Noncentral limit theorems for statistical functionals based on long-memory sequences

Zähle, Henryk
Saarland University, Department of Mathematics
Postfach 151150
D–66041 Saarbrücken, Germany
E-mail: zaehle@math.uni-sb.de

Introduction

Let $\mathbf{V}_T$ be a class of distribution functions on the real line, $\mathbf{V}'$ be a vector space (e.g., $\mathbf{V}' = \mathbb{R}$), and $T : \mathbf{V}_T \to \mathbf{V}'$ be a statistical functional. Let $(X_t)_{t\in\mathbb{N}}$ be a strictly stationary sequence of random variables with distribution function $F \in \mathbf{V}_T$. If $\hat{F}_n$ denotes the empirical distribution functions of $X_1, \ldots, X_n$, i.e., $\hat{F}_n = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[X_i,\infty)}$, then $T(\hat{F}_n)$ can provide a reasonable estimator for $T(F)$. In the context of nonparametric statistics, a central question concerns the asymptotic distribution of $T(\hat{F}_n)$. On the one hand, in the case of *weakly* dependent observations $X_1, X_2, \ldots$ satisfying certain mixing conditions, there are several general results on the asymptotic distribution of $T(\hat{F}_n)$ for various functionals T. See, for instance, [4, 18] for L-functionals, and [5, 10, 11, 25] for V-functionals. On the other hand, in the case of *strongly* dependent observations $X_1, X_2, \ldots$, whose appearance has been observed in numerous scientific areas [2, 3, 20], there are only a few results on the asymptotic distribution of the plug-in estimator $T(\hat{F}_n)$ for some selected functionals T; see, for instance, [9, 17].

This paper (based on [4, 5, 6]) is concerned with a unifying approach for deriving the asymptotic distribution of $T(\hat{F}_n)$ for strongly dependent data. We will avail a version of the Functional Delta Method (FDM). The latter allows to derive the asymptotic distribution of the plug-in estimator $T(\hat{F}_n)$ from the asymptotic distribution of the empirical distribution function $\hat{F}_n$ as long as the functional T is sufficiently regular, more precisely, Hadamard differentiable. The classical FDM [13, 14, 19] was repeatedly criticized for its restricted range of applications. Many tail-dependent statistical functionals T (e.g. general L- or V-functionals) are known to be non-Hadamard differentiable at F. However, recently the concept of *quasi*-Hadamard differentiability was introduced in [4]. This is a weaker concept of differentiability (in particular general L- and V-functionals can be shown to be quasi-Hadamard differentiable), but it is still strong enough to obtain an FDM (referred to as *Modified FDM*); cf. [4, Section 4]. The basic idea of quasi-Hadamard differentiability is to impose a norm only on a suitable subspace $\mathbf{V}_0$ of the space $\mathbb{D}$ ($\supset \mathbf{V}_T$) of all bounded càdlàg functions on $\overline{\mathbb{R}}$ (and not on all of $\mathbb{D}$), and to differentiate only in directions which lie in (some subset of) $\mathbf{V}_0$. It should be stressed that this is not simply the notion of *tangential* Hadamard-differentiability [13, 14, 19] where the tangential space is equipped with the same norm as $\mathbb{D}$. The crucial point is that norms, which assign to F a finite length, are often not strict enough to obtain "differentiability". On the other hand, "differentiability" w.r.t. such good-natured norms is typically not necessary. For details the reader is referred to [4, Section 1].

Upon having established *quasi*-Hadamard differentiability of a given statistical functional T, an application of the Modified FDM typically requires weak convergence of the underlying empirical process w.r.t. a norm being *stricter* than the sup-norm $\|\cdot\|_\infty$, for instance w.r.t. a weighted sup-norm $\|\cdot\|_\lambda := \|(\cdot)\phi_\lambda\|_\infty$ with $\phi_\lambda(x) := (1+|x|)^\lambda$ for some $\lambda > 0$. Here λ depends on the considered statistical functional T. Hence in the context of strongly dependent data, the crucial point is a Noncentral Limit Theorem (NCLT) for the empirical distribution function in a weighted sup-norm. We will first of all present such a result; cf. Theorem 1. Corresponding CLTs can be found in [22] for independent data, in [7] for weakly dependent β-mixing data, in [21] for for weakly dependent α- and ρ-mixing data, and in [24] for weakly dependent causal data.

