SFB 823

A note on conditional versus joint unconditional weak convergence in bootstrap consistency results

Axel Bücher, Ivan Kojadinovic

Nr. 13/2017

Discussion Paper

A note on conditional versus joint unconditional weak convergence in bootstrap consistency results

Axel Bücher [*] and Ivan Kojadinovic [†]

June 4, 2017

Abstract

The consistency of a bootstrap or resampling scheme is classically validated by weak convergence of conditional laws. However, when working with stochastic processes in the space of bounded functions and their weak convergence in the Hoffmann-Jørgensen sense, an obstacle occurs: due to possible non-measurability, neither laws nor conditional laws are well-defined. Starting from an equivalent formulation of weak convergence based on the bounded Lipschitz metric, a classical circumvent is to formulate bootstrap consistency in terms of the latter distance between what might be called a *conditional law* of the (non-measurable) bootstrap process and the law of the limiting process. The main contribution of this note is to provide an equivalent formulation of bootstrap consistency in the space of bounded functions which is more intuitive and easy to work with. Essentially, the equivalent formulation consists of (unconditional) weak convergence of the original process jointly with an arbitrary large number of bootstrap replicates. As a by-product, we provide two equivalent formulations of bootstrap consistency for $\mathbb{R}^d$-valued statistics: the first in terms of (unconditional) weak convergence of the statistic jointly with its bootstrap replicates, the second in terms of convergence in probability of the empirical distribution function of the bootstrap replicates. Finally, the asymptotic validity of bootstrap-based confidence intervals and tests is briefly revisited, with particular emphasis on the, in practice unavoidable, Monte Carlo approximation of conditional quantiles.

Keywords: Bootstrap; conditional weak convergence; confidence Intervals; resampling; stochastic processes; weak convergence.

MSC 2010: 62E20; 62G09.

1 Introduction

It is not uncommon in statistical problems that the limiting distribution of a statistic of interest be intractable. To carry out inference on the underlying quantity, one possibility consists of using a *bootstrap* or *resampling scheme*. Ideally, prior to its use, its *consistency* or *asymptotic validity* should be mathematically demonstrated. For a real or vector-valued statistic $\boldsymbol{S}_n$, the latter classically consists of establishing weak convergence of certain conditional laws. Specifically, a resampling scheme can be considered asymptotically consistent if the distribution function (d.f.) of a *bootstrap replicate* of $\boldsymbol{S}_n$ *given* the available observations is shown to converge in probability

[*]Ruhr-Universität Bochum, Fakultät für Mathematik, Universitätsstr. 150, 44780 Bochum, Germany. E-mail: axel.buecher@rub.de

[†]CNRS / Université de Pau et des Pays de l'Adour, Laboratoire de mathématiques et applications – IPRA, UMR 5142, B.P. 1155, 64013 Pau Cedex, France. E-mail: ivan.kojadinovic@univ-pau.fr

to the d.f. of $\boldsymbol{S}_n$; see, for instance, Bickel and Freedman (1981), van der Vaart (1998, Chapter 23), Horowitz (2001) and the references therein, or Assertion (*b*) in Lemma 2.2 below. A first contribution of this note is to show that, under minimal conditions, the aforementioned weak convergence of conditional laws is actually equivalent to the (unconditional) weak convergence of $\boldsymbol{S}_n$ jointly with an arbitrary large number of bootstrap replicates to independent copies of the same limit. As we shall see in the forthcoming paragraphs, this equivalent formulation is of particular interest when $\boldsymbol{S}_n$ is a "sufficiently smooth" functional of a certain stochastic process (e.g., the general empirical process – see Chapter 2 in van der Vaart and Wellner, 2000) and the resampling scheme for $\boldsymbol{S}_n$ results from a similar resampling scheme at the level of the stochastic process. A third equivalent formulation of the consistency of a bootstrap for $\boldsymbol{S}_n$ is also provided. It roughly states that the empirical d.f. of the bootstrap replicates converges in probability to the unobservable d.f. of $\boldsymbol{S}_n$ as the number of replicates and the sample size increase (see also Beran et al., 1987, Section 4, for a similar result). The latter is particularly meaningful given that most applications of resampling involve at some point approximating the unobservable distribution of $\boldsymbol{S}_n$ by the empirical distribution of bootstrap replicates.

As mentioned above, in many situations, the statistic of interest $\boldsymbol{S}_n$ is a "sufficiently smooth" functional of a certain stochastic process. The latter fact is the main motivation for studying resampling schemes at the stochastic process level. For the general empirical process based on independent and identically distributed (i.i.d.) observations for instance, such an investigation is carried out in van der Vaart and Wellner (2000, Section 3.6) for the so-called *empirical bootstrap* and various other *exchangeable bootstraps*. Following Giné and Zinn (1990), the consistency of a resampling scheme is defined therein by the requirement that the bounded Lipschitz distance between the candidate limiting law and a suitable adaptation of what might be called a conditional law of the bootstrap replicate (even though the latter does not exist in the classical sense due to non-measurability) converges to zero in outer probability. The appeal of working at the stochastic process level then arises from the fact that such bootstrap consistency results can be transferred to the statistic level by means of appropriate extensions of the continuous mapping theorem and the functional delta method.

It may however be argued that the aforementioned generalization of the classical conditional formulation of bootstrap consistency is unintuitive and complicated to use given the subtlety of the underlying mathematical concepts (in particular, relying on "conditional laws" of non-measurable maps). The latter seems all the more true for instance for empirical processes based on estimated or serially dependent observations (see, e.g., Rémillard and Scaillet, 2009; Segers, 2012; Bücher and Kojadinovic, 2016a). The main contribution of this note is to show that the conditional formulation is actually equivalent to the (unconditional) weak convergence of the initial stochastic process jointly with an arbitrary large number of bootstrap replicates. From a practical perspective, using the latter unconditional formulation may have two important advantages. First and most importantly, it may be easier to prove in certain situations than the conditional formulation. For this reason, it was for instance used, as explained above, for empirical processes based on estimated or serially dependent observations; see also Section 3 below for additional references. Second, the unconditional formulation may be transferable to the statistic level for a slightly larger class of functionals of the stochastic process under consideration. The latter follows for instance from the fact that continuous mapping theorems *for the bootstrap*, that is, adapted to the conditional formulation, require more than just continuity of the map that transforms the stochastic process into the statistic of interest (see, e.g., Kosorok, 2008, Section 10.1.4). Furthermore, there does not seem to exist an *extended* continuous mapping theorem (see, e.g., van der Vaart and Wellner, 2000, Theorem 1.11.1) *for the bootstrap*. Once the unconditional formulation is transferred to $\mathbb{R}^d$, the classical conditional statement immediately follows by the equivalence at the vector-valued statistic level mentioned above. Finally, let us mention that the equivalence at the stochastic process level is well-known for the special case of the *multiplier bootstrap* of the general empirical process based on i.i.d. observations using results of van der Vaart and Wellner (2000, Section 2.9). As such, our proven equivalence at the stochastic process level can be seen as an extension of the latter work.

