Empirical process of long-range dependent sequences when parameters are estimated

Rafal Kulik

University of Sydney

Abstract: In this paper we study the asymptotic behaviour of empirical processes when parameters are estimated, assuming that the underlying sequence of random variables is long-range dependent. We show completely different phenomena compared to i.i.d. situation, as well as compared to ordinary empirical processes of long range dependent sequences. Applications include Kolmogorov-Smirnov and Cramer-Smirnov-von Mises goodness-of-fit statistics.

Key words and phrases: long range dependence, linear processes, goodness-of-fit .

1 Introduction and statement of results

Let $\{\epsilon_i, i \geq 1\}$ be a centered sequence of i.i.d. random variables. Consider the class of stationary linear processes

$$X_i = \sum_{k=0}^{\infty} c_k \epsilon_{i-k}, \quad i \geq 1. \tag{1}$$

We assume that the sequence c_k, $k \geq 0$, is regularly varying with index $-\beta$, $\beta \in (1/2, 1)$ (written as $c_k \in RV_{-\beta}$). This means that $c_k \sim k^{-\beta} L_0(k)$ as $k \to \infty$, where L_0 is a slowly varying function at infinity. We shall refer to all such models as long range dependent (LRD) linear processes. In particular, if the variance exists, then the covariances $\rho_k := \mathrm{E} X_0 X_k$ decay at the hyperbolic rate, $\rho_k = L(k) k^{-(2\beta-1)} =: L(k) k^{-D}$, where $\lim_{k\to\infty} L(k)/L_0^2(k) = B(2\beta - 1, 1-\beta)$ and $B(\cdot, \cdot)$ is the beta-function. Consequently, the covariances are not summable (cf. [10]).

Assume that X_1 has a continuous distribution function F. Given $X_1, \ldots, X_n$, let $F_n(x) = n^{-1} \sum_{i=1}^{n} 1_{\{X_i \leq x\}}$ be the empirical distribution function.

Assume that $E\epsilon_1^2 < \infty$. Let r be an integer and define

$$Y_{n,r} = \sum_{i=1}^{n} \sum_{1 \le j_1 < \cdots \le j_r} \prod_{s=1}^{r} c_{j_s} \epsilon_{i-j_s}, \qquad n \ge 1,$$

so that $Y_{n,0} = n$, and $Y_{n,1} = \sum_{i=1}^{n} X_i$. If $p < (2\beta - 1)^{-1}$, then

$$\sigma_{n,p}^2 := \mathrm{Var}(Y_{n,p}) \sim n^{2-p(2\beta-1)} L_0^{2p}(n). \tag{2}$$

From [12] we know that for $p < (2\beta - 1)^{-1}$, as $n \to \infty$,

$$\sigma_{n,p}^{-1} Y_{n,p} \xrightarrow{\mathrm{d}} Z_p, \tag{3}$$

where Z_p is a random variable which can be represented by appropriate multiple Wiener-Itô integrals. In particular, Z_1 is standard normal.

In the present paper we study the asymptotic behaviour of empirical processes when unknown parameters of the underlying distribution function are estimated. The motivation to study such problems comes from Kolmogorov-Smirnov type statistics. From [12] we know that, as $n \to \infty$,

$$\sigma_{n,1}^{-1} n \sup_{x \in \mathbf{R}} |F_n(x) - F(x)| \xrightarrow{\mathrm{d}} |Z_1| \sup_{x \in \mathbf{R}} f(x), \tag{4}$$

where Z_1 is a standard normal random variable and f is the density function of F. The above result can be used, in principle, to test whether data $X_1, \ldots, X_n$ are consistent with a given distribution F. If however F belongs to a one-parameter family $\{F(\cdot, \theta), \theta \in \mathbb{R}\}$ say, then in order to use (4) one needs to know the value of the parameter θ. A straightforward procedure would be to estimate it and use the statistic

$$\sigma_{n,1}^{-1} n \sup_{x \in \mathbf{R}} |F_n(x) - F(x; \hat{\theta}_n)|,$$

where $F(x; \hat{\theta}_n)$ is the distribution function $F(x) = F(x; \theta)$ in which the parameter θ has been replaced with its estimator $\hat{\theta}_n$. However, in the i.i.d. case, it is known that such procedure changes a limiting process. To be more specific, assume for a while that $X_1, \ldots, X_n$ are i.i.d. random variables and consider

$$\sqrt{n} \sup_{x \in \mathbf{R}} |F_n(x) - F(x)|.$$

As it is well-known, the above supremum converges in distribution to the supremum of a Brownian bridge on $[0, 1]$. On the other hand, for a large class of estimators,

$$\sqrt{n}|F_n(x) - F(x;\hat{\theta}_n)|,$$

converges weakly to a Gaussian process, but no longer to a Brownian bridge. The corresponding comments apply to the Cramér-Smirnov-von Mises statistic

$$\sqrt{n}\int_{\mathbf{R}}(F_n(x) - F(x))^2 dF(x)$$

and its 'estimated' version

$$\sqrt{n}\int_{\mathbf{R}}(F_n(x) - F(x;\hat{\theta}_n))^2 dF(x;\hat{\theta}_n).$$

