilon^{-q} C_q \delta^{\frac{2}{1}} - 1 \int_{\mathbb{R}} \mathbb{E}[f_{\gamma}^2(u|\xi_0)]du.$$
By (3), \( \int_R E[f_{\xi}^q/2(u|\xi_0)]du < \infty \). Hence

\[
\limsup_{n \to \infty} \mathbb{P} \left[ \sup_{s,t \in [-r,r], 0 \leq s-t \leq \delta} \left| G_n(s) - G_n(t) \right| > 2\epsilon \right] \leq \epsilon^{-q}Cq\delta^{q-1},
\]

which implies the tightness of \( \{ G_n(s), |s| \leq r \} \) for fixed \( r \). So (10) and (11) entail (ii).

### 5.2. Analysis of \( Q_n \)

It is relatively easier to handle \( Q_n \) since it is a differentiable function. The Hardy-type inequalities (cf Lemma 1) are applicable.

**Lemma 8.** Let \( \gamma' \geq 0 \) and assume (7). Then (i) \( E[\sup_{s \in \mathbb{R}} |Q_n(s)|^2 (1 + |s|)^{\gamma'/2}, s \in \mathbb{R} \} \) is tight.

**Proof.** Let \( r \geq 0 \) and recall \( 0 \leq \mu \leq 1 \). (i) By Lemma 1,

\[
\Lambda_r := \sup_{|s| \geq r} [Q_n^2(s)(1 + |s|)^{\gamma'}] \leq C \int_{|s| \geq r} Q_n^2(s) w_{\gamma'-\mu} \, ds + C \int_{|s| \geq r} [Q_n(s)]^2 w_{\gamma+\mu} \, ds
\]

holds for some constant \( C = C_{\gamma', \mu} \). By Lemma 1,

\[
\frac{\| \Lambda_r \|}{\sqrt{C}} \leq \sum_{j=0}^{\infty} \sqrt{\int_{|s| \geq r} \| P_0 F_{\varepsilon}^j (|\xi_j|) \|^2 w_{\gamma'-\mu} \, d\theta} + \sum_{j=0}^{\infty} \sqrt{\int_{|s| \geq r} \| P_0 F_{\varepsilon}^j (|\xi_j|) \|^2 w_{\gamma+\mu} \, d\theta}.
\]

So (i) follows by letting \( r = 0 \) in (47).

(ii) The argument in (ii) of Lemma 7 in applicable. Let \( 0 < \delta < 1 \). Then

\[
\Psi_{n, \delta}(s,t) := \sup_{s,t \in [-r, r], 0 \leq s-t \leq \delta} \left| Q_n(s)(1 + |s|)^{\gamma'/2} - Q_n(t)(1 + |t|)^{\gamma'/2} \right|
\]

\[
\leq \sup_{s,t \in [-r, r], 0 \leq s-t \leq \delta} \left| (1 + |s|)^{\gamma'/2} [Q_n(s) - Q_n(t)] \right|
\]

\[
+ \sup_{s,t \in [-r, r], 0 \leq s-t \leq \delta} \left| Q_n(t)[(1 + |s|)^{\gamma'/2} - (1 + |t|)^{\gamma'/2}] \right|
\]

\[
\leq C_{r, \gamma', \mu} \sup_{u \in [-r, r]} \left| Q_n(u) \right| + C_{r, \gamma', \mu} \sup_{u \in [-r, r]} \left| Q_n(u) \right|.
\]

By (27), Lemma 2, and (17), there exists a constant \( C = C(r, \gamma', \mu, \nu) \) such that

\[
E \left[ \sup_{|s| \leq r} \left| Q_n'(s) \right|^2 \right] \leq C \int_{-r}^{r} \left\| \frac{d}{ds} Q_n(s) \right\|^2 w_{\gamma'+\mu} \, ds + C \int_{-r}^{r} \left\| Q_n''(s) \right\|^2 w_{\gamma'+\mu} \, ds
\]

\[
\leq C_{\gamma'} \sigma^2(f_{\varepsilon}, w_{\gamma'+\mu}) + C_{\gamma'} \sigma^2(f_{\varepsilon}, w_{\gamma'}) = O(1).
\]
By (i), there exists \( C_1 < \infty \) such that for all \( n \in \mathbb{N} \), \( \mathbb{E}[\Psi^2_{n,r}(\delta)] \leq \delta^2 C_1 \). Notice that the upper bound in (17) goes to 0 as \( r \to \infty \). Hence (ii) obtains.

5.3. Proof of Theorem 1.

We need to verify finite-dimensional convergence and tightness. Let \( j \geq 0 \). Observe that \( (\partial / \partial \theta) \mathcal{P}_0 f_\varepsilon(\theta | \xi_j) = \mathcal{P}_0 f_\varepsilon(\theta | \xi_j) \) and \( \| \mathcal{P}_0 f_\varepsilon(\theta | \xi_j) \| = \| \mathcal{P}_0 \mathbf{1}_{X_{j+1} \leq \theta} \| \).