As an illustration of our results, we revisit the fact that bootstrap consistency implies that bootstrap-based confidence intervals are asymptotically valid in terms of coverage and that bootstrap-based test hold their level asymptotically; see, for instance, van der Vaart (1998, Lemma 23.3) for a related result and Horowitz (2001, Sections 3.3 and 3.4) for more specialized and deeper results. In particular, we provide results which explicitly take into account that (unobservable) conditional quantiles must be approximated by Monte Carlo in practice.

Finally, we would like to stress that the asymptotic results in this note are all of first order. Higher order correctness of a resampling scheme (usually considered for real-valued statistics) may still be important in small samples. The reader is referred to Hall (1992) for more details.

This note is organized as follows. The equivalence between the aforementioned formulations of asymptotic validity of bootstraps of vector-valued statistics is proved in Section 2. Section 3 states conditions under which the results of Section 2 extend to stochastic processes with bounded sample paths. In Section 4, it is formally verified that, as expected, bootstrap consistency implies asymptotic validity of bootstrap-based confidence intervals and tests. A summary of results and concluding remarks are given in the last section.

In the rest of the document, vectors are denoted in bold, the arrow '$\leadsto$' denotes weak convergence, while the arrows '$\xrightarrow{\text{a.s.}}$' and '$\xrightarrow{\text{P}}$' denote almost sure convergence and convergence in probability, respectively.

2 Equivalent statements of bootstrap consistency in $\mathbb{R}^d$

The generic setup considered in this section is as follows. The available data will be denoted by $\boldsymbol{X}_n$. No assumptions are made on the n available observations: they could be univariate or multivariate, serially independent or dependent, etc. The $\mathbb{R}^d$-valued statistic computed from the data $\boldsymbol{X}_n$ will be denoted by $\boldsymbol{S}_n = \boldsymbol{S}_n(\boldsymbol{X}_n)$. *Bootstrap replicates* of $\boldsymbol{S}_n$ on which inference will be based will be denoted by $\boldsymbol{S}_n^{(1)} = \boldsymbol{S}_n^{(1)}(\boldsymbol{X}_n, \boldsymbol{W}_n^{(1)})$, $\boldsymbol{S}_n^{(2)} = \boldsymbol{S}_n^{(2)}(\boldsymbol{X}_n, \boldsymbol{W}_n^{(2)})$, $\dots$, where $\boldsymbol{W}_n^{(1)}$, $\boldsymbol{W}_n^{(2)}$, $\dots$, are identically distributed and represent additional sources of randomness such that $\boldsymbol{S}_n^{(1)}, \boldsymbol{S}_n^{(2)}, \dots$ are independent conditionally on $\boldsymbol{X}_n$.

The previous setup is general enough to encompass most if not all types of resampling procedures. For instance, the classical *empirical* (multinomial) bootstrap of Efron (1979) based on resampling with replacement from some original i.i.d. data set $\boldsymbol{X}_n = (X_1, \dots, X_n)$ can be obtained by letting the $\boldsymbol{W}_n^{(i)} = (W_{n1}^{(i)}, \dots, W_{nn}^{(i)})$ be i.i.d. multinomially distributed with parameter $(n, 1/n, \dots, 1/n)$. Indeed, for fixed $i \in \mathbb{N}$, the sample $\boldsymbol{X}_n^* = (X_1^*, \dots, X_n^*)$ constructed by including the jth original observation X_j exactly $W_{nj}^{(i)}$ times, $j \in \{1, \dots, n\}$, may be identified with a sample being drawn with replacement from the original observations. Many other resampling schemes are included as well: block bootstraps for time series such as the one of Künsch (1989), (possibly dependent) *multiplier* or *wild* bootstraps (see, e.g., Shao, 2010) or the *parametric* bootstrap (see, e.g., Stute et al., 1993; Genest and Rémillard, 2008). For all but the last mentioned resampling scheme, $\boldsymbol{W}_n^{(1)}$, $\boldsymbol{W}_n^{(2)}$, $\dots$, could be interpreted as i.i.d. vectors of *bootstrap weights*, independent of $\boldsymbol{X}_n$. Several examples of such weights when $\boldsymbol{X}_n$ corresponds to n i.i.d. observations are given for instance in van der Vaart and Wellner (2000, Section 3.6.2).

The previous setup is summarized in the following basic assumption.

Condition 2.1 ($\mathbb{R}^d$-valued resampling mechanism)**.** *Let* $\boldsymbol{X}_n : \Omega \to \mathcal{X}_n$ *be a random variable in* *some measurable space* $\mathcal{X}_n$ *and let* $\boldsymbol{S}_n = \boldsymbol{S}_n(\boldsymbol{X}_n)$ *be an* $\mathbb{R}^d$*-valued statistic such that* $\boldsymbol{S}_n \rightsquigarrow \boldsymbol{S}$ *as* $n \to \infty$*, where* $\boldsymbol{S}$ *is absolutely continuous. Furthermore, let* $\boldsymbol{W}_n^{(i)} : \Omega \to \mathcal{W}_n$*,* $i \in \mathbb{N}$*, denote identically distributed random variables in some measurable space* $\mathcal{W}_n$ *and let* $\boldsymbol{S}_n^{(i)} = \boldsymbol{S}_n^{(i)}(\boldsymbol{X}_n, \boldsymbol{W}_n^{(i)})$*,* $i \in \mathbb{N}$*, be* $\mathbb{R}^d$*-valued statistics (to be considered as bootstrap replicates of* $\boldsymbol{S}_n$*) that are independent conditionally on* $\boldsymbol{X}_n$*.*

Subsequently, let $\boldsymbol{S}_n'$ denote a generic copy of $\boldsymbol{S}_n^{(1)}$. Furthermore, let $\mathrm{P}^{\boldsymbol{S}_n' \mid \boldsymbol{X}_n}$ denote (a regular version of) the conditional distribution of $\boldsymbol{S}_n'$ given $\boldsymbol{X}_n$. For arbitrary real-valued functions h such that $\mathrm{E}|h(\boldsymbol{S}_n')| < \infty$, conditional expectations $\mathrm{E}\{h(\boldsymbol{S}_n') \mid \boldsymbol{X}_n\}$ are always to be understood as integration of $h(\boldsymbol{S}_n')$ with respect to $\mathrm{P}^{\boldsymbol{S}_n' \mid \boldsymbol{X}_n}$ (Kallenberg, 2002, Theorem 6.4). Note that this convention makes the variable involving the supremum in Assertion (b) of the forthcoming lemma well-defined.