We refer to [6], [9], [13] and [2] for more details.

Coming back to LRD sequences, the similar problems have been studied in [1] and [11] (See Remark 1.6 for the comparison of results in those papers and in the current one). Here, we will focus on a location family of distributions. We shall assume that $Y_i = X_i + \mu$, where X_i is given by (1). Clearly, if F is the distribution of X_1 and H is the distribution of Y_1, then $H(x) = F(x - \mu)$. Moreover, the empirical processes

$$\beta_n(x) = \sigma_{n,1}^{-1} n(F_n(x) - F(x)), \qquad x \in \mathbb{R}$$

and

$$\gamma_n(x) = \sigma_{n,1}^{-1} n(H_n(x) - H(x)), \qquad x \in \mathbb{R}$$

associated with X_i and Y_i, respectively, are related by

$$\gamma_n(x) = \beta_n(x - \mu). \tag{5}$$

From [12], $\beta_n(x) \Rightarrow f(x)Z_1$, so that $\gamma_n(x) \Rightarrow f(x-\mu)Z_1$. Here and in the sequel, $\Rightarrow$ denotes weak convergence in $D((-\infty,\infty))$. On the contrary, if $\hat{\theta}_n$ is an appropriate sequence of estimators of the mean μ, we will show that, as $n \to \infty$,

$$\hat{\gamma}_n(x) = \sigma_{n,1}^{-1} n(H_n(x) - H(x;\hat{\theta}_n)), \qquad x \in \mathbb{R}$$

converges in probability to 0. Choosing a different scaling one can obtain weak convergence, however the limiting process depends on the choice of the estimator. In particular, using $\hat{\theta}_n = \bar{Y}_n$ (the sample mean of $Y_1, \ldots, Y_n$) or $\hat{\theta}_n = M_n$ (M-estimator), we can obtain different limits, depending on the so-called *second-order*

M-rank of the estimator M_n introduced in [14]. Also, the scaling and the limiting process depend on whether $\beta > 3/4$ or $\beta < 3/4$. In particular, if $\beta > 3/4$, then we obtain $\sqrt{n}$-consistency of a modified Kolmogorov-Smirnov type statistics. The appropriate results are stated in Theorems 1.2 and 1.4.

The proofs of our results will be based on a reduction principle for long-range dependent empirical processes (see Theorem 1.1 below), combined with approximation method as in [2]. The fact, that we were able to use the latter, Hungarian-like approach, shows its extreme power. The Hungarian construction approach was for example employed to obtain the Komlós-Major-Tusnády (KMT) strong approximation of empirical processes. Then, this approach was followed to establish a number of optimal or almost optimal results for functionals of empirical and quantile processes, including the one in [2] for empirical processes with parameters estimated (we refer to [3]). The KMT construction is tailored for the i.i.d. situation. However, a lot of further developments based on this kind of approach, can be applied to long-range dependent sequences. Very recent examples of such an approach include [4], [5].

The reduction principle was obtained first in [7] in case of subordinated Gaussian processes. In more generality, it was obtained in the landmark paper [12]; see also [15] for related studies. The best available result along these lines is due to Wu [17]. To state a particular version of his result, we shall introduce the following assumptions, which will be valid throughout the paper. Let F_ϵ be the distribution function of the centered i.i.d. sequence $\{\epsilon_i, i \geq 1\}$. Assume that for a given integer p, the derivatives $F_\epsilon^{(1)}, \ldots, F_\epsilon^{(p+3)}$ of F_ϵ are bounded and integrable. Note that these properties are inherited by the distribution F as well (cf. [12] or [17]).