NCLT for the empirical distribution function of long-memory sequences

Let

$$(1)\qquad X_t := \sum_{s=0}^{\infty} a_s\, \varepsilon_{t-s}, \qquad t \in \mathbb{N},$$

where $(\varepsilon_i)_{i\in\mathbb{Z}}$ are i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with zero mean and finite variance, and the coefficients a_s satisfy $\sum_{s=0}^{\infty} a_s^2 < \infty$ (so that $(X_t)_{t\in\mathbb{Z}}$ is an L^2-process). We assume that the sequence $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary with distribution function F. Many important time series models, such as ARMA and FARIMA, take this form. If $a_0 = 1$ and $a_1 = a_2 = \cdots = 0$, then the X_t are i.i.d. If a_t decays to zero at a sufficiently fast rate, then the covariances $\mathbb{C}\mathrm{ov}(X_1, X_t)$ are summable over $t \in \mathbb{Z}$ and thus the process exhibits short-range dependence (weak dependence). If a_t decays to zero at a sufficiently slow rate, then the covariances $\mathbb{C}\mathrm{ov}(X_1, X_t)$ are *not* summable over $t \in \mathbb{N}$ and thus the process exhibits long-range dependence (strong dependence).

If the X_t are i.i.d., then it is commonly known that the empirical process $n^{1/2}(\hat{F}_n - F)$ converges in distribution to an F-Brownian bridge, i.e. to a centered Gaussian process with covariance function $\Gamma(s,t) = F(s \wedge t)\overline{F}(s \vee t)$. If the X_t are subject to a certain mixing condition (weak dependence), then the limit in distribution of the empirical process $n^{1/2}(\hat{F}_n - F)$ is known to be a centered Gaussian process with covariance function $\Gamma(s,t) = F(s\wedge t)\overline{F}(s \vee t) + \sum_{k=2}^{\infty}[\mathbb{C}\mathrm{ov}(\mathbb{1}_{\{X_1 \le s\}}, \mathbb{1}_{\{X_k \le t\}}) + \mathbb{C}\mathrm{ov}(\mathbb{1}_{\{X_1 \le t\}}, \mathbb{1}_{\{X_k \le s\}})]$; see [7, 12, 21, 24]. If the X_t exhibit long-range dependence (strong dependence, long-memory), then the situation changes drastically: Assuming a moving average structure (1) with $a_s = s^{-\beta}$, $s \in \mathbb{N}$, for $\beta \in (\frac{1}{2}, 1)$, and some additional regularity and moment conditions on the distribution of ε_0, one has

$$(2)\qquad n^{\beta-1/2}(\hat{F}_n(\cdot) - F(\cdot)) \stackrel{\mathrm{d}}{\longrightarrow} c_\beta\, f(\cdot) Z \qquad (\text{in } (\mathbb{D}, \mathcal{D}, \|\cdot\|_\infty))$$

where Z is a standard normally distributed random variable, f is the Lebesgue density of F, c_β is some constant, and $\mathcal{D}$ is the σ-algebra on $\mathbb{D}$ generated by the usual coordinate projections; see e.g. [8, 15, 16]. Notice the asymptotic degeneracy of the limit process in (2) which shows that the increments of the standardized empirical distribution function $\hat{F}_n$ over disjoint intervals, or disjoint observation sets, are asymptotically completely correlated. Also notice the noncentral rate $\beta - 1/2$ in (2).

As indicated in the Introduction, for our purposes the use of the sup-norm $\|\cdot\|_\infty$ in (2) is insufficient. We need a corresponding result for the weighted sup-norm $\|\cdot\|_\lambda := \|(\cdot)\phi_\lambda\|_\infty$. For $\lambda \ge 0$, let $\mathbb{D}_\lambda$ be the space of all càdlàg functions ψ on $\overline{\mathbb{R}}$ with $\|\psi\|_\lambda < \infty$, and $\mathbb{C}_\lambda$ be the subspace of all continuous functions in $\mathbb{D}_\lambda$. We equip $\mathbb{D}_\lambda$ with the σ-algebra $\mathcal{D}_\lambda := \mathcal{D} \cap \mathbb{D}_\lambda$ to make it a measurable space, where as before $\mathcal{D}$ is the σ-algebra generated by the usual coordinate projections. Without loss of generality we assume $a_0 = 1$. The following theorem is proven in [6] using results from [1] and [23].