Lemma 2.2 below is one of the main result of this note and essentially shows that the unconditional weak convergence of the statistic jointly with its bootstrap replicates is equivalent to the convergence in probability of the conditional law of a bootstrap replicate. The latter (with convergence in probability possibly replaced by almost sure convergence) is the classical mathematical definition of the asymptotic validity of a resampling scheme. A third equivalent formulation, of interest for applications, is also provided. But first recall that the Kolmogorov distance d_K between arbitrary measures P, Q on $\mathbb{R}^d$ is defined by

$$d_K(P, Q) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} \big| P\{(-\boldsymbol{\infty}, \boldsymbol{x}]\} - Q\{(-\boldsymbol{\infty}, \boldsymbol{x}]\} \big|.$$

Lemma 2.2 (Equivalence of unconditional and conditional formulations)**.** *Suppose that Condition 2.1 is met for some arbitrary* $d \in \mathbb{N}$*. Then, the following three assertions are equivalent:*

(a) For any $M \in \mathbb{N}$*, as* $n \to \infty$*,*

$$(\boldsymbol{S}_n, \boldsymbol{S}_n^{(1)}, \dots, \boldsymbol{S}_n^{(M)}) \rightsquigarrow (\boldsymbol{S}, \boldsymbol{S}^{(1)}, \dots, \boldsymbol{S}^{(M)}),$$

where $\boldsymbol{S}, \boldsymbol{S}^{(1)}, \dots, \boldsymbol{S}^{(M)}$ *are i.i.d.*

(b) As $n \to \infty$*, we have*

$$d_K\left(\mathrm{P}^{\boldsymbol{S}_n^{(1)} \mid \boldsymbol{X}_n}, \mathrm{P}^{\boldsymbol{S}_n}\right) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} \left| \mathrm{P}(\boldsymbol{S}_n^{(1)} \le \boldsymbol{x} \mid \boldsymbol{X}_n) - \mathrm{P}(\boldsymbol{S}_n \le \boldsymbol{x}) \right| \overset{\mathrm{P}}{\to} 0.$$

(c) As $n, M \to \infty$*, we have*

$$d_K\left(\tfrac{1}{M}\textstyle\sum_{i=1}^M \delta_{\boldsymbol{S}_n^{(i)}}, \mathrm{P}^{\boldsymbol{S}_n}\right) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} \left| \frac{1}{M} \sum_{i=1}^M \mathbf{1}(\boldsymbol{S}_n^{(i)} \le \boldsymbol{x}) - \mathrm{P}(\boldsymbol{S}_n \le \boldsymbol{x}) \right| \overset{\mathrm{P}}{\to} 0.$$

Before providing a proof of this lemma, let us give an interpretation of the three assertions. The intuition behind Assertion (a) is that a resampling scheme should be considered consistent if the bootstrap replicates $\boldsymbol{S}_n^{(1)}, \boldsymbol{S}_n^{(2)}, \dots$ behave approximately as independent copies of $\boldsymbol{S}_n$, the more so that n is large. Assertion (b) translates mathematically the idea that a resampling scheme should be considered valid if the distribution of a bootstrap replicate given the data is close to the distribution of the original statistic $\boldsymbol{S}_n$, the more so that n is large. Assertion (c) can be regarded as an empirical analogue of Assertion (b): the unobservable conditional d.f. of a bootstrap replicate is replaced by the empirical d.f. of a sample of M bootstrap replicates, providing an approximation of the d.f. of $\boldsymbol{S}_n$ that improves as n, M increase.

Assertion (b) is known to hold for many statistics and resampling schemes, possibly as a consequence of general consistency results such as the one of Beran and Ducharme (1991) (see also Horowitz, 2001, Section 2.1). Assertion (a) is substantially less frequently encountered in the literature and appears essentially as a consequence of a similar assertion at a stochastic process level; see Lemma 3.1 in Section 3 and the references therein.

As we continue, for $n, M \in \mathbb{N}$ and $\boldsymbol{x} \in \mathbb{R}^d$, we will frequently use the following notation:

$$\begin{aligned} F_n^M(\boldsymbol{x}) &= \tfrac{1}{M}\textstyle\sum_{i=1}^M \mathbf{1}(\boldsymbol{S}_n^{(i)} \le \boldsymbol{x}), \qquad & F_n(\boldsymbol{x}) &= \mathrm{P}(\boldsymbol{S}_n^{(1)} \le \boldsymbol{x} \mid \boldsymbol{X}_n), \\ F^M(\boldsymbol{x}) &= \tfrac{1}{M}\textstyle\sum_{i=1}^M \mathbf{1}(\boldsymbol{S}^{(i)} \le \boldsymbol{x}), \qquad & F(\boldsymbol{x}) &= \mathrm{P}(\boldsymbol{S} \le \boldsymbol{x}). \end{aligned} \tag{2.1}$$

Proof of Lemma 2.2. The following two simple consequences of Condition 2.1 will be frequently used throughout the proof: First,

$$\sup_{\boldsymbol{x} \in \mathbb{R}^d} |\mathrm{P}(\boldsymbol{S}_n \le \boldsymbol{x}) - F(x)| \to 0 \qquad \text{as } n \to \infty, \tag{2.2}$$

as a consequence of the weak convergence of $\boldsymbol{S}_n$ to $\boldsymbol{S}$ and continuity of F, see Lemma 2.11 in van der Vaart (1998). Second, for any $n, M \in \mathbb{N}$, $\boldsymbol{x} \in \mathbb{R}^d$ and $\varepsilon > 0$,

$$\begin{aligned} \mathrm{P}\big\{|F_n^M(\boldsymbol{x}) - F_n(\boldsymbol{x})| \ge \varepsilon\big\} &= \mathrm{E}\Big[\mathrm{P}\big\{|F_n^M(\boldsymbol{x}) - \mathrm{E}1(\boldsymbol{S}_n^{(1)} \le \boldsymbol{x})| \ge \varepsilon \mid \boldsymbol{X}_n\big\}\Big] \\ &\le \frac{1}{\varepsilon^2 M^2}\mathrm{E}\left[\mathrm{Var}\left\{\textstyle\sum_{i=1}^M \mathbf{1}(\boldsymbol{S}_n^{(i)} \le \boldsymbol{x}) \,\middle|\, \boldsymbol{X}_n\right\}\right] \\ &\le \frac{1}{\varepsilon^2 M}\mathrm{E}\Big[\mathrm{Var}\big\{\mathbf{1}(\boldsymbol{S}_n^{(1)} \le \boldsymbol{x}) \mid \boldsymbol{X}_n\big\}\Big] \le \frac{1}{\varepsilon^2 M}, \end{aligned} \tag{2.3}$$

by Chebyshev's inequality for conditional probabilities and using the fact that $\boldsymbol{S}_n^{(1)}, \dots, \boldsymbol{S}_n^{(M)}$ are identically distributed and independent conditionally on $\boldsymbol{X}_n$.

$(a) \Rightarrow (b)$: Let $M \in \mathbb{N}$, $\boldsymbol{x} \in \mathbb{R}^d$ and $\psi_{M,\boldsymbol{x}} : (\mathbb{R}^d)^M \to \mathbb{R}$ be the map defined by

$$\psi_{M,\boldsymbol{x}}(\boldsymbol{s}_1, \dots, \boldsymbol{s}_M) = \frac{1}{M}\sum_{i=1}^M \mathbf{1}(\boldsymbol{s}_i \le \boldsymbol{x}), \qquad (\boldsymbol{s}_1, \dots, \boldsymbol{s}_M) \in (\mathbb{R}^d)^M.$$