Theorem 1.1 *Let p be a positive integer. Then, as $n \to \infty$,*

$$\mathrm{E} \sup_{x \in \mathbf{R}} \left| \sum_{i=1}^{n} (1_{\{X_i \leq x\}} - F(x)) + \sum_{r=1}^{p} (-1)^{r-1} F^{(r)}(x) Y_{n,r} \right|^2 = O(\Xi_n + n(\log n)^2),$$

where

$$\Xi_n = \begin{cases} O(n), & (p+1)(2\beta-1) > 1 \\ O(n^{2-(p+1)(2\beta-1)} L_0^{2(p+1)}(n)), & (p+1)(2\beta-1) < 1 \end{cases}.$$

We will a require second-order expansion, thus in the above theorem, $p = 2$.

Let ψ be a real-valued function of bounded variation such that $\mathrm{E}\psi(Y_1 - \mu) = 0$. M-estimators are defined as

$$M = M_n = \arg\min \left\{ \left| \sum_{j=1}^{n} \psi(Y_j - x) \right|, x \in \mathbb{R} \right\}.$$

For $k = 1, 2$, let

$$\lambda_k = \int_{\mathbf{R}} \psi(y) f^{(k)}(y) dy.$$

Let $k^* = k^*(\beta) = [1/(2\beta - 1)]$, where $[\cdot]$ denotes the integer part. The second-order rank $r_M(2)$ of the M-estimator is: $r_M(2) = 2$ if $k^* = 1$ (so that $\beta > 3/4$); $r_M(2) = 2$ if $k^* > 1$ and $\lambda_2 \neq 0$; $r_M(2) > 2$ if $k^* > 1$ and $\lambda_2 = 0$. We refer to [14] for more details.

Let

$$a_n = \sigma_{n,2} \sigma_{n,1}^{-1}.$$

Now, we are ready to state our results. We start with the case $\beta < 3/4$.

Theorem 1.2 *Assume that $\theta_0 = \mu$ and $\beta < 3/4$. Then, under the conditions of* Theorem 1.1, *as $n \to \infty$, we have*

- *If $\hat{\theta}_n = \bar{Y}_n$ or $\hat{\theta}_n = M_n$, then*

$$\sup_{x \in \mathbf{R}} |\hat{\gamma}_n(x)| = o_P(1). \tag{6}$$

- *If $\hat{\theta}_n = \bar{Y}_n$, then*

$$a_n^{-1} \hat{\gamma}_n(x) = \sigma_{n,2}^{-1} n (H_n(x) - H(x; \hat{\theta}_n)) \Rightarrow f^{(1)}(x - \mu) V, \tag{7}$$

  where $V = Z_2 + \frac{1}{2} Z_1^2$.

- *If $\hat{\theta}_n = M_n$, $\mathrm{E}\epsilon_1^{4 \vee 2k^*(\theta)} < \infty$ and $r_M(2) > 2$, then* (7) *holds.*

- *If $\hat{\theta}_n = M_n$, $\mathrm{E}\epsilon_1^{4 \vee 2k^*(\theta)} < \infty$ and $r_M(2) = 2$*

$$a_n^{-1} \hat{\gamma}_n(x) = \sigma_{n,2}^{-1} n (H_n(x) - H(x; \hat{\theta}_n)) \Rightarrow f^{(1)}(x - \mu) V - \frac{\lambda_2}{2\lambda_1} \frac{1}{\sigma} f(x - \mu) V_1, \tag{8}$$

  where V is as in (7) *and $V_1 = Z_1^2 + Z_2$.*

Example 1.3 Assume that $\mu = 0$, f is symmetric and ψ is skew-symmetric. For $\beta < 3/4$, $r_M(2) \geq 3$ (cf. [14]) and the limiting behaviour is described by (7). If, however, f is not symmetric, then $\lambda_2 \neq 0$ and (8) holds.

As for the case $\beta > 3/4$ we have the following theorem.

Theorem 1.4 *Assume that $\theta_0 = \mu$ and $\beta > 3/4$. Then, under the conditions of* Theorem 1.1*, as $n \to \infty$, we have*

- *If $\hat{\theta}_n = \bar{Y}_n$ or $\hat{\theta}_n = M_n$, then*

$$\sup_{x \in \mathbf{R}} |\hat{\gamma}_n(x)| = o_P(1).$$

- *If $\hat{\theta}_n = \bar{Y}_n$, then*

$$\sqrt{n}\sigma_{n,1} n^{-1} \hat{\gamma}_n(x) = \sqrt{n}(H_n(x) - H(x; \hat{\theta}_n)) \Rightarrow W(x - \mu), \tag{9}$$

  where $W(\cdot)$ is a Gaussian process.