By Lemma 1(i)

\[
\sup_{\delta \in \mathbb{R}} \| \mathcal{P}_0 f_\varepsilon(\theta | \xi_j) \| \leq C \sqrt{\int_{\mathbb{R}} \| \mathcal{P}_0 f_\varepsilon(\theta, \xi_j) \|^2 w_\mu(d\theta)} + C \sqrt{\int_{\mathbb{R}} \| \mathcal{P}_0 f_\varepsilon(\theta, \xi_j) \|^2 w_\mu(d\theta)}
\]

which, by (7) and (5), implies \( \sum_{i=0}^{\infty} \| \mathcal{P}_0 \mathbf{1}_{X_i \leq \theta} \| < \infty \). Hence by Theorem 1(i) in Hannan (1973), \( R_n(\theta) \) is asymptotically normal. The finite-dimensional convergence easily follows. Since \( R_n(s) = G_n(s) + Q_n(s) \), the tightness and (i) follow from Lemmas 7 and 8.

Since \( \mathbb{E}[f_\varepsilon(u | \xi_0)] = f(u) \), (5) and the moment condition \( X_1 \in \mathcal{L}^{\gamma+q/2-1} \) imply (6).

5.4. Proof of Theorem 2.

Let \( \Theta_n(a, \delta) = \sup_{0 \leq s < \delta} |G_n(a + s) - G_n(a)| \) and \( \alpha = \max(1, q/4) - q/2 \). Note that \( \log n)^{q_n \alpha} = \mathcal{O}(\delta^{q/2-1}) \). By (10) of Lemma 6, we have, uniformly in \( a \),

\[
\mathbb{E}\left[ \Theta_n^2(a, \delta_n) \right] \leq C_q \log n)^{q_n \alpha} [F(a + \delta_n) - F(a)] + C_q \delta_n^{q-1} \frac{q}{2} - 1 \int_{a}^{a+\delta_n} f(u)du
\]

\[
\leq C \delta_n^{q-1} [F(a + \delta_n) - F(a)].
\]

Here the constant \( C \) only depends on \( \tau, \gamma, q \) and \( \mathbb{E}(|X_1|^{\gamma}) \). Hence

\[
\sum_{k \in \mathbb{Z}} (1 + |k\delta_n|)^\gamma \mathbb{E}\left[ \Theta_n^2(k\delta_n, \delta_n) \right] \leq \sum_{k \in \mathbb{Z}} (1 + |k\delta_n|)^\gamma C \delta_n^{q-1} (F(k\delta_n + \delta_n) - F(k\delta_n))
\]

\[
\leq C \delta_n^{q-1} \mathbb{E}[(1 + |X_1|)^{\gamma}] .
\]

Let \( I_k(\delta) = [k\delta, (k+1)\delta] \) and \( c_\delta = \sup_{|u-t| \leq \delta} [(1 + |t|)/(1 + |u|)] \). Then \( c_\delta \leq 2 \) since \( \delta < 1/2 \). By the inequality \( |G_n(a) - G_n(c)| \leq |G_n(a) - G_n(b)| + |G_n(b) - G_n(c)| \),

\[
\sup_{t \in I_k(\delta_n), 0 \leq s < \delta_n} |G_n(t + s) - G_n(t)| \leq 2 \sup_{0 \leq u < 2\delta_n} |G_n(k\delta_n + u) - G_n(k\delta_n)|.
\]

Therefore,

\[
\mathbb{E}\left[ \sup_{t \in \mathbb{R}} (1 + |t|)^\gamma \Theta_n^q(t, \delta_n) \right] \leq \sum_{k \in \mathbb{Z}} \mathbb{E}\left[ \sup_{t \in I_k(\delta_n)} (1 + |t|)^\gamma \Theta_n^q(t, \delta_n) \right]
\]

\[
\leq 2^{q} c_\delta \sum_{k \in \mathbb{Z}} (1 + |k\delta_n|)^\gamma \mathbb{E}\left[ \Theta_n^q(k\delta_n, 2\delta_n) \right] \leq C \delta_n^{q-1} .
\]
Note that \( R_n(s) = G_n(s) + Q_n(s) \). Then (12) follows if it holds with \( R_n \) replaced by \( G_n \) and \( Q_n \), respectively. The former is an easy consequence of the preceding inequality and Jensen’s Inequality. To show that (12) holds with \( R_n \) replaced by \( Q_n \), recall \( \gamma' = 2\gamma/q \). By (24) of Lemma 1 and Lemma 2,