Note that $\psi_{M,\boldsymbol{x}}$ is continuous at any point $(\boldsymbol{s}_1, \dots, \boldsymbol{s}_M) \in (\mathbb{R}^d)^M$ such that, for all $i \in \{1, \dots, M\}$ and all $j \in \{1, \dots, d\}$, $s_{ij} \ne x_j$. Assertion (a), together with the fact that the weak limit $(\boldsymbol{S}, \boldsymbol{S}^{(1)}, \dots, \boldsymbol{S}^{(M)})$ is absolutely continuous and the continuous mapping theorem, implies that $F_n^M(\boldsymbol{x}) \rightsquigarrow F^M(\boldsymbol{x})$ as $n \to \infty$. Next, the strong law of large numbers implies that, for any $\boldsymbol{x} \in \mathbb{R}^d$, $F^M(\boldsymbol{x}) \xrightarrow{\text{a.s.}} F(\boldsymbol{x})$ as $M \to \infty$. Finally, (2.3) implies that $\lim_{M\to\infty} \limsup_{n\to\infty} \mathrm{P}\big\{|F_n^M(\boldsymbol{x}) - F_n(\boldsymbol{x})| \ge \varepsilon\big\} = 0$. Combining the three previous results with Theorem 3.2 in Billingsley (1999) implies that, for any $\boldsymbol{x} \in \mathbb{R}^d$, $F_n(\boldsymbol{x}) \rightsquigarrow F(\boldsymbol{x})$ as $n \to \infty$, and, since $F(\boldsymbol{x})$ is non-random, the convergence takes place in probability. Absolute continuity of $\boldsymbol{S}$ and Problem 23.1 in van der Vaart (1998), see also Lemma 2.11 in that reference, then imply that $\sup_{\boldsymbol{x} \in \mathbb{R}^d} |F_n(\boldsymbol{x}) - F(\boldsymbol{x})|$ converges to zero in probability as $n \to \infty$. Assertion (b) finally follows from (2.2).

$(b) \Rightarrow (a)$: Fix $M \in \mathbb{N}$ an let $\boldsymbol{S}^{(1)}, \dots, \boldsymbol{S}^{(M)}$ be i.i.d. copies of $\boldsymbol{S}$. From Condition 2.1, for any $n \in \mathbb{N}$, $\boldsymbol{S}_n^{(1)}, \dots, \boldsymbol{S}_n^{(M)}$ are identically distributed. Hence, from (b) and (2.2), for any $i \in \{1, \dots, M\}$,

$$\sup_{\boldsymbol{x} \in \mathbb{R}^d} |\mathrm{P}(\boldsymbol{S}_n^{(i)} \le \boldsymbol{x} \mid \boldsymbol{X}_n) - \mathrm{P}(\boldsymbol{S}^{(i)} \le \boldsymbol{x})| \xrightarrow{\mathrm{P}} 0, \qquad \text{as } n \to \infty. \tag{2.4}$$

To prove (a), since the weak limit $\boldsymbol{S}$ is assumed to be absolutely continuous, it suffices to show that, for any $(\boldsymbol{x}, \boldsymbol{x}_1, \dots, \boldsymbol{x}_M) \in (\mathbb{R}^d)^{M+1}$,

$$\mathrm{P}(\boldsymbol{S}_n \le \boldsymbol{x}, \boldsymbol{S}_n^{(1)} \le \boldsymbol{x}_1, \dots, \boldsymbol{S}_n^{(M)} \le \boldsymbol{x}_M) \to \mathrm{P}(\boldsymbol{S} \le \boldsymbol{x})\mathrm{P}(\boldsymbol{S}^{(1)} \le \boldsymbol{x}_1) \cdots \mathrm{P}(\boldsymbol{S}^{(M)} \le \boldsymbol{x}_M)$$

as $n \to \infty$. Let $(\boldsymbol{x}, \boldsymbol{x}_1, \dots, \boldsymbol{x}_M) \in (\mathbb{R}^d)^{M+1}$. Using the fact that $\boldsymbol{S}_n$ is solely a function of $\boldsymbol{X}_n$ and that $\boldsymbol{S}_n^{(1)}, \dots, \boldsymbol{S}_n^{(M)}$ are independent conditionally on $\boldsymbol{X}_n$, the probability on the left of the previous display can be rewritten as

$$\begin{aligned}
&\mathrm{E}\left\{\mathrm{P}(\boldsymbol{S}_n \le \boldsymbol{x}, \boldsymbol{S}_n^{(1)} \le \boldsymbol{x}_1, \dots, \boldsymbol{S}_n^{(M)} \le \boldsymbol{x}_M \mid \boldsymbol{X}_n)\right\} \\
&\quad = \mathrm{E}\left\{\mathbf{1}(\boldsymbol{S}_n \le \boldsymbol{x})\mathrm{P}(\boldsymbol{S}_n^{(1)} \le \boldsymbol{x}_1, \dots, \boldsymbol{S}_n^{(M)} \le \boldsymbol{x}_M \mid \boldsymbol{X}_n)\right\} \\
&\quad = \mathrm{E}\left\{\mathbf{1}(\boldsymbol{S}_n \le \boldsymbol{x})\mathrm{P}(\boldsymbol{S}_n^{(1)} \le \boldsymbol{x}_1 \mid \boldsymbol{X}_n) \cdots \mathrm{P}(\boldsymbol{S}_n^{(M)} \le \boldsymbol{x}_M \mid \boldsymbol{X}_n)\right\}.
\end{aligned}$$

From (2.4) and the dominated convergence theorem (for convergence in probability), it follows that the expression on the right-hand side of this display converges to $\mathrm{P}(\boldsymbol{S} \le \boldsymbol{x})\mathrm{P}(\boldsymbol{S}^{(1)} \le \boldsymbol{x}_1) \dots \mathrm{P}(\boldsymbol{S}^{(M)} \le \boldsymbol{x}_M)$ as $n \to \infty$.

$(b) \Rightarrow (c)$: As a consequence of (2.3), for any $\boldsymbol{x} \in \mathbb{R}^d$ and $n \in \mathbb{N}$,

$$|F_n^M(\boldsymbol{x}) - F_n(\boldsymbol{x})| \xrightarrow{\mathrm{P}} 0 \qquad \text{as } M \to \infty. \tag{2.5}$$

Combining this convergence with (b) and (2.2), we immediately obtain that, for any $\boldsymbol{x} \in \mathbb{R}^d$, $|F_n^M(\boldsymbol{x}) - F(\boldsymbol{x})|$ converges in probability to zero as $n, M \to \infty$. By a (simple) extension of Problem 23.1 in van der Vaart (1998), we obtain that the latter convergence holds uniformly in $\boldsymbol{x} \in \mathbb{R}^d$. The desired assertion finally follows by (2.2).

