Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes

ERIC BEUTNER[1] and HENRYK ZÄHLE[2]

[1]*Department of Quantitative Economics, Maastricht University, P.O. Box 616, NL–6200 MD Maastricht, The Netherlands. E-mail: e.beutner@maastrichtuniversity.nl*
[2]*Department of Mathematics, Saarland University, Postfach 151150, D–66041 Saarbrücken, Germany. E-mail: zaehle@math.uni-sb.de*

It is commonly acknowledged that V-functionals with an unbounded kernel are not Hadamard differentiable and that therefore the asymptotic distribution of U- and V-statistics with an unbounded kernel cannot be derived by the Functional Delta Method (FDM). However, in this article we show that V-functionals are quasi-Hadamard differentiable and that therefore a modified version of the FDM (introduced recently in (*J. Multivariate Anal.* **101** (2010) 2452–2463)) can be applied to this problem. The modified FDM requires weak convergence of a weighted version of the underlying empirical process. The latter is not problematic since there exist several results on weighted empirical processes in the literature; see, for example, (*J. Econometrics* **130** (2006) 307–335, *Ann. Probab.* **24** (1996) 2098–2127, *Empirical Processes with Applications to Statistics* (1986) Wiley, *Statist. Sinica* **18** (2008) 313–333). The modified FDM approach has the advantage that it is very flexible w.r.t. both the underlying data and the estimator of the unknown distribution function. Both will be demonstrated by various examples. In particular, we will show that our FDM approach covers mainly all the results known in literature for the asymptotic distribution of U- and V-statistics based on dependent data – and our assumptions are by tendency even weaker. Moreover, using our FDM approach we extend these results to dependence concepts that are not covered by the existing literature.

Keywords: Functional Delta Method; Jordan decomposition; quasi-Hadamard differentiability; stationary sequence of random variables; U- and V-statistic; weak dependence; weighted empirical process

1. Introduction

For a distribution function (d.f.) F on the real line, we consider the characteristic

$$U(F) := \int \int g(x_1, x_2) \, dF(x_1) \, dF(x_2) \tag{1}$$

with $g : \mathbb{R}^2 \to \mathbb{R}$ some measurable function, provided the double integral exists. A systematic theory for the nonparametric estimation of $U(F)$ was initiated in [14] and [27]. A natural estimator for $U(F)$ is given by

$$U(F_n) := \int \int g(x_1, x_2) \, dF_n(x_1) \, dF_n(x_2), \tag{2}$$

where F_n denotes some estimate of F based on the first n observations of a sequence $X_1, X_2, \ldots$ of random variables (on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$) being identically distributed according to F. Sometimes $U(F_n)$ is called von-Mises-statistic (or simply V-statistic) with kernel g. If F_n is the empirical d.f. $\hat{F}_n := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[X_i, \infty)}$ of $X_1, \ldots, X_n$, then we obtain

$$U(\hat{F}_n) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} g(X_i, X_j), \tag{3}$$

and we note that $U(\hat{F}_n)$ is closely related to the U-statistic

$$U_n := \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1: j \neq i}^{n} g(X_i, X_j). \tag{4}$$

If $X_1, \ldots, X_n$ are i.i.d., then U_n is an unbiased estimator whereas $U(\hat{F}_n)$ is generally not so. However, U_n and $U(\hat{F}_n)$ typically share the same asymptotic properties; cf. Remark 2.5 below. Also notice that, in the nonparametric setting, U_n is the minimum variance unbiased estimator of $U(F) = \mathbb{E}[g(X_1, X_2)]$ whenever $X_1, \ldots, X_n$ are i.i.d. For background on U-statistics see, for instance, [5,7,14,16,20,21,23].

We note that several features of a d.f. F can be expressed as in (1), for instance, the variance of F, or Gini's mean difference of two independent random variables with d.f. F; for details, see Section 3.

Our objective is the asymptotic distribution of $U(F_n)$, that is, the weak limit of the empirical error $\sqrt{n}(U(F_n) - U(F))$. In the existing literature, the starting point for the derivation of the asymptotic distribution of U-statistics U_n is usually the Hoeffding decomposition [14] of U_n. Using this decomposition, asymptotic normality of U_n was shown in [14] for i.i.d. sequences, in [19] for *-mixing stationary sequences, in [8,31] for β-mixing stationary sequences, in [10] for associated random variables, and recently in [6] for α-mixing stationary sequences (recall from [3], page 109: i.i.d. $\Rightarrow$ *-mixing $\Rightarrow$ β-mixing $\Rightarrow$ α-mixing). Another approach is based on the orthogonal expansion of the kernel g; see, for example, [9] and the references therein.

In this article, we derive the asymptotic distribution of U- and V-statistics by means of a Functional Delta Method (FDM). The use of an FDM is known to be beneficial for the following reason. Provided the functional U can be shown to be Hadamard differentiable at F, it is basically enough to derive the asymptotic distribution of F_n to obtain the asymptotic distribution of $U(F_n)$. Therefore, this method is especially useful for deriving the asymptotic distribution of the estimator $U(\hat{F}_n)$ based on dependent data, because – given the Hadamard differentiability – one "only" has to derive the asymptotic distribution of $\hat{F}_n$ based on data subjected to a certain dependence structure. There are already several respective results on the asymptotic distribution of $\hat{F}_n$ based on dependent data in the literature (e.g., [4,24,30]), and new respective results (combined with the assumed Hadamard differentiability) would immediately yield also the asymptotic distribution of $U(\hat{F}_n)$.

However, one has to be careful with the application of an FDM to our problem. The classical FDM in the sense of [12,13,18] (see also [28,29]) cannot be applied to many interesting statistical functionals depending on the tails of the underlying distribution, because the method typically

relies on Hadamard differentiability w.r.t. the uniform sup-norm. For instance, as pointed out in [28] and [22], whenever F has an unbounded support Hadamard differentiability w.r.t. the uniform sup-norm can be shown neither for an L-statistic with a weight function having one of the endpoints (or both endpoints) of the closed interval $[0, 1]$ in its support nor for a U-statistic with unbounded kernel. However, in [2] a modified version of the FDM was introduced which is suitable also for nonuniform sup-norms (imposed on the tangential space only), and it was in particular shown that this modified version can also be applied to L-statistics with a weight function having one of the endpoints (or both endpoints) of the closed interval $[0, 1]$ in its support. In contrast to the classical FDM, our FDM is based on the notion of quasi-Hadamard differentiability and requires weak convergence of the empirical process $\sqrt{n}(\hat{F}_n - F)$ w.r.t. a nonuniform sup-norm, that is, in other words, weak convergence of a weighted version of the empirical process. Fortunately, the latter is not problematic, because there are many results on the weak convergence of weighted empirical processes in the literature; see [26] for i.i.d. data, and [4,24,30] for dependent data.

In the present article, we demonstrate that the modified version of the FDM can be applied to derive the limiting distribution for U- and V-statistics with an unbounded kernel g. For simplicity of notation, we restrict the derivations to kernels of degree 2. However, in Remark 4.2, we clarify how the results can be extended to kernels of degree $d \geq 3$. Using our FDM approach, we will be able to a great extent to recover the results mentioned above (the conditions imposed by our approach will turn out to be weaker by tendency) and to extend them to other concepts of dependence; cf. Section 3.2. The FDM approach will also turn out to be useful when the empirical d.f. is replaced by a different estimate of F, for instance by a smoothed version of the empirical d.f.; cf. Example 3.4.

The remainder of this article is organized as follows. In Section 2, we state the conditions under which the asymptotic distribution of U- and V-statistics can be derived by the modified version of the FDM and present our main result. The conditions imposed can be divided into two parts: on the one hand conditions on the kernel g and the d.f. F, and on the other hand conditions on an empirical process. In Section 3, we give several examples for both, that is, for kernels g and d.f. F as well as empirical processes fulfilling the conditions imposed. In the Appendix A, we recall the Jordan decomposition of functions of locally bounded variation, which will be beneficial for our applications in Section 3. Finally, in the Appendix B we give an integration-by-parts formula and a sort of weighted Helly-Bray theorem. Both results are needed in Section 4 to show quasi-Hadamard differentiability of V-functionals.

2. Main result

Our main result is Theorem 2.3 below, which provides a CLT for the V-statistic $U(F_n)$ subject to Assumption 2.1. Let $\mathbb{D}_\lambda$ be the space of all càdlàg functions ψ on $\overline{\mathbb{R}}$ with $\|\psi\|_\lambda < \infty$, where $\|\psi\|_\lambda := \|\psi \phi_\lambda\|_\infty$ refers to the nonuniform sup-norm based on the weight function $\phi_\lambda(x) := (1 + |x|)^\lambda$, for $\lambda \in \mathbb{R}$ fixed. As usual, we let $0 \cdot \infty := 0$. If $\lambda \geq 0$, then 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 $\mathcal{D}$ is the σ-algebra generated by the usual coordinate projections $\pi_x : \mathbb{D} \to \mathbb{R}$, $x \in \overline{\mathbb{R}}$, with $\mathbb{D}$ the space of all bounded càdlàg functions on $\overline{\mathbb{R}}$. Further, let $\mathbb{BV}_{\text{loc}}$ be the space of all functions $\psi : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ being real-valued and

of local bounded variation on $\mathbb{R}$. For $\psi \in \mathbb{BV}_{\text{loc}}$, we denote by $\mathrm{d}\psi^+$ and $\mathrm{d}\psi^-$ the unique positive Radon measures induced by the Jordan decomposition of ψ (for details, see the Appendix A), and we set $|\mathrm{d}\psi| := \mathrm{d}\psi^+ + \mathrm{d}\psi^-$. Finally, we will interpret integrals as being over the open interval $(-\infty, \infty)$, that is, $\int = \int_{(-\infty,\infty)}$.