- *If $\hat{\theta}_n = M_n$, $\mathrm{E}\epsilon_1^{4 \vee 2k^*(\theta)} < \infty$, then*

$$\sqrt{n}\sigma_{n,1} n^{-1} \hat{\gamma}_n(x) = \sqrt{n}(H_n(x) - H(x; \hat{\theta}_n)) \Rightarrow W(x - \mu) + \frac{\sigma_\psi^2}{\sigma} f(x - \mu) Z_1, \tag{10}$$

  σ_ψ^2 is given by the formula (1.18) *in* [14].

An immediate corollary to Theorem 1.2 is the following Cramér-Smirnov-von Mises test. An appropriate version can also be stated in terms of Theorem 1.4.

Corollary 1.5 *Let $\theta_0 = \mu$ and $\hat{\theta}_n = \bar{Y}_n$. Under the conditions of* Theorem 1.2,

$$\sigma_{n,2}^{-1} n \int_{\mathbf{R}} (H_n(x) - H(x; \hat{\theta}_n))^2 dH(x; \hat{\theta}_n) \xrightarrow{\mathrm{d}} \frac{1}{\sigma} V^2 \int_{\mathbf{R}} \left(f^{(1)}(x - \mu) \right)^2 f(x - \mu)\, dx.$$

The above result should be compared with a regular situation of non-estimated Cramer-Smirnov-von Mises statistics in [8]. The limiting distribution for the model (1) in case of Gaussian errors ϵ_i, is a random variable Z_1^2 multiplied by a deterministic function.

Remark 1.6 The results established in this paper correspond to the previous research of Beran and Ghosh ([1]) and Ho ([11]). In the first paper it was assumed that ϵ_i, $i \geq 1$, are Gaussian. They considered the location-scale family corresponding to a transformation $Y_i = \sigma X_i + \mu$. If both σ and μ are estimated, then one can conclude from their Theorem 2 that for $\beta > 2/3$ (which corresponds to the Hurst parameter $1/2 < H < 5/6$), the estimated empirical process converges at rate $\sqrt{n}$. That result should be compared with our Theorem 1.4. Estimating μ only we obtain $\sqrt{n}$ rate of convergence for $\beta > 3/4$. Consequently, estimating both μ and σ, we obtain better rates than estimating μ only. In the Gaussian case, if μ is estimated by $\bar{Y}_n$ and σ is either known or estimated, then $\sum_{i=1}^n H_1((Y_i - \bar{Y}_n)/\sigma) = 0$ and the same holds if one replaces σ with its estimator. Thus, second-order Hermite's polynomials describe asymptotic behaviour of estimated empirical process: $\sum_{i=1}^n H_2(X_i)$ growths at rate $\sqrt{n}$ or $\sigma_{n,2}^{-1} n$ according to $\beta > 3/4$ or $\beta < 3/4$. If both μ and σ are estimated, then also $\sum_{i=1}^n H_2((X_i - \bar{X})/\hat{\sigma}_n) = 0$ ($\hat{\sigma}_n$ is a sample variance) and thus third-order Hermite polynomials play major role in the asymptotics. Consequently, $\sqrt{n}$-rate is achieved for $\beta > 2/3$. On the other hand, if μ is known and σ is estimated, then first-order Hermite polynomials do not vanish and the limiting behaviour of the estimated empirical process is the same as of the non-estimated one.

Ho, [11], considered the limiting behaviour of

$$\sum_{i=1}^n K(X_i, \hat{\theta}_n) \tag{11}$$

for a suitable class of functions K. In particular, if $K(x,y) = 1_{\{x<y\}}$, then from his Equation (4.1), applied with $J = 2$ and $f^{(1)}(\mu) \neq 0$, we obtain for $\beta < 3/4$,

$$\sigma_{n,2}^{-1} n (F_n(\bar{Y}_n) - F(\mu)) \xrightarrow{\mathrm{d}} f'(\mu)(Z_2 - Z^2/2).$$

If $f'(\mu) = 0$, then the limit and the scaling factor are different (see Example B in [11]). The situation is somehow parallel to the i.i.d case: depending on $f'(\mu)$ the limiting distribution of $F_n(\bar{Y}_n) - F(\mu)$ can be different, although the scaling remains the same (see [16]). Comparing this with our results, we see that in our case the limiting behaviour of estimated empirical processes and corresponding test statistics depend on a *global* reduction principle only, whereas for the limiting behaviour of (11), a *local* regularity properties of the density f are crucial. In a sense, the situation is comparable to kernel density estimation. The limiting behaviour of Parzen-Rosenblatt estimator at point x_0 depends on how many derivatives $f^{(r)}(x_0)$ vanish (see Theorem 3 in [18]).