$(c) \Rightarrow (b)$: The assertion follows pointwise by combining (c) with (2.5). Uniformity in $\boldsymbol{x} \in \mathbb{R}^d$ is a consequence of (2.2) and, again, of Problem 23.1 in van der Vaart (1998). ∎

3 Extension to stochastic processes with bounded sample paths

As in the previous section, let $\boldsymbol{X}_n$ be some data formally seen as a random variable in some measurable space $\mathcal{X}_n$. Furthermore, let T denote an arbitrary non-empty set and let $\ell^\infty(T)$ denote the set of real-valued bounded functions on T equipped with the supremum distance. In this section, instead of $\mathbb{R}^d$-valued statistics computed from $\boldsymbol{X}_n$, we are interested in stochastic processes $\mathbb{G}_n = \mathbb{G}_n(\boldsymbol{X}_n)$ on T constructed from $\boldsymbol{X}_n$. It is assumed that every *sample path* $t \mapsto \mathbb{G}_n(t, \boldsymbol{X}_n(\omega))$ is a bounded function so that $\mathbb{G}_n$ may formally be regarded as a map from the underlying probability space Ω into $\ell^\infty(T)$ without however imposing any measurability conditions. We additionally suppose that, as $n \to \infty$, $\mathbb{G}_n$ converges weakly in $\ell^\infty(T)$ to some tight, Borel measurable stochastic process $\mathbb{G}$ in the sense of Hoffmann-Jørgensen (see, e.g., van der Vaart and Wellner, 2000, Section 1.3) (which in fact implies that $\mathbb{G}_n$ is asymptotically measurable). Extending the setting of Section 2, we further assume that $\mathbb{G}_n^{(1)} = \mathbb{G}_n^{(1)}(\boldsymbol{X}_n, \boldsymbol{W}_n^{(1)}), \mathbb{G}_n^{(2)} = \mathbb{G}_n^{(2)}(\boldsymbol{X}_n, \boldsymbol{W}_n^{(2)}), \dots$ are *bootstrap replicates* of $\mathbb{G}_n$, that is, stochastic processes on T depending on additional identically distributed random variables $\boldsymbol{W}_n^{(1)}, \boldsymbol{W}_n^{(2)}, \dots$ in some measurable space $\mathcal{W}_n$ that can in many cases be interpreted as *bootstrap weights* and should in general be seen as the additional sources of randomness introduced by the resampling scheme. As for $\mathbb{G}_n$, it is assumed that the sample paths of $\mathbb{G}_n^{(1)}, \mathbb{G}_n^{(2)}, \dots$ also belong to $\ell^\infty(T)$ and, when seen as maps into $\ell^\infty(T)$, no measurability assumptions are made on these bootstrap replicates either. When $\boldsymbol{X}_n$ represents i.i.d. observations and $\mathbb{G}_n$ is the general empirical process constructed from $\boldsymbol{X}_n$, several examples of possible bootstrap replicates of $\mathbb{G}_n$ can for instance be found in van der Vaart and Wellner (2000, Section 3.6). As in Section 3.6 of the latter reference, we assume throughout this section that the underlying probability space is independent of n and has a product structure, that is, $\Omega = \Omega_0 \times \Omega_1 \times \cdots$ with probability measure $\mathrm{P} = \mathrm{P}_0 \otimes \mathrm{P}_1 \otimes \cdots$, where P_i denotes the probability measure on Ω_i, such that, for any $\omega \in \Omega$, $\boldsymbol{X}_n(\omega)$ only depends on the first coordinate of ω and $\boldsymbol{W}_n^{(i)}(\omega)$ only depends on the $(i+1)$-coordinate of ω, implying in particular that $\boldsymbol{X}_n, \boldsymbol{W}_n^{(1)}, \boldsymbol{W}_n^{(2)}, \dots$ are independent. To be able to reuse the results of Section 2, we further assume that the finite dimensional distributions of the limiting stochastic process $\mathbb{G}$ are absolutely continuous.

Some additional notation is needed before our main result can be stated. For any map $Z : \Omega \to \mathbb{R}$, let Z^* be any *minimal measurable majorant of Z with respect to* P, that is, $Z^* : \Omega \to [-\infty, \infty]$ is measurable, $Z^* \geq Z$ and $Z^* \leq U$ almost surely for any measurable function $U : \Omega \to [-\infty, \infty]$ with $U \geq Z$ almost surely. A *maximal measurable minorant of Z with respect to* P is denoted by Z_* and defined by $Z_* = -(-Z)^*$ (see van der Vaart and Wellner, 2000, Section 1.2). Furthermore, for any $i \in \{0, 1, \dots\}$, we define the map $Z^{i*} : \Omega \to [-\infty, \infty]$ such that, for any $(\omega_0, \dots, \omega_{i-1}, \omega_{i+1}, \dots) \in \Omega_0 \times \dots \Omega_{i-1} \times \Omega_{i+1} \times \cdots$, the map $\omega_i \mapsto Z^{i*}(\omega_0, \dots, \omega_{i-1}, \omega_i, \omega_{i+1}, \dots)$ is a minimal measurable majorant of $\omega_i \mapsto Z(\omega_0, \dots, \omega_{i-1}, \omega_i, \omega_{i+1}, \dots)$ with respect to P_i. For a real-valued function Y on $\mathcal{X}_n \times \mathcal{W}_n$ such that $\boldsymbol{w} \mapsto Y(\boldsymbol{x}, \boldsymbol{w})$ is measurable for all $\boldsymbol{x} \in \mathcal{X}_n$, we further use the notation

$$\mathrm{E}(Y \mid \boldsymbol{X}_n) = \int_{\mathcal{W}_n} Y(\boldsymbol{X}_n, \boldsymbol{w}) \, \mathrm{d}\mathrm{P}^{\boldsymbol{W}_n^{(i)}}(\boldsymbol{w}),$$

provided the integral exists. Note that if Y is jointly Borel measurable, the right-hand side of the last displays defines a version of the conditional expectation of Y given $\boldsymbol{X}_n$, whence the notation. Finally, for any metric space (D, d), let $\mathrm{BL}_1(D)$ denote the set of functions $h : D \to \mathbb{R}$ with Lipschitz norm bounded by 1, that is, such that $\sup_{x \in D} |h(x)| \leq 1$ and $|h(x) - h(y)| \leq d(x, y)$ for all $x, y \in D$.

Lemma 3.1. *With the previous notation and under the above assumptions, the following two assertions are equivalent:*

(a) For any $M \in \mathbb{N}$, as $n \to \infty$,

$$(\mathbb{G}_n, \mathbb{G}_n^{(1)}, \dots, \mathbb{G}_n^{(M)}) \rightsquigarrow (\mathbb{G}, \mathbb{G}^{(1)}, \dots, \mathbb{G}^{(M)}) \qquad \text{in } \{\ell^\infty(T)\}^{M+1}, \tag{3.1}$$

where $\mathbb{G}, \mathbb{G}^{(1)}, \dots, \mathbb{G}^{(M)}$ are i.i.d.

(b) As $n \to \infty$,

$$\sup_{h \in \mathrm{BL}_1(\ell^\infty(T))} \left| \mathrm{E}\{h(\mathbb{G}_n^{(1)})^{1*} \mid \boldsymbol{X}_n\} - \mathrm{E}\{h(\mathbb{G})\} \right| \overset{\mathrm{P}^*}{\to} 0, \tag{3.2}$$

and $\mathbb{G}_n^{(1)}$ is asymptotically measurable, where $\overset{\mathrm{P}^*}{\to}$ denotes convergence in outer probability.