Assumption 2.1. *We assume that for some $\lambda > \lambda' \geq 0$ the following assertions hold*:

(a) *For every $x_2 \in \mathbb{R}$ fixed, the function $g_{x_2}(\cdot) := g(\cdot, x_2)$ lies in $\mathbb{BV}_{\text{loc}} \cap \mathbb{D}_{-\lambda'}$. Moreover, the function $x_2 \mapsto \int \phi_{-\lambda}(x_1)|\mathrm{d}g_{x_2}|(x_1)$ is measurable and finite w.r.t. $\|\cdot\|_{-\lambda'}$.*

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

(c) *F is continuous, the double integral in (1) exists, and $\int \phi_{\lambda'}(x)\,\mathrm{d}F(x) < \infty$.*

(d) *$F_n : \Omega \to \mathbb{D}$ is $(\mathcal{F}, \mathcal{D})$-measurable, and every realization of F_n is nonnegative and nondecreasing, has variation bounded by 1, the double integral in (2) exists and $\int \phi_{\lambda'}(x)\,\mathrm{d}F_n(x) < \infty$, for every $n \in \mathbb{N}$.*

(e) *The process $\sqrt{n}(F_n - F)$ is a random element of $(\mathbb{D}_\lambda, \mathcal{D}_\lambda)$ for all $n \in \mathbb{N}$, and there is some random element B° of $(\mathbb{D}_\lambda, \mathcal{D}_\lambda)$ with continuous samples such that*

$$\sqrt{n}(F_n - F) \xrightarrow{d} B^\circ \qquad in\ (\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda). \tag{5}$$

The assumptions (a) and (b) will allow us to prove quasi-Hadamard differentiability of the functional U (defined in (1)) at F; see Section 4. At first glance, they seem to be awkward but in an application their verification is often straightforward, see Section 3.1. To understand the meaning of conditions (a) and (b), let us suppose that we want to derive the asymptotic distribution of U- and V-statistics by means of the classical FDM in the sense of [12,13,18]. Then we would have to prove Hadamard differentiability of the functional U given by (1) at F. If F has an unbounded support this could be done by imposing Assumptions 2.1(a) and (b) with $\lambda' = 0$, that is, with the uniform sup-norm. Thus, as pointed out in the Introduction, an application of the classical FDM for the derivation of the asymptotic distribution of U- and V-statistics would, inter alia, require a uniformly bounded kernel g (cf. [22]). On the other hand, the modified FDM only requires that this boundedness holds w.r.t. the weaker nonuniform sup-norm $\|\cdot\|_{-\lambda'}$ for some $\lambda' \geq 0$.

Remark 2.2. Notice that

(a)′ Assumption 2.1(a) could, alternatively, be imposed on g_{x_1} defined similar as g_{x_2}. Further notice that the second requirement in Assumption 2.1(a) is rather weak. Indeed: In the examples to be given in Section 3.1 the function $x_2 \mapsto \int \phi_{-\lambda}(x_1)|\mathrm{d}g_{x_2}|(x_1)$ even lies in $\mathbb{D}$.

(b)′ The last part of Assumption 2.1(b) implies $g_{1,F}, g_{2,F} \in \mathbb{D}_{-\lambda'}$.

(c)′ Continuity of F is required for the application of the modified FDM.

(d)′ Assumption 2.1(d) is always fulfilled if F_n is the empirical d.f. $\hat{F}_n$.

(e)′ Assumption 2.1(e) does not require that F lies in $\mathbb{D}_\lambda$ or that F_n is a random element of $(\mathbb{D}_\lambda, \mathcal{D}_\lambda)$. These conditions would actually fail to hold.

Theorem 2.3. *Under Assumption* 2.1, *we have*

$$\sqrt{n}\bigl(U(F_n) - U(F)\bigr) \xrightarrow{d} \dot{U}(B^\circ) \qquad in \ (\mathbb{R}, \mathcal{B}(\mathbb{R}), |\cdot|) \tag{6}$$

with

$$\dot{U}_F(B^\circ) := -\int B^\circ(x)\,dg_{1,F}(x) - \int B^\circ(x)\,dg_{2,F}(x). \tag{7}$$

Proof. First of all, notice that the integrals in (7) exist by Assumptions 2.1(b) and (e). Now, let $\mathbb{BV}_{1,d}$ be the space of all càdlàg functions in $\mathbb{BV}_{\mathrm{loc}}$ with variation bounded by 1, and $\mathbb{U}$ be the class of all nonnegative and nondecreasing functions $f \in \mathbb{BV}_{1,d}$ for which the integral on the right-hand side of equation (8) below and the integral $\int \phi_{\lambda'}(x)\,df(x)$ exist. We define a functional $U : \mathbb{U} \to \mathbb{R}$ by setting

$$U(f) := \int \int g(x_1, x_2)\,df(x_1)\,df(x_2), \qquad f \in \mathbb{U}, \tag{8}$$

so that $U(F)$ and $U(F_n)$ defined in (1)–(2) can be written as $U(f)$ with $f := F$ and $f_n := F_n$, respectively. We are going to apply an FDM to the functional U. The version of the FDM we need for our purposes is given in [2], Theorem 4.1. It is based on the notion of quasi-Hadamard differentiability which is also introduced in [2], Definition 2.1.

Let $\mathbb{C}_\lambda$ be the space of all continuous functions in $\mathbb{D}_\lambda$, and notice that $\mathbb{C}_\lambda$ is separable w.r.t. $\|\cdot\|_\lambda$. For every f in U's domain $\mathbb{U}$ we define a functional $\dot{U}_f : \mathbb{C}_\lambda \to \mathbb{R}$ by setting

$$\dot{U}_f(v) := -\int v(x)\,dg_{1,f}(x) - \int v(x)\,dg_{2,f}(x), \qquad v \in \mathbb{C}_\lambda, \tag{9}$$

where $g_{i,f}$ is defined analogously to $g_{i,F}$ (cf. Assumption 2.1(b)). Lemma 4.1 below shows that, subject to Assumption 2.1(a)–(c), the functional U is quasi-Hadamard differentiable at $f := F$ tangentially to $\mathbb{C}_\lambda \langle \mathbb{D}_\lambda \rangle$ with quasi-Hadamard derivative $\dot{U}_F$. Thus, assumption (iv) of Theorem 4.1 in [2] (with $f = U$, $V_f = \mathbb{U}$, $(\mathbf{V}', \|\cdot\|_{\mathbf{V}'}) = (\mathbb{R}, |\cdot|)$, $(\mathbf{V}_0, \|\cdot\|_{\mathbf{V}_0}) = (\mathbb{D}_\lambda, \|\cdot\|_\lambda)$, $\mathbb{C}_0 = \mathbb{C}_\lambda$, $\theta = F$ and $T_n = F_n$) is fulfilled. Therefore, the statement of Theorem 2.3 would follow from the FDM given in Theorem 4.1 in [2] if we could verify that also the conditions (i)–(iii) of this theorem are satisfied. Conditions (i) and (ii) are satisfied by Assumption 2.1(d) and (e), respectively. It thus remains to verify (iii), that is, that the mapping $\widetilde{\omega} \mapsto U(W(\widetilde{\omega}) + F)$ is $(\widetilde{\mathcal{F}}, \mathcal{B}(\mathbb{R}))$-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 $W(\widetilde{\omega}) + F \in \mathbb{U}$ for all $\widetilde{\omega} \in \widetilde{\Omega}$. Since W is $(\widetilde{\mathcal{F}}, \mathcal{D}_\lambda)$-measurable and $\mathcal{D}_\lambda$ is the projection σ-field, we obtain in particular $(\widetilde{\mathcal{F}}, \mathcal{B}(\mathbb{R}))$-measurability of $\widetilde{\omega} \mapsto W(x, \widetilde{\omega})$ for every $x \in \overline{\mathbb{R}}$. Along with the representation (1), this yields $(\widetilde{\mathcal{F}}, \mathcal{B}(\mathbb{R}))$-measurability of $\widetilde{\omega} \mapsto U(W(\widetilde{\omega}) + F)$. $\square$

We emphasize that Theorem 2.3 is quite a flexible tool to derive the asymptotic distribution of the plug-in estimate $U(F_n)$. In fact: Apart from checking the technical Assumptions 2.1(a)–(d), it is enough to establish the CLT (5) for F_n in order to obtain the CLT (6) for $U(F_n)$. Section 3 below demonstrates this flexibility by various examples.

Remark 2.4. If B° in Theorem 2.3 is a Gaussian process with zero mean and measurable covariance function Γ and if $\int \int \Gamma(x, y) \, dg_{i,F}(x) \, dg_{j,F}(y)$ exists for every $i, j \in \{1, 2\}$, then the random variable $\dot{U}_F(B^\circ)$ defined in (7) is normally distributed with mean 0 and variance

$$\sigma^2 := \sum_{i=1}^{2} \sum_{j=1}^{2} \int \int \Gamma(x, y) \, dg_{i,F}(x) \, dg_{j,F}(y). \tag{10}$$

Remark 2.5. If $\mathbb{E}[|g(X_1, X_1)|] < \infty$ (in Examples 3.1 and 3.2 below we even have $g(x, x) = 0$ for all $x \in \mathbb{R}$), then the particular V-statistic $U(\hat{F}_n)$ and the U-statistic U_n (defined in (3) and (4), resp.) have the same asymptotic distribution. To see this, we first of all note that (for $n \geq 2$)

$$\begin{aligned}
&\sqrt{n}\big(U_n - U(F)\big) \\
&= \sqrt{n}\big(U_n - U(\hat{F}_n)\big) + \sqrt{n}\big(U(\hat{F}_n) - U(F)\big) \\
&= \frac{\sqrt{n}}{n-1} U(\hat{F}_n) - \frac{\sqrt{n}}{n(n-1)} \sum_{i=1}^{n} g(X_i, X_i) + \sqrt{n}\big(U(\hat{F}_n) - U(F)\big) \\
&=: S_1(n) - S_2(n) + \sqrt{n}\big(U(\hat{F}_n) - U(F)\big).
\end{aligned} \tag{11}$$

As $\sqrt{n}(U(\hat{F}_n) - U(F))$ converges weakly to some nondegenerate limit, we obtain by Slutzky's lemma that $S_1(n) = \frac{1}{n-1}\sqrt{n}(U(\hat{F}_n) - U(F)) + \frac{\sqrt{n}}{n-1}U(F)$ converges in probability to zero. Further, by the Markov inequality we know that, for every $\varepsilon > 0$ fixed, $\mathbb{P}[|S_2(n)| > \varepsilon]$ is bounded above by $\frac{1}{\varepsilon}\mathbb{E}[|S_2(n)|]$ which, in turn, is bounded above by $\frac{\sqrt{n}}{n-1}\frac{1}{\varepsilon}\mathbb{E}[|g(X_1, X_1)|]$. So we also have that $S_2(n)$ converges in probability to zero. Slutzky's lemma and (11) thus imply that $\sqrt{n}(U_n - U(F))$ has indeed the same limit distribution as $\sqrt{n}(U(\hat{F}_n) - U(F))$.