Furthermore, Example C in [11] shows that the limiting behaviour of $\sum_{i=1}^n |Y_i - \bar{Y}_n|$ and $\sum_{i=1}^n |Y_i - M_n|$, where M_n is a M-estimator of μ, is the same. In case of estimated empirical processes, using M_n instead of $\bar{Y}_n$, we can change the limit.

In what follows C will denote a generic constant which may be different at each of its appearance. Also, for any sequences a_n and b_n, we write $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$. Moreover, $f^{(k)}$ denotes the kth order derivative of f.

2 Proofs

Let p be a positive integer. Recall that

$$a_n = \sigma_{n,2}\sigma_{n,1}^{-1}L_0(n),$$

and let

$$d_{n,p} = \begin{cases} n^{-(1-\beta)}L_0^{-1}(n)(\log n)^{5/2}(\log\log n)^{3/4}, & (p+1)(2\beta-1) > 1 \\ n^{-p(\beta-\frac{1}{2})}L_0^{p}(n)(\log n)^{1/2}(\log\log n)^{3/4}, & (p+1)(2\beta-1) < 1 \end{cases}$$

Note that $d_{n,2} = o(a_n)$ provided $\beta < \frac{3}{4}$,

Put

$$\begin{aligned} S_{n,p}(x) &= \sum_{i=1}^n (1_{\{X_i \le x\}} - F(x)) + \sum_{r=1}^p (-1)^{r-1}F^{(r)}(x)Y_{n,r} \\ &=: \sum_{i=1}^n (1_{\{X_i \le x\}} - F(x)) + V_{n,p}(x). \end{aligned}$$

Using Theorem 1.1 we obtain

$$\sigma_{n,p}^{-1}\sup_{x\in\mathbf{R}}|S_{n,p}(x)| =$$
$$\begin{cases} O_{a.s.}(n^{-(\frac{1}{2}-p(\beta-\frac{1}{2}))}L_0^{-p}(n)(\log n)^{5/2}(\log\log n)^{3/4}), & (p+1)(2\beta-1) > 1 \\ O_{a.s.}(n^{-(\beta-\frac{1}{2})}L_0(n)(\log n)^{1/2}(\log\log n)^{3/4}), & (p+1)(2\beta-1) < 1 \end{cases}.$$

Since (see (2))

$$\frac{\sigma_{n,p}}{\sigma_{n,1}} \sim n^{-(\beta-\frac{1}{2})(p-1)}L_0^{p-1}(n), \tag{12}$$

we obtain

$$\sup_{x\in\mathbf{R}} |\beta_n(x) + \sigma_{n,1}^{-1} V_{n,p}(x)| = \tag{13}$$
$$= \frac{\sigma_{n,p}}{\sigma_{n,1}} \sup_{x\in\mathbf{R}} \left| \sigma_{n,p}^{-1} \sum_{i=1}^{n} (1_{\{X_i \le x\}} - F(x)) + \sigma_{n,p}^{-1} V_{n,p}(x) \right| = o_{a.s.}(d_{n,p}).$$

For a function $g(x;\theta)$ denote by $\nabla_\theta^r g(x;\theta_0)$ its rth order derivative with respect to θ, evaluated at $\theta = \theta_0$. In particular, $\nabla = \nabla^1$.

Recall (5). For an arbitrary unknown parameter θ_0 and its estimator $\hat{\theta}_n$ we have by (13)

$$\begin{aligned}
\hat{\gamma}_n(x) &= \gamma_n(x) + \sigma_{n,1}^{-1} n (H(x;\theta_0) - H(x;\hat{\theta}_n)) \\
&= \beta_n\,(x-\mu) + \sigma_{n,1}^{-1} n (H(x;\theta_0) - H(x;\hat{\theta}_n)) \\
&= o_p(d_{n,2}) - \sigma_{n,1}^{-1} V_{n,2}\,(x-\mu) + \sigma_{n,1}^{-1} n (\theta_0 - \hat{\theta}_n) \nabla_\theta H(x;\theta_0) \\
&\quad + \frac{1}{2}\sigma_{n,1}^{-1} n (\theta_0 - \hat{\theta}_n)^2 \nabla_\theta^2 H(x;\theta_0) + \frac{1}{6}\sigma_{n,1}^{-1} n (\theta_0 - \hat{\theta}_n)^3 \nabla_\theta^3 H(x;\hat{\theta}_n^*) \\
&= o_p(d_{n,2}) - \sigma_{n,1}^{-1} f\,(x-\mu) \sum_{i=1}^{n} X_i + \sigma_{n,1}^{-1} f^{(1)}\,(x-\mu)\, Y_{n,2} \\
&\quad + \sigma_{n,1}^{-1} n (\theta_0 - \hat{\theta}_n) \nabla_\theta H(x;\theta_0) + \frac{1}{2}\sigma_{n,1}^{-1} n (\theta_0 - \hat{\theta}_n)^2 \nabla_\theta^2 H(x;\theta_0) \\
&\quad + \frac{1}{6}\sigma_{n,1}^{-1} n (\theta_0 - \hat{\theta}_n)^3 \nabla_\theta^3 H(x;\hat{\theta}_n^*),
\end{aligned} \tag{14}$$