Let us make a few comments on this result:

- Assertion (b) is the extension of the conditional formulation of bootstrap consistency in $\mathbb{R}^d$ to the space $\ell^\infty(T)$. As explained in the introduction, following Giné and Zinn (1990), it can be seen as a "conditional adaptation" of the formulation of weak convergence based on the bounded Lipschitz metric. Section 3.6 in van der Vaart and Wellner (2000) and Chapter 10 in Kosorok (2008) in particular provide proofs of Assertion (b) for various bootstraps of the general empirical process constructed from i.i.d. observations along with continuous mapping theorems *for the bootstrap* and a functional delta method *for the bootstrap* that can be used to transfer (3.2) to the statistic level in certain situations.
- In van der Vaart and Wellner (2000), van der Vaart (1998) and Kosorok (2008), the expression on the left-hand side of (3.2) appears without the minimal measurable majorant with respect to the "weights". This is a consequence of the fact that, for all the resampling schemes considered in these monographs, the function $\boldsymbol{w} \mapsto \mathbb{G}_n^{(1)}(\boldsymbol{x}, \boldsymbol{w})$ is continuous for

all $\boldsymbol{x} \in \mathcal{X}_n$, implying that $\boldsymbol{w} \mapsto h\{\mathbb{G}_n^{(1)}(\boldsymbol{x}, \boldsymbol{w})\}$ is measurable for all $\boldsymbol{x} \in \mathcal{X}_n$ and all $h \in \mathrm{BL}_1(\ell^\infty(T))$. However, the minimal measurable majorant becomes for instance necessary if one wishes to apply Lemma 3.1 to certain stochastic processes appearing when using the *parametric bootstrap* (e.g., for goodness-of-fit testing, see, Stute et al., 1993; Genest and Rémillard, 2008). To see this, suppose that $\boldsymbol{X}_n$ is an i.i.d. sample of size n from some d.f. G on the real line, with G from some parametric family $\{G_\theta\}$. A natural stochastic process, from which one may for instance construct classical goodness-of-fit statistics, is then $\mathbb{G}_n(t) = \sqrt{n}\{G_n(t) - G(t)\}$, $t \in \mathbb{R}$, where G_n is the empirical d.f. of $\boldsymbol{X}_n$. Bootstrap samples are generated by sampling from G_{θ_n}, where $\theta_n = \theta_n(\boldsymbol{X}_n)$ is an estimator of θ. Note in passing that the latter way of proceeding is compatible with the product-structure condition on the underlying probability space since bootstrap samples can equivalently be regarded as obtained by applying $G_{\theta_n}^{-1}$ component-wise to independent random vectors $\boldsymbol{W}_n^{(1)}, \boldsymbol{W}_n^{(2)}, \dots$ independent of $\boldsymbol{X}_n$ and whose components are i.i.d. standard uniform. Now, corresponding parametric bootstrap replicates of $\mathbb{G}_n$ are given by $\mathbb{G}_n^{(i)} = \sqrt{n}(G_n^{(i)} - G_n)$, where $G_n^{(i)}$ is the empirical d.f. of the sample $(G_{\theta_n}^{-1}(W_{n1}^{(i)}), \dots, G_{\theta_n}^{-1}(W_{nn}^{(i)}))$. The need for the minimal measurable majorant with respect to the "weights" in (3.2) is then a consequence of the fact that the function from $\mathbb{R}^n$ to $\mathbb{R}$ defined by

$$\boldsymbol{w}^{(i)} \mapsto h\{\mathbb{G}_n^{(i)}(\boldsymbol{x}, \boldsymbol{w}^{(i)})\} = h\biggl(\frac{1}{\sqrt{n}} \sum_{j=1}^n \Bigl[\mathbf{1}\{G_{\theta_n(\boldsymbol{x})}^{-1}(w_j^{(i)}) \le \cdot\} - \mathbf{1}(x_j \le \cdot)\Bigr]\biggr)$$

is not measurable for all $h \in \mathrm{BL}_1(\ell^\infty(\mathbb{R}))$ and all $\boldsymbol{x} \in \mathcal{X}_n$, as can for instance be verified by adapting arguments from Billingsley (1999, Section 15).

- Bootstrap asymptotic validity in the form of Assertion (a) is less frequently encountered in the literature, although, as discussed in the introduction, it may be argued that this unconditional formulation is more intuitive and easy to work with. It is proved for example in Genest and Rémillard (2008) (for $M = 1$), Rémillard and Scaillet (2009), Segers (2012), Genest and Nešlehová (2014), Berghaus and Bücher (2017) and Bücher and Kojadinovic (2016a,b), among many others, for various stochastic processes arising in statistical tests on copulas or for assessing stationarity (after possibly transferring a weak convergence result with respect to the Skorohod topology into a result with respect to the supremum distance, which, by the continuous mapping theorem, is possible whenever the weak limit has continuous sample paths almost surely).
- As mentioned in the introduction, note that Assertions (a) and (b) are known to be equivalent for the special case of the *multiplier bootstrap* for the general empirical process based on i.i.d. observations and, in this case, it is even sufficient to consider $M = 1$ in (a): Corollary 2.9.3 in van der Vaart and Wellner (2000) corresponds to Assertion (a), while Theorem 2.9.6 corresponds to Assertion (b). The equivalence between the two follows by combining Theorem 2.9.6 with Theorem 2.9.2.

Before proving Lemma 3.1, we provide a useful corollary which is an immediate consequence of Lemma 3.1 and Lemma 2.2. It may be regarded as an analogue of Theorem 1.5.4 in van der Vaart and Wellner (2000) in a conditional setting and, roughly speaking, states that conditional weak convergence of a sequence of stochastic processes is equivalent to the conditional weak convergence of finite-dimensional distributions and (unconditional) asymptotic tightness.

Corollary 3.2. *Suppose that the assumptions of Lemma 3.1 are met. Then, any of the equivalent assertions in that lemma is equivalent to the fact that the finite dimensional distributions of $\mathbb{G}_n^{(1)}$ conditionally weakly converge to those of $\mathbb{G}$ in probability, that is, for any $k \in \mathbb{N}$ and*

$s_1, \ldots, s_k \in T$,

$$d_K\left(\mathrm{P}^{(\mathbb{G}_n^{(1)}(s_1),\ldots,\mathbb{G}_n^{(k)}(s_k))|\boldsymbol{X}_n}, \mathrm{P}^{(\mathbb{G}(s_1),\ldots,\mathbb{G}(s_k))}\right)$$
$$= \sup_{\boldsymbol{x}\in\mathbb{R}^k} \left|\mathrm{P}\{\mathbb{G}_n^{(1)}(s_1) \le x_1, \ldots, \mathbb{G}_n^{(1)}(s_k) \le x_k \mid \boldsymbol{X}_n\} - \mathrm{P}\{\mathbb{G}(s_1) \le x_1, \ldots, \mathbb{G}(s_k) \le x_k\}\right| \overset{\mathrm{P}}{\to} 0$$

as $n \to \infty$, and that $\mathbb{G}_n^{(1)}$ is (unconditionally) asymptotically tight.