Remark 2.6. The linear part of the Hoeffding decomposition of $U_n - U(F)$ (cf. [23], page 178) multiplied by $\sqrt{n}$ can be written as $\sum_{i=1}^{2} \int g_{i,F} \, d(\sqrt{n}(\hat{F}_n - F))$, for example, using the integration-by-parts formula (22), as $-\sum_{i=1}^{2} \int \sqrt{n}(\hat{F}_n - F) \, dg_{i,F}$. Then, if we could show that the degenerate part of U_n converges in probability to zero (which is nontrivial for dependent data), we could recover (6) with U_n in place of $U(F_n)$ by using (5) and the Continuous Mapping theorem.

3. Examples

In this section, we give some examples for g, F and F_n satisfying Assumption 2.1. At first, in Section 3.1, we provide examples for g (and F) satisfying Assumptions 2.1(a)–(b). Thereafter, in Section 3.2, we will give examples for F_n (and F) satisfying Assumptions 2.1(d)–(e) for various types of data. We assume throughout this section that Assumption 2.1(c) is fulfilled because its meaning is rather obvious and the conditions imposed by it are fairly weak.

3.1. Examples for *g*

In [1], one can find a number of examples for kernels g for which $U(F)$ corresponds to a popular characteristic of F. By means of two popular examples, we now illustrate how to verify the Assumptions 2.1(a)–(b). It will be seen that the verification of these assumptions is easy, though, at first glance, it may seem cumbersome. We will use the notion of Jordan decomposition $\psi = \psi(c) + \psi_c^+ - \psi_c^-$ centered at some point $c \in \mathbb{R}$. For the reader's convenience, we have recalled the essentials in the Appendix A.

Example 3.1 (Gini's mean difference). If $g(x_1, x_2) = |x_1 - x_2|$ and F has a finite first moment, then $U(F)$ equals Gini's mean difference $\mathbb{E}[|X_1 - X_2|]$ of two i.i.d. random variables X_1 and X_2 on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with d.f. F. Then the Assumptions 2.1(a)–(b) are fulfilled for $\lambda' = 1$. Indeed: We have $g_{x_2}(x_1) = (x_1 - x_2)\mathbb{1}_{(x_2, \infty]}(x_1) - (x_1 - x_2)\mathbb{1}_{[-\infty, x_2]}(x_1)$, so that the first part of Assumption 2.1(a) obviously holds. Further, the Jordan decomposition (18) of g_{x_2} centered at $c = x_2$ reads as $g_{x_2}(x_1) = 0 + g_{x_2 x_2}^+(x_1) - g_{x_2 x_2}^-(x_1)$, where $g_{x_2 x_2}^+(x_1) = (x_1 - x_2)\mathbb{1}_{(x_2, \infty]}(x_1)$ and $g_{x_2 x_2}^-(x_1) = (x_1 - x_2)\mathbb{1}_{[-\infty, x_2]}(x_1)$, and so, in view of Lemma A.1, $\mathrm{d}g_{x_2}^+(x_1) = \mathbb{1}_{(x_2, \infty]}(x_1)\,\mathrm{d}x_1$ and $\mathrm{d}g_{x_2}^-(x_1) = \mathbb{1}_{[-\infty, x_2]}(x_1)\,\mathrm{d}x_1$. Now it can be seen easily that also the second part of Assumption 2.1(a) holds; we omit the details. Let us now turn to Assumption 2.1(b). We have

$$
\begin{aligned}
g_{1,F}(x_1) \\
&= \mathbb{E}\big[X_2 \mathbb{1}_{(x_1, \infty]}(X_2)\big] - x_1 \mathbb{P}[X_2 > x_1] + x_1 \mathbb{P}[X_2 \le x_1] - \mathbb{E}\big[X_2 \mathbb{1}_{[-\infty, x_1]}(X_2)\big] \\
&= x_1\big(2F(x_1) - 1\big) - \mathbb{E}[X_2] + 2\mathbb{E}\big[X_2 \mathbb{1}_{(x_1, \infty]}(X_2)\big] \\
&= K + x_1 + 2\big(-x_1\big(1 - F(x_1)\big) + \mathbb{E}\big[X_2 \mathbb{1}_{(x_1, \infty]}(X_2)\big]\big) \\
&= K + x_1 + 2\int_{x_1}^{\infty} \big(1 - F(x)\big)\,\mathrm{d}x
\end{aligned}
$$

with $K := -\mathbb{E}[X_2]$. The same representation holds for $g_{2,F}$. So we obviously have $g_{i,F} = \overline{g_{i,F}} \in \mathbb{D}_{-1} \cap \mathbb{BV}_{\mathrm{loc}}$ for $i = 1, 2$. Moreover, we have $g_{i,F}'(x) = 2F(x) - 1$, and so there is some constant $c \in \mathbb{R}$ such that $g_{i,F}$ is nonincreasing on $(-\infty, c)$ and is nondecreasing on (c, ∞), for $i = 1, 2$. Since the density of $|\mathrm{d}g_{i,F}|$ on $(-\infty, c)$ and the density of $|\mathrm{d}g_{i,F}|$ on (c, ∞) are bounded, we also have $\int \phi_{-\lambda}(x)|\mathrm{d}g_{i,F}|(x) < \infty$ for $i = 1, 2$ and every $\lambda > 1$. That is, all parts of Assumption 2.1(b) hold true. Thus, Assumptions 2.1(a)–(b) hold true.

If also Assumptions 2.1(d)–(e) hold true, then we obtain from Theorem 2.3 for the kernel $g(x_1, x_2) = |x_1 - x_2|$ that $\dot{U}(B^\circ) = 2\int B^\circ(x)(1 - 2F(x))\,\mathrm{d}x$, because $\mathrm{d}g_{1,F}(x) = \mathrm{d}g_{2,F}(x) = (2F(x) - 1)\,\mathrm{d}x$.

Example 3.2 (Variance). If $g(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$ and F has a finite second moment, then $U(F)$ equals the variance of F. In this case, the Assumptions 2.1(a)–(b) are fulfilled for $\lambda' = 2$. The verification of this is even easier than the elaborations in Example 3.1. We note that this time, we obtain $\mathrm{d}g_{x_2}^+(x_1) = (x_1 - x_2)\mathbb{1}_{(x_2, \infty]}(x_1)\,\mathrm{d}x_1$ and $\mathrm{d}g_{x_2}^-(x_1) = (x_2 - x_1)\mathbb{1}_{[-\infty, x_2]}(x_1)\,\mathrm{d}x_1$ as well

as $dg_{i,F}^+(x_i) = (x_i - \mathbb{E}[X_j])\mathbb{1}_{(\mathbb{E}[X_j],\infty]}(x_i)\,dx_i$ and $dg_{i,F}^-(x_i) = (\mathbb{E}[X_j] - x_i)\mathbb{1}_{[-\infty,\mathbb{E}[X_j]]}(x_i)\,dx_i$ for $i, j \in \{1, 2\}$ with $i \neq j$.

If also Assumptions 2.1(d)–(e) hold true, then we obtain from Theorem 2.3 for the kernel $g(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$ that $\dot{U}(B^\circ) = 2\int B^\circ(x)(\mathbb{E}[X_1] - x)\,dx$, because $dg_{1,F}(x) = dg_{2,F}(x) = (x - \mathbb{E}[X_1])\,dx$.

3.2. Examples for F_n

Here we will give some examples for estimators F_n for F that satisfy Assumption 2.1(d)–(e). We first consider the case of i.i.d. data.

Example 3.3 (Empirical d.f. of i.i.d. data). Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with d.f. F, and let $\lambda \geq 0$. If F has a finite γ-moment for some $\gamma > 2\lambda$, then Theorem 6.2.1 in [26] shows that for the empirical d.f. $\hat{F}_n$ of $X_1, \ldots, X_n$,

$$\sqrt{n}(\hat{F}_n - F) \xrightarrow{d} B_F^\circ \qquad (\text{in } (\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)), \tag{12}$$

where B_F° is an F-Brownian bridge, that is, a centered Gaussian process with covariance function $\Gamma(x, y) = F(x \wedge y)\overline{F}(x \vee y)$. Thus, if $\lambda > 0$, if F has a finite γ-moment for some $\gamma > 2\lambda$, and if g is a kernel satisfying Assumptions 2.1(a)–(b) for F and some $\lambda' \in [0, \lambda)$, then Theorem 2.3 shows that the law of $\sqrt{n}(U(\hat{F}_n) - U(F))$ converges weakly to the normal distribution with mean 0 and variance given by (10) with $\Gamma(x, y) = F(x \wedge y)\overline{F}(x \vee y)$. Alternatively, the result can be stated as follows: If g is a fixed kernel and $\mathbb{F}_{g,\lambda'}$ denotes the class of all d.f. F for which Assumptions 2.1(a)–(b) hold with $\lambda' \geq 0$, then $\sqrt{n}(U(\hat{F}_n) - U(F))$ converges weakly to the above mentioned normal distribution for every $F \in \mathbb{F}_{g,\lambda'}$ having a finite γ-moment for some $\gamma > 2\lambda'$. Indeed: In this case, we can choose $\lambda \in (\lambda', \gamma/2)$ in Assumption 2.1(e).