Theorem 1 (NCLT for $\hat{F}_n$) *Let $\lambda \ge 0$, and assume that*

(i) $a_s = s^{-\beta}\,\ell(s)$, $s \in \mathbb{N}$, where $\beta \in (\frac{1}{2}, 1)$ and ℓ is slowly varying at infinity,

(ii) $\mathbb{E}[|\varepsilon_0|^{2+2\lambda}] < \infty$,

(iii) the distribution function G of ε_0 is twice differentiable and $\sum_{j=1}^{2}\int |G^{(j)}(x)|^2 \phi_{2\lambda}(x)\,dx < \infty$.

Then we have the following analogue of (2):

$$n^{\beta-1/2}\,\ell(n)^{-1}(\hat{F}_n(\cdot) - F(\cdot)) \stackrel{\mathrm{d}}{\longrightarrow} c_{1,\beta}\, f(\cdot) Z \qquad (\mathit{in}\ (\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)),$$

where f is the Lebesgue density of F, Z is a standard normally distributed random variable, and $c_{1,\beta} := \{\mathbb{E}[\varepsilon_0^2](1-(\beta-\frac{1}{2}))(1-(2\beta-1))/(\int_0^\infty (x+x^2)^{-\beta}dx)\}^{1/2}$.

NCLT for plug-in estimators based on long-memory sequences

We now turn to the application of the Modified FDM to $T(\hat{F}_n)$. First of all we recall from [4] the notion of quasi-Hadamard differentiability and the Modified FDM. Let $\mathbf{V}$ and $\mathbf{V}'$ be vector spaces, and $\mathbf{V}_0$ be a subspace of $\mathbf{V}$. Let $\|\cdot\|_{\mathbf{V}_0}$ and $\|\cdot\|_{\mathbf{V}'}$ be norms on $\mathbf{V}_0$ and $\mathbf{V}'$, respectively.

Definition 2 (Quasi-Hadamard differentiability) *Let $T : \mathbf{V}_T \to \mathbf{V}'$ be a mapping defined on a subset $\mathbf{V}_T$ of $\mathbf{V}$, and $\mathbb{C}_0$ be a subset of $\mathbf{V}_0$. Then T is said to be quasi-Hadamard differentiable at $\theta \in \mathbf{V}_T$ tangentially to $\mathbb{C}_0\langle\mathbf{V}_0\rangle$ if there is some continuous mapping $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}T : \mathbb{C}_0 \to \mathbf{V}'$ such that*

$$(3) \qquad \lim_{n\to\infty} \Big\| D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}T\,(v) - \frac{T(\theta + h_n v_n) - T(\theta)}{h_n} \Big\|_{\mathbf{V}'} = 0$$

holds for each triplet $(v, (v_n), (h_n))$, with $v \in \mathbb{C}_0$, $(v_n) \subset \mathbf{V}_0$ satisfying $\|v_n - v\|_{\mathbf{V}_0} \to 0$ as well as $\theta + h_n v_n \in \mathbf{V}_T$ for every $n \in \mathbb{N}$, and $(h_n) \subset (0,\infty)$ satisfying $h_n \to 0$. In this case the mapping $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}T$ is called quasi-Hadamard derivative of T at θ tangentially to $\mathbb{C}_0\langle\mathbf{V}_0\rangle$.

Let $\mathcal{V}_0$ and $\mathcal{V}'$ be σ-algebras on $\mathbf{V}_0$ and $\mathbf{V}'$, respectively. Suppose that $\mathcal{V}_0$ is nested between the open-ball and the Borel σ-algebra on $\mathbf{V}_0$, and that $\mathcal{V}'$ is not larger than the Borel σ-algebra on $\mathbf{V}'$. For every $n \in \mathbb{N}$, let $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ be a probability space, and $\hat{\theta}_n$ be a mapping from Ω_n to $\mathbf{V}$.