\[
\mathbb{E} \left[ \sup_{x \in \mathbb{R}} (1 + |x|)^{\gamma'} |Q'_n(x)|^2 \right] \leq C \int_\mathbb{R} |Q'_n(x)|^2 w_{\gamma'}(dx) + C \int_\mathbb{R} |Q''_n(x)|^2 w_{\gamma'}(dx)
\]

\[
\leq C \sigma^2(f, w_{\gamma'}) + C \sigma^2(f, w_{\gamma'}) < \infty,
\]

which completes the proof in view of the fact that

\[
(1 + |t|)^{\gamma'} \sup_{|s| \leq \delta_n} |Q_n(t + s) - Q_n(t)|^2 \leq \delta_n^2 (1 + |t|)^{\gamma'} \sup_{|s| \leq \delta_n} |Q'_n(t + s)|^2
\]

\[
\leq C \delta_n^2 \sup_{u - t \leq \delta_n} [(1 + |u|)^{\gamma'} |Q'_n(u)|^2].
\]

Remark 6. It is worthwhile to note that the modulus of continuity of \( G_n \) has the order \( \delta_n^{1-2/q} \), while that of \( Q_n \) has the higher order \( \delta_n \).

6. Proof of Theorem 3.

(i) If (17) holds, then \( \sigma(h, m) \leq 2E_{h,m} \). To prove (17), let \( Z_n(\theta) = h_n(\theta, \xi_0) - h_n(\theta, \xi_n) \), \( h(\theta, Y_n) = h(\theta, Y_n) - h(\theta, Y_n) = \int_{Y_n}^h \frac{\partial}{\partial y} h(\theta, y) dy \).

Let \( \lambda(y) = [H_{h,m}(y)]^{1/2} \) and \( U = \int_{Y_n}^\infty \lambda(y) dy \). By the Cauchy-Schwarz Inequality,

\[
\int_\mathbb{R} V_n^2(\theta) m(d\theta) \leq \int_\mathbb{R} \int_{Y_n}^\infty \frac{1}{\lambda(y)} \left[ \frac{\partial}{\partial y} h(\theta, y) \right]^2 dy \times \int_{Y_n}^\infty \lambda(y) dy \ m(d\theta) = U^2.
\]

Note that \( \mathbb{E}[h(\theta, Y_n)|\xi_\theta] = \mathbb{E}[h(\theta, Y_n^*)|\xi_\theta] \). By (14), \( \|Z_n(\theta)\| \leq 2\|P_0 h(\theta, Y_n)\| \leq 2\|V_n(\theta)\| \). So we have (17). (ii) Let \( \delta = q/2 - 1 \) and \( W = \int_{Y_n}^\infty w_\delta(dy) \). If \( q < 0 \), then \( \int_\mathbb{R} w_\delta(dy) = -4/q \) and \( |W| \leq \min(|Y_n - Y_n^*|, -4/q) \). If \( \delta > 0 \), by Hölder’s Inequality,

\[
\|W\| \leq \|(1 + |Y_n|)^\delta + (1 + |Y_n^*|)^\delta\|Y_n - Y_n^*\|
\]

\[
\leq \|(1 + |Y_n|)^\delta + (1 + |Y_n^*|)^\delta\|q\|Y_n - Y_n^*\| = O(\|Y_n - Y_n^*\|_q).
\]

For the case \(-1 < \delta \leq 0\), we need to prove the inequality \( \int_v^w w_\delta(dy) \leq 2^{-\delta} |u - v|^{1+\delta}/(1 + \delta) \). For the latter, it suffices to consider cases (a) \( u \geq v \geq 0 \), and (b) \( u \geq 0 \geq v \). For (a),

\[
\int_v^w w_\delta(dy) = \frac{(1 + u)^{1+\delta} - (1 + v)^{1+\delta}}{1 + \delta} \leq \frac{(u - v)^{1+\delta}}{1 + \delta}.
\]
For (b), let \( t = (u - v)/2 \). Then
\[
\int_v^u w_\delta(dy) = \frac{(1+u)^{1+\delta} - 1 + (1+|v|)^{1+\delta} - 1}{1+\delta} \leq \frac{2(1+t)^{1+\delta} - 2}{1+\delta} \leq \frac{2t^{1+\delta}}{1+\delta}.
\]
Therefore \( \| W \| = O(\| Y_n - Y_n^* \|_{\delta+1}) = O(\| Y_n - Y_n^* \|_q^{\delta+1} / q) \).

Acknowledgements

The author is grateful to three referees for their valuable comments. The author also thanks Professors Sándor Csörgő and Jan Mielniczuk for useful suggestions. The work was supported in part by NSF grant DMS-0478704.

References


Department of Statistics, The University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, U.S.A.

E-mail: wbwu@galton.uchicago.edu

(Received September 2005; accepted May 2006)