with some $\hat{\theta}_n^*$ such that $|\hat{\theta}_n^* - \hat{\theta}_n| \le |\theta_0 - \hat{\theta}_n^*|$.

If $\theta_0 = \mu$, then

$$\nabla_\theta^r H(x) = \nabla_\mu^r F\,(x-\mu) = (-1)^r \frac{1}{\sigma^r} f^{(r-1)}\,(x-\mu)\,. \tag{15}$$

Also, if $\hat{\theta}_n = \bar{Y}_n$, then

$$\hat{\theta}_n - \theta_0 = \sigma \bar{X}_n \tag{16}$$

Hence, using uniform boundness of $f^{(2)}$,

$$\begin{aligned}
\hat{\gamma}_n(x) &= o_p(d_{n,2}) - \sigma_{n,1}^{-1} f\,(x-\mu) \sum_{i=1}^{n} X_i + \sigma_{n,1}^{-1} f^{(1)}\,(x-\mu)\, Y_{n,2} + \\
&\quad \sigma_{n,1}^{-1} f\,(x-\mu) \sum_{i=1}^{n} X_i + \frac{1}{2}\sigma_{n,1}^{-1} n f^{(1)}\,(x-\mu)\, \bar{X}_n^2 + O_P\left(\sigma_{n,1}^{-1} n \bar{X}_n^3\right).
\end{aligned}$$

Since $\beta < 3/4$, note that $\sigma_{n,1} Y_{n,2} = o_p(1)$ (cf. (3)), $\sigma_{n,1}^{-1} n \bar{X}_n^2 = o_P(1)$ and $\sigma_{n,1}^{-1} n \bar{X}_n^3 = o_P(1)$. Thus, we conclude that $\sup_x |\hat{\gamma}_n(x)| \xrightarrow{p} 0$ for $\hat{\theta}_n = \bar{Y}_n$.

Further,

$$\begin{aligned} & a_n^{-1} \sup_x \left| \hat{\gamma}_n(x) - f^{(1)}(x-\mu) \left[\sigma_{n,1}^{-1} Y_{n,2} + \frac{1}{2} \sigma_{n,1}^{-1} n \bar{X}_n^2 \right] \right| \\ &= o_p(d_{n,2} a_n^{-1}) + O_P(a_n^{-1} \sigma_{n,1}^{-1} n \bar{X}_n^3) = o_p(1) + O_P(a_n^{-1} \sigma_{n,1}^{-1} n n^{-3} \sigma_{n,1}^3) \\ &= o_P(1). \end{aligned}$$

Thus, (7) follows.

If $\hat{\theta}_n = M_n$ then, as in (14) and (15),

$$\begin{aligned} \hat{\gamma}_n(x) &= o_p(d_{n,2}) - \sigma_{n,1}^{-1} f(x-\mu) \sum_{i=1}^n X_i + \sigma_{n,1}^{-1} f^{(1)}(x-\mu) Y_{n,2} + \\ & \quad -\frac{1}{\sigma} \sigma_{n,1}^{-1} n(\mu - \bar{Y}_n) f(x-\mu) - \frac{1}{\sigma} \sigma_{n,1}^{-1} n(\bar{Y}_n - M_n) f(x-\mu) + \\ & \quad \frac{1}{2\sigma^2} \sigma_{n,1}^{-1} n f^{(1)}(x-\mu)(\mu - M_n)^2 + O_P(\sigma_{n,1}^{-1} n(\mu - M_n)^3) \\ &= o_p(d_{n,2}) + \sigma_{n,1}^{-1} f^{(1)}(x-\mu) Y_{n,2} - \frac{1}{\sigma} \sigma_{n,1}^{-1} n(\bar{Y}_n - M_n) f(x-\mu) \\ & \quad + \frac{1}{2\sigma^2} \sigma_{n,1}^{-1} n f^{(1)}(x-\mu)(\mu - M_n)^2 + O_P(\sigma_{n,1}^{-1} n(\mu - M_n)^3). \end{aligned}$$

From [14],

$$\sigma_{n,1}^{-1} n(M_n - \mu) = \sigma_{n,1}^{-1} n(\bar{Y}_n - \mu) + o_P(1) \xrightarrow{\mathrm{d}} \sigma^2 Z_1 \tag{17}$$

and $\sigma_{n,1}^{-1} n(\bar{Y}_n - M_n) = o_P(1)$. Thus, $\sup_x |\hat{\gamma}_n(x)| \xrightarrow{p} 0$ for $\hat{\theta}_n = M_n$.