Proof of Lemma 3.1. We start by proving an equivalence that will be used both in the necessity and the sufficiency part of the proof. For any $s_1, \ldots, s_k \in T$, let

$$A_n(s_1, \ldots, s_k)$$
$$= \sup_{\boldsymbol{x}\in\mathbb{R}^k} \left|\mathrm{P}\{\mathbb{G}_n^{(1)}(s_1) \le x_1, \ldots, \mathbb{G}_n^{(1)}(s_k) \le x_k \mid \boldsymbol{X}_n\} - \mathrm{P}\{\mathbb{G}(s_1) \le x_1, \ldots, \mathbb{G}(s_k) \le x_k\}\right| \quad (3.3)$$

and let

$$B_n(s_1, \ldots, s_k) = \sup_{h\in\mathrm{BL}_1(\mathbb{R}^k)} \left|\mathrm{E}[h\{\mathbb{G}_n^{(1)}(s_1), \ldots, \mathbb{G}_n^{(1)}(s_k)\} \mid \boldsymbol{X}_n] - \mathrm{E}[h\{\mathbb{G}(s_1), \ldots, \mathbb{G}(s_k)\}]\right|. \quad (3.4)$$

By our convention on conditional expectations after Condition 2.1, both $A_n(s_1, \ldots, s_k)$ and $B_n(s_1, \ldots, s_k)$ should be considered as depending only on a regular version of the conditional distribution of $(\mathbb{G}_n^{(1)}(s_1), \ldots, \mathbb{G}_n^{(1)}(s_k))$ given $\boldsymbol{X}_n$. Under the considered assumptions, we claim that, for any $s_1, \ldots, s_k \in T$, as $n \to \infty$,

$$A_n(s_1, \ldots, s_k) \overset{\mathrm{P}}{\to} 0 \qquad \text{if and only if} \qquad B_n(s_1, \ldots, s_k) \overset{\mathrm{P}}{\to} 0. \quad (3.5)$$

We shall drop the dependence on $s_1, \ldots, s_k$ in the notation when there is no risk of confusion. To prove (3.5), fix $s_1, \ldots, s_k \in T$ and assume first that A_n converges to zero in probability. Hence, every subsequence $A_{n'}$ of A_n has a further subsequence $A_{n''}$ along which the latter convergence takes place almost surely, that is, $A_{n''}(\omega) \to 0$ for all ω in a measurable set Ω_0 with $\mathrm{P}(\Omega_0) = 1$. Since weak convergence on $\mathbb{R}^k$ may be metrized by the bounded Lipschitz metric (see, e.g., van der Vaart and Wellner, 2000, Chapter 1.12), we obtain that $A_{n''}(\omega) \to 0$ if and only if $B_{n''}(\omega) \to 0$ for all $\omega \in \Omega_0$. Since the subsequence we started with was arbitrary, we obtain that B_n converges to zero in probability. For the converse, assume that B_n converges to zero in probability and exchange the roles of A_n and B_n in the previous lines to obtain that A_n converges to zero in probability.

We can now proceed with the proof, closely following the proof of Theorem 2.9.6 of van der Vaart and Wellner (2000) and relying on Lemma 2.2 where necessary.

(*a*) $\Rightarrow$ (*b*): Asymptotic measurability of $\mathbb{G}_n^{(1)}$ is an immediate consequence of the weak convergence of $\mathbb{G}_n^{(1)}$ to $\mathbb{G}^{(1)}$ in $\ell^\infty(T)$ (van der Vaart and Wellner, 2000, Lemma 1.3.8). Next, by Theorems 1.5.4 and 1.5.7 in van der Vaart and Wellner (2000), the latter convergence implies that there exists a semimetric ρ on T such that (T, ρ) is totally bounded and such that, for any $\varepsilon > 0$,

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathrm{P}^*\left\{\sup_{\rho(s,t)<\delta} |\mathbb{G}_n^{(1)}(s) - \mathbb{G}_n^{(1)}(t)| > \varepsilon\right\} = 0. \quad (3.6)$$

Fix $\ell \in \mathbb{N}$. For any $s \in T$, let $B(s, 1/\ell) = \{t \in T : \rho(s,t) < 1/\ell\}$ denote the ball of radius $1/\ell$ centered at s. Since (T, ρ) is totally bounded, there exists $k = k(\ell) \in \mathbb{N}$ and $s_i = s_i(\ell) \in T$, $i \in \{1, \ldots, k\}$, such that T is included in the union of all balls $B(s_i, 1/\ell)$, $i \in \{1, \ldots, k\}$. The latter allows us to define a mapping $\Pi_\ell : T \to T$ defined, for any $s \in T$, by $\Pi_\ell(s) = s_{i^*}$ where s_{i^*} is the center of a ball containing s. Now, to prove (3.2), we consider the decomposition

$$\sup_{h \in \mathrm{BL}_1(\ell^\infty(T))} |\mathrm{E}\{h(\mathbb{G}_n^{(1)})^{1*} \mid \boldsymbol{X}_n\} - \mathrm{E}\{h(\mathbb{G})\}| \leq I_n(\ell) + J_n(\ell) + K(\ell), \qquad \ell \in \mathbb{N},$$

where

$$\begin{aligned} I_n(\ell) &= \sup\nolimits_{h \in \mathrm{BL}_1(\ell^\infty(T))} \left| \mathrm{E}\{h(\mathbb{G}_n^{(1)})^{1*} \mid \boldsymbol{X}_n\} - \mathrm{E}\{h(\mathbb{G}_n^{(1)} \circ \Pi_\ell)^{1*} \mid \boldsymbol{X}_n\} \right|, \\ J_n(\ell) &= \sup\nolimits_{h \in \mathrm{BL}_1(\ell^\infty(T))} \left| \mathrm{E}\{h(\mathbb{G}_n^{(1)} \circ \Pi_\ell)^{1*} \mid \boldsymbol{X}_n\} - \mathrm{E}\{h(\mathbb{G} \circ \Pi_\ell)\} \right|, \\ K(\ell) &= \sup\nolimits_{h \in \mathrm{BL}_1(\ell^\infty(T))} \left| \mathrm{E}\{h(\mathbb{G} \circ \Pi_\ell)\} - \mathrm{E}\{h(\mathbb{G})\} \right|. \end{aligned}$$

Some thought reveals that (3.2) is proved if, for any $\varepsilon > 0$,

$$\lim_{\ell \to \infty} \limsup_{n \to \infty} \mathrm{P}^* \left\{ I_n(\ell) > \varepsilon \right\} = 0, \tag{3.7}$$

and similarly for $J_n(\ell)$ and $K(\ell)$.

Term $I_n(\ell)$: By Markov's inequality for outer probabilities (Lemma 6.10 in Kosorok, 2008), it suffices to show (3.7) with $\mathrm{P}^* \left\{ I_n(\ell) > \varepsilon \right\}$ replaced by $\mathrm{E}^*\{I_n(\ell)\}$. For any $\ell \in \mathbb{N}$, we have, by Lemma 1.2.2 (iii) in van der Vaart and Wellner (2000),

$$\begin{aligned} I_n(\ell) &\leq \sup_{h \in \mathrm{BL}_1(\ell^\infty(T))} \mathrm{E}\Big\{ |h(\mathbb{G}_n^{(1)} \circ \Pi_\ell) - h(\mathbb{G}_n^{(1)})|^{1*} \mid \boldsymbol{X}_n \Big\} \\ &\leq \mathrm{E}\Big[\Big\{ \sup_{s \in T} |\mathbb{G}_n^{(1)} \circ \Pi_\ell(s) - \mathbb{G}_n^{(1)}(s)| \wedge 1 \Big\}^* \mid \boldsymbol{X}_n \Big] \leq \mathrm{E}\Big\{ L_n(\ell)^* \mid \boldsymbol{X}_n \Big\}, \end{aligned}$$