Example 3.4 (Smoothed empirical d.f. of i.i.d. data). Suppose that in the setting of Example 3.3 the empirical d.f. $\hat{F}_n$ is smoothed out by the heat kernel $p_{\varepsilon_n}(\cdot)$ with bandwidth $\varepsilon_n \geq 0$, that is, that $\hat{F}_n$ is replaced by $P_{\varepsilon_n}\hat{F}_n$ with $(P_\varepsilon)_{\varepsilon \geq 0}$ the heat semigroup (i.e., $P_\varepsilon\psi := \int_\mathbb{R} \psi(y)p_\varepsilon(\cdot - y)\,dy$ for $\varepsilon > 0$, and $P_0 := \mathbb{I}$). Then, if F is also Lipschitz continuous and $\sqrt{n}\varepsilon_n^{(\gamma-\lambda)/(2\gamma)} \to 0$, the CLT (12) (with $\hat{F}_n$ replaced by $P_{\varepsilon_n}\hat{F}_n$) still holds (cf. Corollary A.2 in [2]), and therefore the weak limit of the law of $\sqrt{n}(U(P_{\varepsilon_n}\hat{F}_n) - U(F))$ is still the normal distribution with mean 0 and variance given by (10) with $\Gamma(x, y) = F(x \wedge y)\overline{F}(x \vee y)$. Of course, at this point we have to ensure that under the imposed assumptions the expression $U(P_{\varepsilon_n}\hat{F}_n)$ is well defined, that is, that Assumption 2.1(d) is satisfied. Now, it can be easily deduced from Lemma 3.2 in [32] that in our setting $P_{\varepsilon_n}\hat{F}_n$ lies in $\mathbb{D}_\lambda$. Thus, if we assume that, for example, $\sup_{x_1,x_2\in\mathbb{R}} |g(x_1, x_2)|\phi_{-\lambda'}(x_1)\phi_{-\lambda'}(x_2) < \infty$, Assumption 2.1(d) follows easily.

Let us now turn to the case of dependent data, which is our actual objective. Throughout the examples presented below, we consider a strictly stationary sequence $(X_i) = (X_i)_{i \geq 1}$ of random

variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with continuous d.f. F, and let as before $\hat{F}_n$ denote the corresponding empirical d.f. at stage n. By strict stationarity, we mean that the joint distribution of $X_{i+1}, \ldots, X_{i+m}$ does not depend on i for every fixed positive integer m. We will consider three popular dependency structures (α-, β- and ρ-mixing) in more detail in Examples 3.5, 3.6, and 3.7, respectively. There, we will also provide a comparison of the results obtained by the approach considered here and the results obtained up to now. For the definition of α-, β- and ρ-mixing (and other) mixing conditions and for examples of strictly stationary α-, β- and ρ-mixing sequences see, for example, [3,11,17]. As usual, the corresponding mixing coefficients will be referred to as $\alpha(n)$, $\beta(n)$ and $\rho(n)$, respectively. The application of our method to other dependence concepts will be discussed in Example 3.8. Notice that the condition of α-mixing is weaker than the condition of β-mixing (absolute regularity) under which CLTs for U-statistics have been established in [8,31]. A CLT for strictly stationary α-mixing (strongly mixing) sequences of random variables has been given in [6].

Example 3.5 (Empirical d.f. of α-mixing data). Let (X_i) be α-mixing with $\alpha(n) = \mathcal{O}(n^{-\theta})$ for some $\theta > 1 + \sqrt{2}$, and let $\lambda \geq 0$. If F has a finite γ-moment for some $\gamma > \frac{2\theta\lambda}{\theta-1}$, then it can easily be deduced from Theorem 2.2 in [24] that

$$\sqrt{n}(\hat{F}_n - F) \xrightarrow{d} \widetilde{B}_F^\circ \qquad (\text{in } (\mathbb{D}_\lambda, \mathcal{D}_\lambda, \|\cdot\|_\lambda)) \tag{13}$$

with $\widetilde{B}_F^\circ$ a continuous centered Gaussian process with covariance function

$$\Gamma(s, t) = F(s \wedge t)\bar{F}(s \vee t) \tag{14}$$
$$+ \sum_{k=2}^{\infty} \left[\mathbb{Cov}\left(\mathbb{1}_{\{X_1 \leq s\}}, \mathbb{1}_{\{X_k \leq t\}}\right) + \mathbb{Cov}\left(\mathbb{1}_{\{X_1 \leq t\}}, \mathbb{1}_{\{X_k \leq s\}}\right) \right]$$

(cf. Section 3.3 in [2]). Thus, if g is a fixed kernel and $\mathbb{F}_{g,\lambda'}$ denotes the class of all d.f. satisfying Assumptions 2.1(a)–(b) for some $\lambda' \geq 0$, then Theorem 2.3 shows that the law of $\sqrt{n}(U(\hat{F}_n) - U(F))$ converges weakly to the normal distribution with mean 0 and variance given by (10), with Γ as in (14), for every d.f. $F \in \mathbb{F}_{g,\lambda'}$ having a finite γ-moment for some $\gamma > \frac{2\theta\lambda'}{\theta-1}$. Indeed: In this case we can choose $\lambda \in (\lambda', \gamma(\theta - 1)/(2\theta))$ in Assumption 2.1(e).

To compare our result with that of Theorem 1.8 in [6], we consider the kernel $g(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$. For Theorem 1.8 in [6] to be applicable, we must assume that F has a finite γ-moment for some $\gamma > 4$ (the same condition is necessary to ensure that the approach considered here works). In this case, both integrability conditions in Theorem 1.8 in [6] are fulfilled, and the condition on the mixing coefficients reads as follows: $\alpha(n) = \mathcal{O}(n^{-\theta})$ for some $\theta > \frac{3}{2} + \frac{1}{2\gamma} + \frac{5}{\gamma-4} + \frac{2}{\gamma(\gamma-4)} = \frac{3\gamma-1}{2\gamma-8}$. On the other hand, if F has a finite γ-moment for some $\gamma > 4$, in our setting we may choose $\lambda' = 2$, and so $\theta > \frac{\gamma}{\gamma-4}$ (and $\lambda \in (2, \frac{\gamma(\theta-1)}{2\theta})$). Hence, our condition on the mixing coefficients reads as follows: $\alpha(n) = \mathcal{O}(n^{-\theta})$ for some $\theta > \frac{\gamma}{\gamma-4}$. Notice that $\frac{3\gamma-1}{2\gamma-8} > \frac{\gamma}{\gamma-4}$ holds for all $\gamma > 4$. Taking into account that in our setting, we must choose $\theta > 1 + \sqrt{2}$ for the result of [24] to be applicable we find that our result relies on a weaker assumption on the mixing coefficients than Theorem 1.8 in [6] whenever $\frac{3\gamma-1}{2\gamma-8} > 1 + \sqrt{2}$, that is, $\gamma < \frac{7+8\sqrt{2}}{2\sqrt{2}-1}$.

Example 3.6 (Empirical d.f. of β-mixing data). Let (X_i) be β-mixing with $\beta(n) = O(n^{-\theta})$ for some $\theta > \frac{\kappa}{\kappa-1}$ with $\kappa > 1$, and let $\lambda \geq 0$. If F has a finite γ-moment for some $\gamma > 2\lambda\kappa$, then it can easily be deduced from Lemma 4.1 in [4] that the CLT (13) still holds and that the covariance function is again given by (14). Thus, if g is a fixed kernel and $\mathbb{F}_{g,\lambda'}$ denotes the class of all d.f. satisfying Assumptions 2.1(a)–(b) for some $\lambda' \geq 0$, then Theorem 2.3 shows that the law of $\sqrt{n}(U(\hat{F}_n) - U(F))$ converges weakly to the normal distribution with mean 0 and variance given by (10), with Γ as in (14), for every d.f. $F \in \mathbb{F}_{g,\lambda'}$ having a finite γ-moment for some $\gamma > 2\lambda'\kappa$. Indeed: In this case we can choose $\lambda \in (\lambda', \frac{\gamma}{2\kappa})$.

To compare our result with that of Theorem 3.1 in [31] (see also Theorem 1.8 in [6]), we consider the kernel $g(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$. For this theorem to be applicable, we must again assume that F has a finite γ-moment for some $\gamma > 4$ (the same condition is again necessary to ensure that the approach considered here works). In this case, both integrability conditions in Theorem 3.1 in [31] (see also Theorem 1.8 in [6]) are fulfilled, and the condition on the mixing coefficients reads as follows: $\beta(n) = \mathcal{O}(n^{-\theta})$ for some $\theta > \frac{\gamma}{\gamma-4}$. On the other hand, if F has a finite γ-moment for some $\gamma > 4$, in our setting we may choose $\lambda' = 2$, and so $\kappa < \gamma/4$ (and $\lambda \in (2, \frac{\gamma}{2\kappa})$). Hence, in view of $\theta > \frac{\kappa}{\kappa-1}$, our condition on the mixing coefficients reads as follows: $\beta(n) = \mathcal{O}(n^{-\theta})$ for some $\theta > \frac{\gamma}{\gamma-4}$. That is, both results impose the same condition on the mixing coefficients.