Theorem 3 (Modified Functional Delta Method) *Let $T : \mathbf{V}_T \to \mathbf{V}'$ be a mapping defined on some subset $\mathbf{V}_T$ of $\mathbf{V}$, let $\theta \in \mathbf{V}_T$, let $\mathbb{C}_0$ be some subset of $\mathbf{V}_0$ being separable w.r.t. $\|\cdot\|_{\mathbf{V}_0}$ (we regarded $\|\cdot\|_{\mathbf{V}_0}$ as a metric if $\mathbb{C}_0$ is not a vector space), and suppose that*

(i) $\hat{\theta}_n$ takes values only in $\mathbf{V}_T$,

(ii) $\hat{\theta}_n - \theta$ takes values only in $\mathbf{V}_0$, is $(\mathcal{F}_n, \mathcal{V}_0)$-measurable and satisfies

$$r_n(\hat{\theta} - \theta) \xrightarrow{d} V \qquad (\textit{in } (\mathbf{V}_0, \mathcal{V}_0, \|\cdot\|_{\mathbf{V}_0}))$$

for some sequence $(r_n) \subset (0,\infty)$ with $r_n \uparrow \infty$, and some random element V of $(\mathbf{V}_0, \mathcal{V}_0)$, on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values only in $\mathbb{C}_0$,

(iii) $\widetilde{\omega} \mapsto T(W(\widetilde{\omega}) + \theta)$ is $(\widetilde{\mathcal{F}}, \mathcal{V}')$-measurable whenever W is a measurable mapping from some measurable space $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ to $(\mathbf{V}_0, \mathcal{V}_0)$ such that $W(\widetilde{\omega}) + \theta \in \mathbf{V}_T$ for all $\widetilde{\omega} \in \widetilde{\Omega}$,

(iv) T is quasi-Hadamard differentiable at θ tangentially to $\mathbb{C}_0\langle\mathbf{V}_0\rangle$ with quasi-Hadamard derivative $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}T$.

Then

$$r_n(T(\hat{\theta}_n) - T(\theta)) \xrightarrow{d} D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}T\,(V) \qquad (\textit{in } (\mathbf{V}', \mathcal{V}', \|\cdot\|_{\mathbf{V}'})).$$

As an immediate consequence of Theorems 1 and 3 we now obtain the following NCLT for the plug-in estimator $T(\hat{F}_n)$. We choose $\mathbf{V} := \mathbb{D}$, $\mathbf{V}_0 := \mathbb{D}_\lambda$, $\mathbb{C}_0 := \mathbb{C}_\lambda$, and assume that $\mathbf{V}_T$ is a class of distribution functions on the real line containing F.

Theorem 4 (NCLT for $T(\hat{F}_n)$) *Let $\lambda \geq 0$, and assume that*

(i) $\hat{F}_n$ takes values only in $\mathbf{V}_T$,

(ii) the assumptions of Theorem 1 are fulfilled,

(iii) $\widetilde{\omega} \mapsto T(W(\widetilde{\omega}) + F)$ is $(\widetilde{\mathcal{F}}, \mathcal{V}')$-measurable whenever W is a measurable mapping from some measurable space $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ to $(\mathbb{D}_\lambda, \mathcal{D}_\lambda)$ such that $W(\widetilde{\omega}) + F \in \mathbf{V}_T$ for all $\widetilde{\omega} \in \widetilde{\Omega}$,

(iv) *T is quasi-Hadamard differentiable at F tangentially to $\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle$ with quasi-Hadamard derivative $D^{\rm Had}_{F;\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle}T$.*

Then

$$n^{\beta-1/2}\,\ell(n)^{-1}(T(\hat{F}_n(\cdot))-T(F(\cdot)))\stackrel{\rm d}{\longrightarrow} D^{\rm Had}_{F;\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle}T(c_{1,\beta}\,f(\cdot)Z)\qquad (\textit{in } (\mathbf{V}',\mathcal{V}',\|\cdot\|_{\mathbf{V}'})),$$

where β, $c_{1,\beta}$, f and Z are as in Theorem 1.