If $r_M(2) > 2$, then from [14, Theorem 1.1],

$$a_n^{-1} \sigma_{n,1}^{-1} n(\bar{Y}_n - M_n) = o_P(1),$$

thus in this case

$$\begin{aligned} & a_n^{-1} \sup_x \left| \hat{\gamma}_n(x) - f^{(1)}(x-\mu) \left[\sigma_{n,1}^{-1} Y_{n,2} + \frac{1}{2\sigma^2} \sigma_{n,1}^{-1} n(\mu - M_n)^2 \right] \right| \\ &= o_p(d_{n,2} a_n^{-1}) + o_P(1) + O_P(a_n^{-1} \sigma_{n,1}^{-1} n(\mu - M_n)^3) = o_P(1). \end{aligned}$$

Therefore, in view of (17), (7) follows.

If $r_M(2) = 2$, then $a_n^{-1}\sigma_{n,1}^{-1}n$ is the proper scaling for $(\bar{Y}_n - M_n)$ and thus

$$\begin{aligned} a_n^{-1} \sup_x \Bigg| \hat{\gamma}_n(x) - f^{(1)}(x-\mu) \left[\sigma_{n,1}^{-1} Y_{n,2} + \frac{n(\mu - M_n)^2}{2\sigma^2 \sigma_{n,1}} \right] \\ + \frac{n}{\sigma \sigma_{n,1}} f(x-\mu)(\bar{Y}_n - M_n) \Bigg| \\ = \quad o_p(d_{n,2} a_n^{-1}) + O_P(a_n^{-1}\sigma_{n,1}^{-1} n(\mu - M_n)^3) = o_P(1), \end{aligned}$$

and hence (8) follows using (17) and Corollary 1.1 in [14].

$\odot$

3 Proof of Corollary 1.5

Write

$$\begin{aligned} \int \hat{\gamma}_n(x)^2 dH(x;\hat{\theta}_n) = \int \hat{\gamma}_n(x)^2 h(x;\theta_0) dx \\ + \int \hat{\gamma}_n(x)^2 (h(x;\hat{\theta}_n - h(x;\theta_0)) dx. \end{aligned}$$

As for the second term, we have

$$\int \hat{\gamma}_n(x)^2 \nabla_\theta h(x;\theta_0)(\hat{\theta}_n) - \theta_0) dx + R_n,$$

where $R_n = O_P((\hat{\theta}_n - \theta_0)^2) = o_P(\hat{\theta}_n - \theta_0)$. Thus, the second term is of a smaller rate than the first one and the limiting behaviour of $a_n^{-1} \int \hat{\gamma}_n(x)^2 dH(x;\hat{\theta}_n)$ is the same as that of $\int \hat{\gamma}_n(x)^2 h(x;\theta_0) dx$. Thus, Corollary 1.5 follows from Theorem 1.2.

$\odot$

4 Proof of Theorem 1.4

Recall that $\beta > 3/4$. Then

$$\begin{aligned} \sqrt{n}\sigma_{n,1} n^{-1} \hat{\gamma}_n(x) = \sqrt{n}\sigma_{n,1} n^{-1} \beta_n(x-\mu) + \sqrt{n}\left(F(x-\mu) - F\left(x-\mu, \hat{\theta}_n\right)\right) \\ = \quad \sqrt{n}\left(F_n(x-\mu) - F(x-\mu) + f(x-\mu)\sum_{i=1}^n X_i/n\right) \end{aligned}$$

$$
\begin{aligned}
& -f\,(x-\mu)\,\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} - \frac{1}{\sigma}\sqrt{n}(\theta_0 - \hat{\theta}_n) f\,(x-\mu) + O(\sqrt{n}(\theta_0 - \hat{\theta}_n)^2) \\
:= \quad & W_n\,(x-\mu) - f\,(x-\mu)\,\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} - \frac{1}{\sigma}\sqrt{n}(\theta_0 - \hat{\theta}_n) f\,(x-\mu) \\
& +O(\sqrt{n}(\theta_0 - \hat{\theta}_n)^2).
\end{aligned}
$$

If $\theta_0 = \mu$ and $\hat{\theta}_n = \bar{Y}_n$, then via (16),

$$\sup_{x \in \mathbf{R}} \left| \sqrt{n}\sigma_{n,1} n^{-1} \hat{\gamma}_n(x) - W_n\,(x-\mu) \right| = O_P(\sqrt{n}(\mu - \hat{\theta}_n)^2) = o_P(1).$$

Thus, using [17, Theorem 3], we obtain (9).