Example 3.7 (Empirical d.f. of ρ-mixing data). Let (X_i) be ρ-mixing with $\sum_{n=1}^{\infty} \rho(2^n) < \infty$, suppose $\sum_{k=2}^{\infty} |\mathbb{C}ov(\mathbb{1}_{\{X_1 \leq s\}}, \mathbb{1}_{\{X_k \leq t\}}) + \mathbb{C}ov(\mathbb{1}_{\{X_1 \leq t\}}, \mathbb{1}_{\{X_k \leq s\}})| < \infty$, and let $\lambda \geq 0$. If F has a finite γ-moment for some $\gamma > \lambda(2 + \varepsilon)$ with $\varepsilon > 0$, then it can easily be deduced from Theorem 2.3 in [24] that the CLT (13) still holds and that the covariance function is again given by (14) (cf. Section 3.3 in [2]). Hence, we again have in this case: If g is a fixed kernel and if we denote by $\mathbb{F}_{g,\lambda'}$ the class of all d.f. for which Assumptions 2.1(a)–(b) hold for some $\lambda' \geq 0$, then Theorem 2.3 yields that the law of $\sqrt{n}(U(\hat{F}_n) - U(F))$ converges weakly to the normal distribution with mean 0 and variance given by (10) with Γ as in (14) for every $F \in \mathbb{F}_{g,\lambda'}$ having a finite γ-moment for some $\gamma > \lambda'(2 + \varepsilon)$. Indeed: In this case, we can choose $\lambda \in (\lambda', \gamma/(2 + \varepsilon))$.

Up to our best knowledge, the asymptotic distribution of U- and V-statistics of ρ-mixing data has not been studied explicitly so far. Of course, every ρ-mixing sequence is also α-mixing (since $\alpha(n) \leq \frac{1}{4}\rho(n)$; see [3], Inequality (1.12)), but the condition on the mixing coefficients imposed in Example 3.7 is considerably weaker than the condition on the mixing coefficients imposed in Example 3.5. Similar statements apply to further dependence concepts, and one also obtains that further dependence concepts are also covered by our approach.

Example 3.8 (Further examples). Recently, a new dependence structure for sequences of random variables was introduced in [30]. Thus, not surprising, limit distributions for U- and V-statistics under this dependence concepts have not been derived so far. Anyhow, in [30] it was also proved that, subject to certain conditions, the weighted empirical process $\sqrt{n}(\hat{F}_n - F)\phi_\gamma$ converges weakly to a tight Gaussian process. Here $\hat{F}_n$ is the empirical d.f. based on a sequence of random variables fulfilling this dependence condition. From our Theorem 2.3 one can thus (along the lines of Examples 3.5, 3.6, and 3.7) derive the limit distribution of U- and V-statistics when the data fulfills the dependence structure in [30]. We omit the details.

In [10], the limit distribution of U-statistics for associated sequences was derived using the Hoeffding decomposition. To prove asymptotic normality of U-statistics for stationary and associated sequences, it was required there that the partial derivatives of g are uniformly bounded. This clearly excludes the variance of a random variable. On the other hand, our approach also covers the variance for the case of stationary and associated sequences. Indeed: Let (X_i) be a stationary, associated sequence with $\mathbb{C}\text{ov}(X_1, X_n) = \mathrm{O}(n^{-\nu-\varepsilon})$ for some $\nu \geq (3 + \sqrt{33})/2$ and $\varepsilon > 0$. Then, we can deduce from Theorem 2.4 in [24] that the CLT (13) still holds and the covariance function is again given by (14) whenever F has a finite γ-moment for some $\gamma > \frac{2\lambda\nu}{\nu-3}$ ($\lambda \geq 0$ fixed). Hence, we obtain from Theorem 2.3 (recall from Example 3.2 that Assumptions 2.1(a)–(b) are fulfilled for the variance with $\lambda' = 2$) that the variance is included in our method of proof whenever F has a finite γ-moment for some $\gamma > \frac{4\nu}{\nu-3}$; in this case we can choose $\lambda \in (2, \gamma(\nu - 3)/(2\nu))$.

4. Quasi-Hadamard differentiability of *U*

This section is concerned with the quasi-Hadamard differentiability (in the sense of Definition 2.1 in [2]) of the functional U defined in (8). Recall that quasi-Hadamard differentiability is needed in the proof of Theorem 2.3. Recall also that $\mathbb{BV}_{1,\mathrm{d}}$ is the space of all càdlàg functions in $\mathbb{BV}_{\mathrm{loc}}$ with variation bounded by 1, and that $\mathbb{U}$ is the class of all nonnegative and nondecreasing functions $f \in \mathbb{BV}_{1,\mathrm{d}}$ for which the integral on the right-hand side of equation (8) and the integral $\int \phi_{\lambda'}(x)\,\mathrm{d}f(x)$ exist. Moreover, we let $\mathbb{BV}_{\mathrm{loc},\mathrm{d}}$ be the space of all càdlàg functions in $\mathbb{BV}_{\mathrm{loc}}$.

Lemma 4.1. *Under Assumptions* 2.1(a)–(c) (*the continuity of F is actually superfluous at this point*), *the functional U defined in* (8) *is quasi-Hadamard differentiable at* $f := F$ *tangentially to* $\mathbb{C}_\lambda \langle \mathbb{D}_\lambda \rangle$ *with quasi-Hadamard derivative given by* $\dot{U}_f$ *defined in* (9) *with* $f := F$.

Proof. To prove the claim, we have to show that

$$\lim_{n\to\infty}\left| \dot{U}_f(v) - \frac{U(f + h_n v_n) - U(f)}{h_n} \right| = 0 \tag{15}$$

holds for each triplet $(v, (v_n), (h_n))$ with $v \in \mathbb{C}_\lambda$, $(v_n) \subset \mathbb{D}_\lambda$ satisfying $f + h_n v_n \in \mathbb{U}$ (for all $n \in \mathbb{N}$) as well as $\|v_n - v\|_\lambda \to 0$, and $(h_n) \subset \mathbb{R}_0 := \mathbb{R} \setminus \{0\}$ satisfying $h_n \to 0$. Let $f_n := f + h_n v_n$. We stress the fact that f_n lies in $\mathbb{U}$ which is a subset of $\mathbb{BV}_{1,\mathrm{d}}$, and that consequently $h_n v_n$ is the difference of two functions which both lie in $\mathbb{U}$ (notice that f lies in $\mathbb{U}$ by Assumption 2.1(c)). For the verification of (15), we now proceed in two steps.

Step 1. To justify the analysis in Step 2 below, we first of all show that the three integrals

$$\int |g_{1,f}|(x_1)|\mathrm{d}v_n|(x_1), \qquad \int |g_{2,f}|(x_2)|\mathrm{d}v_n|(x_2), \qquad \int\int |g(x_1, x_2)||\mathrm{d}v_n|(x_1)|\mathrm{d}v_n|(x_2)$$

are finite for all $n \in \mathbb{N}$. For the finiteness of these integrals, it suffices to show that for every $n \in \mathbb{N}$

$$\int\int |g(x_1, x_2)|\,\mathrm{d}f_n(x_1)\,\mathrm{d}f(x_2) < \infty \quad \text{and} \quad \int\int |g(x_1, x_2)|\,\mathrm{d}f(x_1)\,\mathrm{d}f_n(x_2) < \infty, \tag{16}$$

since $|g_{1,f}| \leq \int |g(\cdot, x_2)| \, df(x_2)$ and $|g_{2,f}| \leq \int |g(x_1, \cdot)| \, df(x_1)$, since $h_n|dv_n| = df_n + df$, and since $f, f_n \in \mathbb{U}$ implies

$$\int \int |g(x_1, x_2)| \, df(x_1) \, df(x_2) < \infty \quad \text{and} \quad \int \int |g(x_1, x_2)| \, df_n(x_1) \, df_n(x_2) < \infty.$$

(Notice that (16) by itself is also needed in Step 2 below.) We clearly have

$$\int \int |g(x_1, x_2)| \, df(x_1) \, df_n(x_2) \leq \|\overline{g_{2,f}}\|_{-\lambda'} \int \phi_{\lambda'}(x_2) \, df_n(x_2).$$

From the second part of Assumption 2.1(b) we have $\|\overline{g_{2,f}}\|_{-\lambda'} < \infty$, and $\int \phi_{\lambda'}(x_2) \, df_n(x_2) < \infty$ holds since $f_n \in \mathbb{U}$. That is, $\|\overline{g_{2,f}}\|_{-\lambda'} \int \phi_{\lambda'}(x_2) \, df_n(x_2) < \infty$. Similar arguments show that the first inequality in (16) holds.

Step 2. By Step 1 and the triangular inequality we have

$$\left| \dot{U}_f(v) - \frac{U(f + h_n v_n) - U(f)}{h_n} \right|$$

$$= \left| - \int v(x_1) \, dg_{1,f}(x_1) - \int v(x_2) \, dg_{2,f}(x_2) \right.$$

$$- \frac{1}{h_n} \left(\int \int g(x_1, x_2) \, d(f + h_n v_n)(x_1) \, d(f + h_n v_n)(x_2) \right.$$

$$\left. \left. - \int \int g(x_1, x_2) \, df(x_1) \, df(x_2) \right) \right| \tag{17}$$

$$\leq \sum_{i=1}^{2} \left| - \int v(x_i) \, dg_{i,f}(x_i) - \int g_{i,f}(x_i) \, dv_n(x_i) \right|$$

$$+ \left| h_n \int \int g(x_1, x_2) \, dv_n(x_1) \, dv_n(x_2) \right|$$

$$=: \sum_{i=1}^{2} S_{1,i}(n) + S_2(n).$$

In order to show that $S_{1,1}(n)$ converges to zero, we will apply the integration-by-parts formula (22) to $\int g_{1,f}(x_1) \, dv_n(x_1)$. At first, we have to make clear that formula (22) can be applied, that is, that the assumptions of Lemma B.1 are fulfilled.