Example 5 *(L-functionals)* Let K be the distribution function on $[0,1]$, and $\mathbf{V}_K$ be the class of all distribution functions F on the real line for which $\int|x|\,dK(F(x))<\infty$. The functional $\mathcal{L}$, defined by

$$\mathcal{L}(F):=\mathcal{L}_K(F):=\int x\,dK(F(x)),\qquad F\in\mathbf{V}_K,$$

is called L-functional associated with K. It was shown in [4] that if K is continuous and piecewise differentiable, the (piecewise) derivative K' is bounded above and $F\in\mathbf{V}_K$ takes the value $d\in(0,1)$ at most once if K is not differentiable at d, then for every $\lambda>1$ the functional $\mathcal{L}:\mathbf{V}_K\to\mathbb{R}$ is quasi-Hadamard differentiable at F tangentially to $\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle$ with quasi-Hadamard derivative

$$D^{\rm Had}_{F;\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle}\mathcal{L}\,(v)\;=\;\int K'(F(x))\,v(x)\,dx\qquad\forall\, v\in\mathbb{C}_\lambda.$$

Thus, if also the assumptions of Theorem 1 are fulfilled with $f\in\mathbb{C}_\lambda$, Theorem 4 (with $\mathbf{V}'=\mathbb{R}$) yields

$$n^{\beta-1/2}\,\ell(n)^{-1}(\mathcal{L}(\hat{F}_n)-\mathcal{L}(F))\stackrel{\rm d}{\longrightarrow}\widetilde{Z}\qquad(\text{in }(\mathbb{R},\mathcal{B}(\mathbb{R}),|\cdot|)),$$

where $\widetilde{Z}$ is normally distributed with mean zero and variance $c_{1,\beta}^2(\int K'(F(x))f(x)dx)^2$, and β and $c_{1,\beta}$ are as in Theorem 1. ◇

Example 6 *(V-functionals)* Let $g:\mathbb{R}^2\to\mathbb{R}$ be a measurable function, and $\mathbf{V}_g$ be the class of all distribution functions F on the real line for which $\int\int|g(x_1,x_2)|dF(x_1)dF(x_2)<\infty$. The functional $\mathcal{U}$, defined by

$$\mathcal{U}(F):=\mathcal{U}_g(F):=\int\int g(x_1,x_2)\,dF(x_1)dF(x_2),\qquad F\in\mathbf{V}_g,$$

is called von Mises-functional (V-functional) associated with g. Let $\mathbb{BV}_{\rm loc}$ be the space of all functions $\psi:\mathbb{R}\to\mathbb{R}$ of local bounded variation. For $\psi\in\mathbb{BV}_{\rm loc}$, we denote by $d\psi^+$ and $d\psi^-$ the unique positive Radon measures induced by the Jordan decomposition of ψ, and we set $|d\psi|:=d\psi^++d\psi^-$. Suppose that, for some $\lambda>\lambda'\geq 0$, the following two assertions hold:

(a) For every $x_2\in\mathbb{R}$ fixed, the function $g_{x_2}(\cdot):=g(\,\cdot\,,x_2)$ lies in $\mathbb{BV}_{\rm loc}\cap\mathbb{D}_{-\lambda'}$. Moreover, the function $x_2\mapsto\int\phi_{-\lambda}(x_1)|dg_{x_2}|(x_1)$ lies in $\mathbb{D}_{-\lambda'}$.

(b) The functions $g_{1,F}(\cdot):=\int g(\,\cdot\,,x_2)dF(x_2)$ and $g_{2,F}(\cdot):=\int g(x_1,\,\cdot\,)dF(x_1)$ lie in $\mathbb{BV}_{\rm loc}$, and we have $\int\phi_{-\lambda}(x)\,|dg_{i,F}|(x)<\infty$ for $i=1,2$. Moreover, the functions $\overline{g_{1,F}}(\cdot):=\int|g(\,\cdot\,,x_2)|dF(x_2)$ and $\overline{g_{2,F}}(\cdot):=\int|g(x_1,\,\cdot\,)|dF(x_1)$ lie in $\mathbb{D}_{-\lambda'}$.