If $\theta_0 = \mu$ and $\hat{\theta}_n = M_n$, then

$$\sup_{x \in \mathbf{R}} \left| \sqrt{n}\sigma_{n,1} n^{-1} \hat{\gamma}_n(x) - W_n(x) + \frac{1}{\sigma} f\,(x-\mu)\,\sqrt{n}(M_n - \bar{Y}_n) \right| = o_P(1).$$

If $\beta > 3/4$, then from [14, Theorem 1.1], $\sqrt{n}(M_n - \bar{Y}_n) \xrightarrow{\mathrm{d}} N(0, \sigma_\phi^2)$. Thus, (10) follows.

$\odot$

Acknowledgment

This work was initiated during my stay at Carleton University. I am thankful to Professors Barbara Szyszkowicz and Miklós Csörgő for their support and helpful remarks.

References

[1] Beran, J. and Ghosh, S. (1991). Slowly Decaying Correlations, Testing Normality, Nuisance Parameters. *J. Amer. Statist. Assoc.* **86**, 785–791.

[2] Burke, M. D., Csörgő, M., Csörgő, S., Révész, P. (1979). Approximations of the empirical process when parameters are estimated. *Ann. Probab.* **7**, 790–810.

[3] Csörgő, M., Horváth, L. (1993). *Weighted approximations in probability and statistics.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester, 1993.

[4] Csörgő, M., Szyszkowicz, B. and Wang, L. (2006). Strong Invariance Principles for Sequential Bahadur-Kiefer and Vervaat Error Processes of Long-Range Dependence Sequences. *Ann. Statist.* **34**, 1013–1044.

[5] Csörgő, M. and Kulik, R. (2006). Reduction principles for quantile and Bahadur-Kiefer processes of long-range dependent linear sequences. *Preprint.*

[6] Darling, D. A. (1955). The Cramér-Smirnov test in the parametric case. *Ann. Math. Statist.* **26**, 1–20.

[7] Dehling, H. and Taqqu, M. (1989). The Empirical Process of some Long-Range Dependent Sequences with an Applications to U-Statisitcs. *Ann. Statist.* **17**, 1767–1783.

[8] Dehling, H. and Taqqu, M. (1991). Bivariate symmetric statistics of long-range dependent observations. *J. Statist. Pl. Inf.* **28**, 153–165.

[9] Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279–290.

[10] Giraitis, L. and Surgailis, D. (2002). The reduction principle for the empirical process of a long memory linear process. *Empirical process techniques for dependent data*, 241–255, Birkhäuser Boston, Boston, MA.

[11] Ho, H.-C. (2002). On functionals of linear processes with estimated parameters. *Statistica Sinica* **12**, 1171–1190.

[12] Ho, H.-C. and Hsing, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24**, 992–1024.

[13] Kac, M., Kiefer, J., Wolfowitz, J. (1955). On tests of normality and other tests of goodness of fit based on distance methods. *Ann. Math. Statist.* **26**, 189–211.

[14] Koul, H.L. and Surgailis, D. (1996). Asymptotic expansion of M-estimators with long memory moving errors. *Ann. Statist.* **25**, 818–850.

[15] Koul, H.L. and Surgailis, D. (2002). Asymptotic expansion of the empirical process of long memory moving averages. *Empirical process techniques for dependent data*, 213–239, Birkhäuser Boston, Boston, MA.

[16] Ralescu, S. and Puri,M.L. (1984). On the Berry-Esseen rates, the law of the iterated logarithm and an invariance principle for the proportion of the sampel below the sample mean. *J. Multiv. Anal.* **14**, 231–247.

[17] Wu, W.B. (2003). Empirical processes of long-memory sequences. *Bernoulli* **9**, 809–831.

[18] Wu, W.B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *Ann. Statist.* **30**, 1441–1459.

University of Sydney

E-mail: rkuli@maths.usyd.edu.au