It follows from Step 1 that the second condition in (21) holds true (where $g_{1,f}$ and v_n play the roles of u and v, resp.). Moreover, by the continuity of $\phi_{-\lambda}$ we have

$$\int |v_n(x_1-)| |dg_{1,f}|(x_1)$$

$$= \int |v_n(x_1-)\phi_\lambda(x_1-)\phi_{-\lambda}(x_1-)| |dg_{1,f}|(x_1)$$

$$= \int |v_n(x_1-)\phi_\lambda(x_1-)|\phi_{-\lambda}(x_1)|dg_{1,f}|(x_1)$$

$$\leq \|v_n\|_\lambda \int \phi_{-\lambda}(x_1)|dg_{1,f}|(x_1).$$

By Assumption 2.1(b) and the fact that $v_n \in \mathbb{D}_\lambda$, the latter bound is finite, so that also the first condition in (21) holds true. We finally note that $\lim_{|x_1|\to\infty} v_n(x_1)g_{1,f}(x_1) = 0$. Indeed: On one hand, $|g_{1,f}(x_1)\phi_{-\lambda'}(x_1)|$ is bounded above uniformly in x_1 by Assumption 2.1(b) and Remark 2.2(b)′. On the other hand, $|v_n(x_1)\phi_{\lambda'}(x_1)|$ converges to 0 as $|x_1| \to \infty$ because $|v_n(x_1)\phi_\lambda(x_1)|$ is bounded above uniformly in x_1 (recall $\lambda > \lambda'$). That is, the assumptions of Lemma B.1 are indeed fulfilled.

Now, we may apply the integration-by-parts formula (22) to obtain

$$S_{1,1}(n) = \left| -\int v(x_1)\,dg_{1,f}(x_1) + \int v_n(x_1-)\,dg_{1,f}(x_1) \right|$$

$$\leq \left| \int (v_n - v)(x_1)\,dg_{1,f}(x_1) \right| + \left| \int (v_n(x_1-) - v_n(x_1))\,dg_{1,f}(x_1) \right|$$

$$\leq (\|v_n - v\|_\lambda + \|v_n - v\|_\lambda + \|v - v_n\|_\lambda) \int \phi_{-\lambda}(x_1)|dg_{1,f}|(x_1).$$

The latter bound converges to zero by Assumption 2.1(b) and $\|v - v_n\|_\lambda \to 0$. That is, $S_{1,1}(n) \to 0$. In the same way we obtain $S_{1,2}(n) \to 0$.

Thus, it remains to show $S_2(n) \to 0$. We will apply the integration-by-parts formula (22) to the inner integral in $S_2(n)$. So at first we will verify that formula (22) can be used, that is, that the assumptions of Lemma B.1 are fulfilled. By Assumption 2.1(a), we have $g_{x_2} \in \mathbb{BV}_{\text{loc,d}}$, and as mentioned above we also have $v_n \in \mathbb{BV}_{\text{loc,d}}$. Further, the integrals $\int g(x_1, x_2)\,df(x_1)$ and $\int g(x_1, x_2)\,df_n(x_1)$ exist by the fact that $f_n, f \in \mathbb{U}$ and Fubini's theorem. This and the representation $v_n = (f_n - f)/h_n$ imply $\int |g_{x_2}(x_1)||dv_n|(x_1) < \infty$, that is, that the second condition in (21) holds true. Moreover, by the continuity of $\phi_{-\lambda}$ we have as above

$$\int |v_n(x_1-)||dg_{x_2}|(x_1) = \int |v_n(x_1-)\phi_\lambda(x_1-)\phi_{-\lambda}(x_1-)||dg_{x_2}|(x_1)$$

$$= \int |v_n(x_1-)\phi_\lambda(x_1-)|\phi_{-\lambda}(x_1)||dg_{x_2}|(x_1)$$

$$\leq \|v_n\|_\lambda \int \phi_{-\lambda}(x_1)|dg_{x_2}|(x_1).$$

By Assumption 2.1(a) and the fact that $v_n \in \mathbb{D}_\lambda$, this bound is finite, so that also the first condition in (21) holds true. We finally note that $\lim_{|x_1|\to\infty} v_n(x_1)g_{x_2}(x_1) = 0$. Indeed: On one hand, $|g_{x_2}(x_1)\phi_{-\lambda'}(x_1)|$ is bounded above uniformly in x_1 by Assumption 2.1(a). On the other hand, $|v_n(x_1)\phi_{\lambda'}(x_1)|$ converges to 0 as $|x_1| \to \infty$ since $|v_n(x_1)\phi_\lambda(x_1)|$ is bounded above uniformly in x_1 (recall $\lambda > \lambda'$). That is, the assumptions of Lemma B.1 are indeed fulfilled.

Now, we may apply the integration-by-parts formula (22) to the inner integral in $S_2(n)$ to obtain

$$
\begin{aligned}
S_2(n) &= \left| - \int \int v_n(x_1-) \, dg_{x_2}(x_1) \, d(f_n - f)(x_2) \right| \\
&\leq \left| - \int \int (v_n(x_1-) - v(x_1-)) \, dg_{x_2}(x_1) \, d(f_n - f)(x_2) \right| \\
&\quad + \left| \int \int v(x_1-) \, dg_{x_2}(x_1) \, d(f_n - f)(x_2) \right|.
\end{aligned}
$$

Since f_n and f generate positive (probability) measures, and v and $\phi_{-\lambda'}$ are continuous, we may continue with

$$
\begin{aligned}
&\leq \|v_n - v\|_\lambda \int \left(\int \phi_{-\lambda}(x_1) |dg_{x_2}|(x_1) \phi_{-\lambda'}(x_2) \right) \phi_{\lambda'}(x_2) \, df_n(x_2) \\
&\quad + \|v_n - v\|_\lambda \int \left(\int \phi_{-\lambda}(x_1) |dg_{x_2}|(x_1) \phi_{-\lambda'}(x_2) \right) \phi_{\lambda'}(x_2) \, df(x_2) \\
&\quad + \left| \int \left(\int v(x_1) \, dg_{x_2}(x_1) \right) df_n(x_2) - \int \left(\int v(x_1) \, dg_{x_2}(x_1) \right) df(x_2) \right| \\
&\leq \|v_n - v\|_\lambda \int C \phi_{\lambda'}(x_2) \, df_n(x_2) + \|v_n - v\|_\lambda \int C \phi_{\lambda'}(x_2) \, df(x_2) \\
&\quad + \left| \int \left(\int v(x_1) \, dg_{x_2}(x_1) \right) df_n(x_2) - \int \left(\int v(x_1) \, dg_{x_2}(x_1) \right) df(x_2) \right| \\
&=: S_{2,1}(n) + S_{2,2}(n) + S_{2,3}(n)
\end{aligned}
$$

with $C := \sup_{x_2} \int \phi_{-\lambda}(x_1) |dg_{x_2}|(x_1) \phi_{-\lambda'}(x_2)$ (which is finite by the second part of Assumption 2.1(a)). By Lemma B.2, which can be applied due to Assumption 2.1(a), and the facts that $v \in \mathbb{D}_\lambda$, $\|f_n - f\|_\lambda \to 0$, and that $\int \phi_{\lambda'}(x_2) \, df(x_2)$ and $\int \phi_{\lambda'}(x_2) \, df_n(x_2)$ exist, the summand $S_{2,3}(n)$ converges to 0. Since $\|v_n - v\|_\lambda \to 0$, and since $\int \phi_{\lambda'}(x_2) \, df(x_2)$ is finite because $f \in \mathbb{U}$, we also obtain $S_{2,2}(n) \to 0$. It remains to show $S_{2,1}(n) \to 0$. As $\|v_n - v\|_\lambda \to 0$, it suffices to show that $\int \phi_{\lambda'}(x_2) \, df_n(x_2)$ is uniformly bounded from above. The latter follows from the finiteness of $\int \phi_{\lambda'}(x_2) \, df(x_2)$ and Lemma B.2 which is applicable since we clearly have $\phi_{\lambda'} \in \mathbb{D}_{-\lambda'}$, and for every $n \in \mathbb{N}$ the integral $\int \phi_{\lambda'}(x_2) \, df_n(x_2)$ exists due to $f_n \in \mathbb{U}$. This proves the claim of Lemma 4.1. $\qquad\square$

Remark 4.2. We note that the proof of Lemma 4.1 basically applies also to V-functionals of the shape $U(F) = \int \cdots \int g(x_1, \ldots, x_d) \, dF(x_1) \cdots dF(x_d)$ with arbitrary $d \geq 2$, provided Assumptions 2.1(a)–(b) (which ensure the quasi-Hadamard differentiability of U in the case $d = 2$) are modified suitably and the definition of $\dot{U}_f$ in (9) is replaced by $\dot{U}_f(v) := -\sum_{i=1}^d \int v(x) \, dg_{i,f}(x)$ with $g_{i,f}(x_i) := \int \cdots \int g(x_1, \ldots, x_d) \, df(x_1) \cdots df(x_{i-1}) \, df(x_{i+1}) \cdots df(x_d)$. In particular, Theorem 2.3 then still holds for such general V-functionals. Let us exemplify the validity of

the analogue of Lemma 4.1 for the case $d = 3$. To do so, we let $\mathbb{M}_{(\lambda,\lambda)}$ be the space of all measurable functions $h: \mathbb{R}^2 \to \mathbb{R}$ such that $\sup_{x_1,x_2} |h(x_1,x_2)\phi_\lambda(x_1)\phi_\lambda(x_2)|$ is finite. To ensure the existence of the integrals as in Step 1 in the above proof, it is sufficient to require that the functions $\overline{g_{i,j,f}}(x_i,x_j) := \int |g(x_1,x_2,x_3)| \, df(x_k)$, $i, j, k \in \{1,2,3\}$, $i < j$, $k \neq i$, $k \neq j$, are in $\mathbb{M}_{(-\lambda',-\lambda')}$, and that the functions $\overline{g_{i,f}}(x_i) := \int |g(x_1,x_2,x_3)| \, df(x_j) \, df(x_k)$, $i, j, k \in \{1,2,3\}$ pairwise disjoint, lie in $\mathbb{D}_{-\lambda'}$ (cf. the second part of Assumption 2.1(b)). Then Step 1 still holds. Let us turn to Step 2 in the above proof. In (17), we now obtain the bound