It is shown in [5] that under assumptions (a)–(b) the functional $\mathcal{U}$ is quasi-Hadamard differentiable at F tangentially to $\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle$ with quasi-Hadamard derivative

$$D^{\rm Had}_{F;\mathbb{C}_\lambda\langle\mathbb{D}_\lambda\rangle}\mathcal{U}\,(v)\;=\;-\int v(x)dg_{1,F}(x)-\int v(x)dg_{2,F}(x)\qquad\forall\, v\in\mathbb{C}_\lambda. \tag{4}$$

Thus, if also the assumptions of Theorem 1 are fulfilled with $f\in\mathbb{C}_\lambda$, Theorem 4 (with $\mathbf{V}'=\mathbb{R}$) yields

$$n^{\beta-1/2}\,\ell(n)^{-1}(\mathcal{U}(\hat{F}_n)-\mathcal{U}(F))\stackrel{\rm d}{\longrightarrow}\widetilde{Z}\qquad(\text{in }(\mathbb{R},\mathcal{B}(\mathbb{R}),|\cdot|)), \tag{5}$$

where $\widetilde{Z}$ is normally distributed with mean zero and variance $c_{1,\beta}^2(\int f(x)dg_{1,F}(x) + \int f(x)dg_{2,F}(x))^2$, and β and $c_{1,\beta}$ are as in Theorem 1.

It is easy to show that the variance kernel $g(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$ and Gini's mean difference kernel $g(x_1, x_2) = |x_1 - x_2|$ satisfy conditions (a)–(b) for $\lambda' = 2$ and $\lambda' = 1$ (respectively), where $dg_{1,F}(x) = dg_{2,F}(x) = (x - \mathbb{E}[X_1])dx$ and $dg_{1,F}(x) = dg_{2,F}(x) = (2F(x) - 1)dx$ (respectively); cf. [5]. In the former case, however, it is straightforwardly seen that the asymptotic variance in (5) vanishes, so that the right-hand side in (5) degenerates to zero. This is consistent with Example 1 in [9]. ◇

Remark 7 *(Degenerate V-functionals)* Among V-functionals —introduced in Example 6— the functionals with a so-called degenerate kernel have attracted special interest; see, e.g., [8, 9]. A kernel g is called *degenerate* w.r.t. $F \in \mathbf{V}_g$ if the functions $g_{1,F}$ and $g_{2,F}$ defined in part (b) in Example 6 are identically zero. In this case, $\mathcal{U}$ is called *degenerate* V-functional w.r.t. F. Moreover, in this case the right-hand side in (4) vanishes and thus the right-hand side in (5) degenerates to zero. That is, an application of Theorem 4 yields little. However, in this case one can exploit the Continuous Mapping Theorem (CMT) instead of the Modified FDM. Indeed: By the degeneracy of the kernel g we have the representation $\mathcal{U}(\hat{F}_n) = \int\int g(x_1, x_2)\, d(\hat{F}_n - F)(x_1)d(\hat{F}_n - F)(x_2)$ and it was pointed out in [8, Section 2] that, under certain conditions on g and F, integration-by-parts yields

$$\mathcal{U}(\hat{F}_n) = \int\int (\hat{F}_n - F)(x_1)(\hat{F}_n - F)(x_2)\, dg(x_1, x_2). \tag{6}$$

To apply integration-by-parts, it was assumed in [8] that the kernel g is right-continuous and has bounded total variation. However, as the assumption that g be of bounded total variation is too restrictive, the result of [8, Section 2] was extended in [9] to more general kernels. A related, slightly stronger result can be found in [6]. Now, if the assumptions of Theorem 1 hold for some $\lambda \geq 0$ for which the integral $\int\int \phi_{-\lambda}(x_1)\phi_{-\lambda}(x_2)\, |dg|(x_1, x_2)$ is finite, then we immediately obtain from (6), Theorem 1, $\mathcal{U}(F) = 0$ (which holds by the degeneracy of g) and the CMT that

$$n^{2\beta-1}\,\ell(n)^{-2}\,\mathcal{U}(\hat{F}_n) \stackrel{\mathrm{d}}{\longrightarrow} \Big(c_{1,\beta}^2 \int\int f(x_1)f(x_2)dg(x_1, x_2)\Big) Z^2 \qquad (\text{in } (\mathbb{R}, \mathcal{B}(\mathbb{R}), |\cdot|)),$$

where Z^2 is χ_1^2-distributed, and β and $c_{1,\beta}$ are as in Theorem 1. For details, in particular for the conditions on g and F ensuring the representation (6), see [6]. ◇