$$S_1(n) + S_2(n) + S_3(n) := \sum_{i=1}^{3} \left| -\int v(x_i) \, dg_{i,f}(x_i) - \int g_{i,f}(x_i) \, dv_n(x_i) \right|$$

$$+ h_n \sum_{i,j=1: i<j}^{3} \left| \int \int g_{i,j,f}(x_i,x_j) \, dv_n(x_i) \, dv_n(x_j) \right|$$

$$+ h_n^2 \left| \int \int \int g(x_1,x_2,x_3) \, dv_n(x_1) \, dv_n(x_2) \, dv_n(x_3) \right|,$$

where $g_{i,j,f}(x_i,x_j) := \int g(x_1,x_2,x_3) \, df(x_k)$, $i, j, k \in \{1,2,3\}$, $i < j$, $k \neq i$, $k \neq j$. To obtain $S_1(n) \to 0$, it suffices to assume that the functions $g_{i,f}$ satisfy the first part of Assumption 2.1(b). To ensure that $h_n^{-1} S_2(n)$ is bounded above, it suffices to assume that, similar to the case $d = 2$, the functions $g_{i,j,f}$ satisfy Assumption 2.1(a) (with g replaced by $g_{i,j,f}$). Assuming that for every fixed x_2, x_3 the function $g_{x_2,x_3}(\cdot) := g(\cdot, x_2, x_3)$, lies in $\mathbb{BV}_{\text{loc}} \cap \mathbb{D}_{-\lambda'}$, and that $(x_2, x_3) \mapsto \int \phi_{-\lambda}(x_1) |dg_{x_2,x_3}|(x_1)$ lies in $\mathbb{M}_{(-\lambda',-\lambda')}$ (cf. Assumption 2.1(a)), ensures that $h_n^{-2} S_3(n)$ is bounded above. Thus, $S_1(n) + S_2(n) + S_3(n) \to 0$.

Finally, we note that the case $d = 1$ is even easier. Here, we only need to assume $g \in \mathbb{BV}_{\text{loc}} \cap \mathbb{D}_{-\lambda'}$ (instead of Assumptions 2.1(a)–(b)) and to replace (9) by $\dot{U}_f(v) := -\int v(x) \, dg(x)$.

Appendix A: Jordan decomposition of functions in $\mathbb{BV}_{\textbf{loc}}$

Recall that for $\psi \in \mathbb{BV}_{\text{loc}}$ and $c \in \mathbb{R}$, the Jordan decomposition of ψ centered at c,

$$\psi = \psi(c) + \psi_c^+ - \psi_c^-, \tag{18}$$

is characterized as follows: ψ_c^+ and ψ_c^- are the unique nondecreasing functions satisfying

$$\psi_c^+(x) = V^+([c,x], \psi), \qquad \psi_c^-(x) = V^-([c,x], \psi) \qquad \forall x \geq c, \tag{19}$$

$$\psi_c^+(x) = -V^+([x,c], \psi), \qquad \psi_c^-(x) = -V^-([x,c], \psi) \qquad \forall x < c, \tag{20}$$

where $V^+([a,b], \psi)$ and $V^-([a,b], \psi)$ denote the positive and the negative variation of ψ on the interval $[a,b]$, respectively. For details see, for example, [15], page 34. In our applications, we are mainly concerned with the positive measures $d\psi_c^+$ and $d\psi_c^-$ induced by ψ_c^+ and ψ_c^-, respectively (provided ψ_c^+ and ψ_c^- are right-continuous). The following lemma shows that $d\psi_c^+$ and $d\psi_c^-$ are independent of c, although ψ_c^+ and ψ_c^- typically do depend on c. In particular, the definition $|d\psi| := d\psi_c^+ + d\psi_c^-$ of the absolute value measure $|d\psi|$ is independent of c.

Lemma A.1. *Let $\psi \in \mathbb{BV}_{\mathrm{loc}}$ and $c \in \mathbb{R}$. Then ψ_c^+, ψ_c^- differ from ψ_0^+, ψ_0^- only by constants K_c^+, K_c^-, respectively. In particular, the positive measures $\mathrm{d}\psi_c^+$ and $\mathrm{d}\psi_c^-$ are independent of c.*

Proof. Let $c > 0$. Then, in view of (19)–(20), we have

$$\psi_0^+(x) = V^+([0, x], \psi) = V^+([0, c], \psi) + V^+([c, x], \psi) = V^+([0, c], \psi) + \psi_c^+(x)$$

for $x \in (c, \infty)$, and similar we obtain $\psi_0^+(x) = V^+([0, c], \psi) + \psi_c^+(x)$ for the cases $x \in [0, c]$ and $x \in (-\infty, 0)$. That is, $\psi_c^+ = \psi_0^+ + K_c^+$ for some constant K_c^+. Analogously, we obtain $\psi_c^+ = \psi_0^+ + K_c^+$ for $c \le 0$, and $\psi_c^- = \psi_0^- + K_c^-$ for $c \le 0$ as well as $c > 0$. $\square$

Appendix B: Integration theoretical auxiliaries

Recall our convention $\int = \int_{(-\infty, \infty)}$ and that $\mathbb{BV}_{\mathrm{loc,d}}$ denotes the space of all càdlàg functions in $\mathbb{BV}_{\mathrm{loc}}$.

Lemma B.1. *Let $u, v \in \mathbb{BV}_{\mathrm{loc,d}}$ such that $\lim_{x \to \pm\infty} u(x)v(x) = c_\pm$ for some constants c_+, $c_- \in \mathbb{R}$. Then, if*

$$\int |v(x-)||\mathrm{d}u|(x) < \infty \quad and \quad \int |u(x)||\mathrm{d}v|(x) < \infty, \tag{21}$$

we have the integration-by-parts formula

$$\int u(x)\,\mathrm{d}v(x) = c_+ - c_- - \int v(x-)\,\mathrm{d}u(x). \tag{22}$$

Proof. If $-\infty < a < b < \infty$, then one can proceed as in the proof of Theorem II.6.11 in [25] to obtain

$$\int_{(a,b]} u(x)\,\mathrm{d}v(x) = u(b)v(b) - u(a)v(a) - \int_{(a,b]} v(x-)\,\mathrm{d}u(x), \tag{23}$$

because $\int_{(a,b]} |v(x-)||\mathrm{d}u|(x) < \infty$ and $\int_{(a,b]} |u(x)||\mathrm{d}v|(x) < \infty$. Now, choosing sequences $(a_n), (b_n) \subset (-\infty, \infty)$ with $a_n \downarrow -\infty$ and $b_n \uparrow \infty$, the statement of the lemma follows from (23), the continuity from below of the finite measures $\int u^+(x)\,\mathrm{d}v^+(x)$, $\int u^-(x)\,\mathrm{d}v^+(x), \dots$ on $(-\infty, \infty)$, and the assumption $\lim_{x \to \pm\infty} u(x)v(x) = c_\pm$. $\square$

Next, we give a sort of Helly–Bray theorem. Recall that $\mathbb{BV}_{1,\mathrm{d}}$ denotes the space of all càdlàg functions on $\overline{\mathbb{R}}$ with variation bounded by 1.

Lemma B.2. *Let $\lambda > \lambda' \ge 0$, let $\psi \in \mathbb{D}_{-\lambda'}$ and suppose that $f, f_1, f_2, \dots \in \mathbb{BV}_{1,\mathrm{d}}$ are nondecreasing and satisfy $\lim_{n \to \infty} \| f_n - f \|_\lambda = 0$. Let $\int \phi_{\lambda'}(x)\,\mathrm{d}f(x) < \infty$ and $\int \phi_{\lambda'}(x)\,\mathrm{d}f_n(x) < \infty$ for every $n \in \mathbb{N}$. Then the integrals $\int \psi(x)\,\mathrm{d}f(x)$ and $\int \psi(x)\,\mathrm{d}f_n(x)$ exist and we have*

$$\lim_{n \to \infty} \int \psi(x)\,\mathrm{d}f_n(x) = \int \psi(x)\,\mathrm{d}f(x).$$

Proof. The first claim follows from

$$\int |\psi(x)| \, df(x) = \int |\psi(x)\phi_{\lambda'}(x)\phi_{-\lambda'}(x)| \, df(x) \leq \|\psi\|_{-\lambda'} \int \phi_{\lambda'}(x) \, df(x)$$

and the analogous bound for $\int |\psi(x)| \, df_n(x)$, $n \in \mathbb{N}$.

Now let us turn to the second claim. Since $\psi\phi_{-\lambda'}$ is a bounded càdlàg function on the *compact* interval $\overline{\mathbb{R}}$, we may and do choose for each $\varepsilon > 0$ a step function $\widetilde{\psi}_\varepsilon \in \mathbb{D}$ with a finite number of jumps and satisfying $\|\psi\phi_{-\lambda'} - \widetilde{\psi}_\varepsilon\|_\infty \leq \varepsilon$. For $\psi_\varepsilon := \widetilde{\psi}_\varepsilon \phi_{\lambda'}$, we thus have $\|\psi - \psi_\varepsilon\|_{-\lambda'} \leq \varepsilon$. Of course,

$$
\begin{aligned}
&\left| \int \psi(x) \, df_n(x) - \int \psi(x) \, df(x) \right| \\
&\leq \left| \int \psi(x) \, d(f_n - f)(x) - \int \psi_\varepsilon(x) \, d(f_n - f)(x) \right| \\
&\quad + \left| \int \psi_\varepsilon(x) \, d(f_n - f)(x) \right| \\
&=: S_1(n, \varepsilon) + S_2(n, \varepsilon).
\end{aligned}
\tag{24}
$$

For the first summand, we obtain

$$
\begin{aligned}
S_1(n, \varepsilon) &= \left| \int \phi_{-\lambda'}(x)\phi_{\lambda'}(x)\psi(x) \, d(f_n - f)(x) \right. \\
&\qquad \left. - \int \phi_{-\lambda'}(x)\phi_{\lambda'}(x)\psi_\varepsilon(x) \, d(f_n - f)(x) \right| \\
&\leq \left(\int \phi_{\lambda'}(x) \, df_n(x) + \int \phi_{\lambda'}(x) \, df(x) \right) \|\psi - \psi_\varepsilon\|_{-\lambda'} \\
&\leq \left(\int \phi_{\lambda'}(x) \, df_n(x) + \int \phi_{\lambda'}(x) \, df(x) \right) \varepsilon \\
&\leq C\varepsilon
\end{aligned}
\tag{25}
$$

for some finite constant $C > 0$ being independent of n and ε. For the last step, we used the assumption $\int \phi_{\lambda'}(x) \, df(x) < \infty$ and the fact that $\sup_{n \in \mathbb{N}} \int \phi_{\lambda'}(x) \, df_n(x) < \infty$. The latter fact is not completely obvious, so that we give the details: Because of $\int \phi_{\lambda'}(x) \, df(x) < \infty$, it is clearly sufficient to show that $\sup_{n \in \mathbb{N}} |\int \phi_{\lambda'}(x) \, d(f - f_n)(x)|$ is bounded above by some finite constant. By our assumptions and the bound (26) below, we can apply the integration by parts formula (22) to the functions $f - f_n$ and $\phi_{\lambda'}$ to obtain

$$\left| \int \phi_{\lambda'}(x) \, d(f - f_n) \right| \leq 2\|f - f_n\|_{\lambda'} + \left| \int (f - f_n)(x-) \, d\phi_{\lambda'}(x) \right|.$$

By our assumptions, the first summand tends to 0 since $\|f_n - f\|_{\lambda'} \le \|f_n - f\|_{\lambda}$. The second summand is less than or equal to $\int |(f - f_n)(x-)||d\phi_{\lambda'}|(x)$ and we have

$$
\int |(f - f_n)(x-)||d\phi_{\lambda'}|(x) = \int |\phi_\lambda(x)(f - f_n)(x-)|\phi_{-\lambda}(x)|d\phi_{\lambda'}|(x)
$$
$$
\le 2\|f - f_n\|_\lambda \int_{\mathbb{R}_+} \phi_{-\lambda}(x)\,d\phi_{\lambda'}(x). \tag{26}
$$

Since $\|f - f_n\|_\lambda \to 0$ by assumption, and $\int_{\mathbb{R}_+} \phi_{-\lambda}(x)\,d\phi_{\lambda'}(x) < \infty$ by $\lambda > \lambda' \ge 0$, the left-hand side of (26) converges to 0. In particular, the left-hand side of (26) is bounded above uniformly in n. This completes the proof of (25).

Now, the second claim of the lemma would follow from (24) and (25) if we could show that $S_2(n, \varepsilon)$ converges to 0 as $n \to \infty$ uniformly in $\varepsilon \in (0, 1]$. By our assumptions and formula (27) below, we can apply the integration by parts formula (22) to obtain

$$
\begin{aligned}
&S_2(n, \varepsilon) \\
&= \left| \int \psi_\varepsilon(x)\phi_{\lambda'}(x)\phi_{-\lambda'}(x)\,d(f_n - f)(x) \right| \\
&\le 2\|\psi_\varepsilon\|_{-\lambda'}\|f_n - f\|_{\lambda'} + \left| \int (f_n - f)(x-)\,d\psi_\varepsilon(x) \right| \\
&\le 2(\|\psi_\varepsilon - \psi\|_{-\lambda'} + \|\psi\|_{-\lambda'})\|f_n - f\|_{\lambda'} + \left| \int (f_n - f)(x-)\,d\psi_\varepsilon(x) \right|.
\end{aligned}
$$

The first summand converges to 0 by our assumptions and $\|\psi_\varepsilon - \psi\|_{-\lambda'} \le \varepsilon \le 1$. Furthermore, the second summand is less than or equal to $\int |(f_n - f)(x-)||d\psi_\varepsilon|(x)$. Recalling $\psi_\varepsilon = \widetilde{\psi}_\varepsilon \phi_{\lambda'}$ and that $\widetilde{\psi}_\varepsilon$ is a step function with a finite number of jumps, we now obtain

$$
\begin{aligned}
&\int |(f_n - f)(x-)||d\psi_\varepsilon|(x) \\
&\le \|\widetilde{\psi}_\varepsilon\|_\infty \int |(f_n - f)(x-)||d\phi_{\lambda'}|(x) \\
&= \|\widetilde{\psi}_\varepsilon\|_\infty \int |(f_n - f)(x-)\phi_\lambda(x)|\phi_{-\lambda}(x)|d\phi_{\lambda'}|(x) \\
&\le 2(\|\psi\phi_{-\lambda'}\|_\infty + 1)\|f_n - f\|_\lambda \int_{\mathbb{R}_+} \phi_{-\lambda}(x)\,d\phi_{\lambda'}(x),
\end{aligned} \tag{27}
$$

and this expression converges to 0 because $\|f_n - f\|_\lambda \to 0$ and $\lambda > \lambda' \ge 0$. That is, $S_2(n, \varepsilon)$ indeed converges to 0 as $n \to \infty$ uniformly in $\varepsilon \in (0, 1]$. $\qquad \square$

Acknowledgements

References

[1] Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T. and Weiss, B. (1996). Strong laws for *L*- and *U*-statistics. *Trans. Amer. Math. Soc.* **348** 2845–2866. MR1363941

[2] Beutner, E. and Zähle, H. (2010). A modified functional delta method and its application to the estimation of risk functionals. *J. Multivariate Anal.* **101** 2452–2463. MR2719874

[3] Bradley, R.C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144 (electronic). MR2178042

[4] Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *J. Econometrics* **130** 307–335. MR2211797

[5] Dehling, H. (2006). Limit theorems for dependent *U*-statistics. In *Dependence in Probability and Statistics*. *Lecture Notes in Statist.* **187** 65–86. New York: Springer. MR2283250

[6] Dehling, H. and Wendler, M. (2010). Central limit theorem and the bootstrap for *U*-statistics of strongly mixing data. *J. Multivariate Anal.* **101** 126–137. MR2557623

[7] Denker, M. (1985). *Asymptotic Distribution Theory in Nonparametric Statistics*. *Advanced Lectures in Mathematics*. Braunschweig: Friedr. Vieweg & Sohn. MR0889896

[8] Denker, M. and Keller, G. (1986). Rigorous statistical procedures for data from dynamical systems. *J. Statist. Phys.* **44** 67–93. MR0854400

[9] Dewan, I. and Prakasa Rao, B.L.S. (2001). Asymptotic normality for *U*-statistics of associated random variables. *J. Statist. Plann. Inference* **97** 201–225. MR1861150

[10] Dewan, I. and Prakasa Rao, B.L.S. (2002). Central limit theorem for *U*-statistics of associated random variables. *Statist. Probab. Lett.* **57** 9–15. MR1911807

[11] Doukhan, P. (1994). *Mixing*: *Properties and Examples*. *Lecture Notes in Statistics* **85**. New York: Springer. MR1312160

[12] Fernholz, L.T. (1983). *Von Mises Calculus for Statistical Functionals*. *Lecture Notes in Statistics* **19**. New York: Springer. MR0713611

[13] Gill, R.D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. I. *Scand. J. Statist.* **16** 97–128. MR1028971

[14] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics* **19** 293–325. MR0026294

[15] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd ed. *Probability and Its Applications* (*New York*). New York: Springer. MR1876169

[16] Lee, A.J. (1990). *U-statistics*: *Theory and Practice*. *Statistics*: *Textbooks and Monographs* **110**. New York: Marcel Dekker Inc. MR1075417

[17] Lindner, A.M. (2008). Stationarity, Mixing, Distributional Properties and Moments of GARCH(p, q) Processes. In *Handbook of Financial Time Series* (T.G. Andersen, R.A. Davis, J.-P. Kreiß, T. Mikosch, eds.) 41–69. Berlin: Springer.

[18] Reeds, J.A. (1976). On the definition of von Mises functionals. Research Report S-44. Dept. Statistics, Harvard Univ.

[19] Sen, P.K. (1972/73). Limiting behavior of regular functionals of empirical distributions for stationary mixing processes. *Z. Wahrsch. Verw. Gebiete* **25** 71–82. MR0329003

[20] Sen, P.K. (1981). *Sequential Nonparametrics*: *Invariance principles and statistical inference*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0633884

[21] Sen, P.K. (1991). Introduction to Hoeffding (1948): A class of statistics with asymptotically normal distribution. In *Breakthroughs in Statistics* (S. Kotz and N.L. Johnson, eds.) **I** 299–307. Springer, New York.

[22] Sen, P.K. (1996). Statistical functionals, Hadamard differentiability and martingales. In *Probability Models and Statistics* 29–47. New Delhi: New Age. MR1410514

[23] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0595165

[24] Shao, Q.M. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann. Probab.* **24** 2098–2127. MR1415243

[25] Shiryaev, A.N. (1996). *Probability*, 2nd ed. *Graduate Texts in Mathematics* **95**. New York: Springer. MR1368405

[26] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. New York: Wiley. MR0838963

[27] von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statistics* **18** 309–348. MR0022330

[28] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247

[29] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. New York: Springer. MR1385671

[30] Wu, W.B. (2008). Empirical processes of stationary sequences. *Statist. Sinica* **18** 313–333. MR2384990

[31] Yoshihara, K.i. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Z. Wahrsch. Verw. Gebiete* **35** 237–252. MR0418179

[32] Zähle, H. (2011). Rates of almost sure convergence of plug-in estimates for distortion risk measures. *Metrika* **74** 267–285.