References

[1] Avram, F. and Taqqu, M. (1987). Noncentral limit theorems and Appell polynomials. *Annals of Probability*, 15, 767–775.

[2] Baillie, R.T. (1996). Long memory processes and fractional integration in econometics. *Journal of Econometrics*, 73, 5–59.

[3] Beran J. (1994). *Statistics for Long-Memory Processes.* Chapman & Hall, New York.

[4] Beutner, E. and Zähle, H. (2010). A modified functional delta method and its application to the estimation of risk functionals. *Journal of Multivariate Analysis*, 101, 2452–2463.

[5] Beutner, E. and Zähle, H. (2011). Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes, *Bernoulli*, forthcoming.

[6] Beutner, E., Wu, W.B. and Zähle, H. (2011). Asymptotics for statistical functionals of long-memory sequences. *Submitted.*

[7] Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130, 307–335.

[8] Dehling, H. and Taqqu, M.S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *Annals of Statistics*, 17, 1767–1783.

[9] Dehling, H. and Taqqu, M.S. (1991). Bivariate symmetric statistics of long-range dependent observations. *Journal of Statistical Planning and Inference*, 28, 153–165.

[10] Dehling, H. and Wendler, M. (2010). Central limit theorem and the bootstrap for U-statistics of strongly mixing data. *Journal of Multivariate Analysis*, 101(1), 126–137.

[11] Dewan, I. and Prakasa Rao, B.L.S. (2002). Central limit theorems of U-statistics of associated random variables. *Statistics and Probability Letters*, 57, 9–15.

[12] Doukhan, P., Massart, P. and Rio, E. (1995). The invariance principle for the empirical measure of a weakly dependent process. *Annales de l'Institut Henri Poincar* B, 31, 393–427.

[13] Fernholz, L.T. (1983). Von Mises calculus for statistical functionals, *Lecture Notes in Statistics*, 19, Springer-Verlag, New York.

[14] Gill, R.D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part I). *Scandinavian Journal of Statistics*, 16, 97–128.

[15] Giraitis, L. and Surgailis, D. (1999). Central limit theorem for the empirical process of a linear sequence with long memory. *Journal of Statistical Planning and Inference*, 80, 81–93.

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

[17] Hsing, T. and Wu, W.B. (2004). On weighted U-Statistics for stationary processes. *Annals of Probability*, 32, 1600–1631.

[18] Mehra, K.L. and Rao, M.S. (1975). On functions of order statistics for mixing processes. *Annals of Statistics*, 3, 874-883.

[19] Reeds, J.A. (1976). On the definition of von Mises functionals. *Research Report S-44. Department of Statistics, Harvard University.*

[20] Robinson, J. (1994). Time series with strong dependence. *In:* Sims, E. (Ed.), *Advances in econometric*, Econometric Society Monographs 23, Cambridge, 47–94.

[21] Shao, Q.-M. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability*, 24, 2098–2127.

[22] Shorack, G.R. and Wellner, J.A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

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

[24] Wu, W.B. (2008). Empirical processes of stationary sequences. *Statistica Sinica*, 18, 313–333.

[25] Yoshihara, K. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35, 237–252.

ABSTRACT

Noncentral limit theorems for statistical functionals based on strictly stationary time series exhibiting long-range dependence are presented. The key tool is a noncentral limit theorem for empirical processes of long-memory data with respect to nonuniform sup-norms. Using a modified Functional Delta Method, based on the new concept of quasi-Hadamard differentiability, one can easily derive the asymptotic distribution of fairly general statistics, including L- and V-statistics.