where $\wedge$ denotes the minimum operator and $L_n(\ell) = \sup_{\rho(s,t) < 1/\ell} |\mathbb{G}_n^{(1)}(s) - \mathbb{G}_n^{(1)}(t)| \wedge 1$. It follows that $\mathrm{E}^*\{I_n(\ell)\} \leq \mathrm{E}\{L_n(\ell)^*\}$. Note that, by Lemma 1.2.2 (viii) in van der Vaart and Wellner (2000), we may choose $L_n(\ell)^*$ in such a way that $\ell \mapsto L_n(\ell)^*$ is nonincreasing almost surely. Then $\ell \mapsto \mathrm{P}\{L_n(\ell)^* > \varepsilon\}$ is nonincreasing as well, and from (3.6) and Problem 2.1.5 in van der Vaart and Wellner (2000, see also Section 2.1.2), we have that $L_n(\ell_n)^* \to 0$ in probability as $n \to \infty$ for any sequence $\ell_n \to \infty$, which, by dominated convergence for convergence in probability, implies that $\mathrm{E}\{L_n(\ell_n)^*\} \to 0$. Hence, $\lim_{\ell \to \infty} \limsup_{n \to \infty} \mathrm{E}[L_n(\ell)^*] = 0$ by invoking Problem 2.1.5 in van der Vaart and Wellner (2000) again.

Term $J_n(\ell)$: Fix $\ell \in \mathbb{N}$ and recall that the centers of the balls defining Π_ℓ were denoted by $s_1, \dots, s_k$. Since the weak convergence stated in (a) implies weak convergence of the respective finite dimensional distributions, we may invoke Lemma 2.2 to conclude that $A_n(s_1, \dots, s_k)$ in (3.3) converges in probability to zero. From (3.5), we then obtain that $B_n(s_1, \dots, s_k)$ in (3.4) converges in probability to zero as well. Next, let $h \in \mathrm{BL}_1(\ell^\infty(T))$ be arbitrary. Define $f : \mathbb{R}^k \to \ell^\infty(T)$ such that, for any $\boldsymbol{x} \in \mathbb{R}^k$ and $s \in T$, $f(\boldsymbol{x})(s) = x_i$ if $\Pi_\ell(s) = s_i$. Furthermore, let $g : \mathbb{R}^k \to \mathbb{R}$ be defined as $g(\boldsymbol{x}) = h(f(\boldsymbol{x}))$ implying that $h(\mathbb{G} \circ \Pi_\ell) = g(\mathbb{G}(s_1), \dots, \mathbb{G}(s_k))$. Some thought reveals that $g \in \mathrm{BL}_1(\mathbb{R}^k)$, whence $J_n(\ell) \leq B_n(s_1, \dots, s_k) \to 0$ in probability as $n \to \infty$ for all $\ell \in \mathbb{N}$, implying the analogue of (3.7) for $J_n(\ell)$.

Term $K(\ell)$: For any $\ell \in \mathbb{N}$, we have

$$K(\ell) \leq \mathrm{E}\Big\{ \sup_{s \in T} |\mathbb{G} \circ \Pi_\ell(s) - \mathbb{G}(s)| \wedge 1 \Big\} \leq \mathrm{E}\Big\{ \sup_{\rho(s,t) < 1/\ell} |\mathbb{G}(s) - \mathbb{G}(t)| \wedge 1 \Big\}.$$

By tightness of $\mathbb{G}$, Addendum 1.5.8 in van der Vaart and Wellner (2000) and dominated convergence, the expectation on the right converges to zero as $\ell \to \infty$, implying the analogue of (3.7) for $K(\ell)$.

$(b) \Rightarrow (a)$: To prove (3.1), we need to show the weak convergence of the finite-dimensional distributions and marginal asymptotic tightness. We start with the former. Let $M, k \in \mathbb{N}$ and $s_1, \dots, s_k \in T$. It suffices to show that, as $n \to \infty$,

$$\big(\mathbb{G}_n(s_1), \dots, \mathbb{G}_n(s_k), \mathbb{G}_n^{(1)}(s_1), \dots, \mathbb{G}_n^{(1)}(s_k), \dots, \mathbb{G}_n^{(M)}(s_1), \dots, \mathbb{G}_n^{(M)}(s_k)\big) \\ \rightsquigarrow \big(\mathbb{G}(s_1), \dots, \mathbb{G}(s_k), \mathbb{G}^{(1)}(s_1), \dots, \mathbb{G}^{(1)}(s_k), \dots, \mathbb{G}^{(M)}(s_1), \dots, \mathbb{G}^{(M)}(s_k)\big) \quad (3.8)$$

in $\mathbb{R}^{(M+1)k}$. Now, for any $g \in \mathrm{BL}_1(\mathbb{R}^k)$, the function $h : \ell^\infty(T) \to \mathbb{R}$ defined by $h(f) = g(f(s_1), \dots, f(s_k))$ is an element of $\mathrm{BL}_1(\ell^\infty(T))$. From (3.2), we then obtain that $B_n(s_1, \dots, s_k)$ in (3.4) converges in probability to zero, which, by (3.5), implies the same for $A_n(s_1, \dots, s_k)$ in (3.3). We may hence apply Lemma 2.2 to obtain (3.8).

It remains to show marginal tightness. Since $\mathbb{G}_n \rightsquigarrow \mathbb{G}$ in $\ell^\infty(T)$ and $\mathbb{G}_n^{(1)}, \dots, \mathbb{G}_n^{(M)}$ are identically distributed, it is sufficient to show that $\mathbb{G}_n^{(1)} \rightsquigarrow \mathbb{G}^{(1)}$ in $\ell^\infty(T)$. Then, as in the proof of Theorem 2.9.6 of van der Vaart and Wellner (2000), for any $h \in \mathrm{BL}_1(\ell^\infty(T))$,

$$|\mathrm{E}^*\{h(\mathbb{G}_n^{(1)})\} - \mathrm{E}\{h(\mathbb{G}^{(1)})\}| \le \Big|\mathrm{E}\Big[\mathrm{E}\{h(\mathbb{G}_n^{(1)})^* \mid \boldsymbol{X}_n\}\Big] - \mathrm{E}^*\Big[\mathrm{E}\{h(\mathbb{G}_n^{(1)})^{1*} \mid \boldsymbol{X}_n\}\Big]\Big| \\ + \Big|\mathrm{E}^*\Big[\mathrm{E}\{h(\mathbb{G}_n^{(1)})^{1*} \mid \boldsymbol{X}_n\} - \mathrm{E}\{h(\mathbb{G}^{(1)})\}\Big]\Big|.$$

By dominated convergence for convergence in outer probability and (3.2), the second term converges to zero. Since $h(\mathbb{G}_n^{(1)})^{1*} \ge \{h(\mathbb{G}_n^{(1)})_*\}^{1*} = h(\mathbb{G}_n^{(1)})_*$ almost surely, the first term is bounded above by

$$\mathrm{E}\Big[\mathrm{E}\{h(\mathbb{G}_n^{(1)})^* \mid \boldsymbol{X}_n\}\Big] - \mathrm{E}\Big[\mathrm{E}\{h(\mathbb{G}_n^{(1)})_* \mid \boldsymbol{X}_n\}\Big] = \mathrm{E}\{h(\mathbb{G}_n^{(1)})^*\} - \mathrm{E}\{h(\mathbb{G}_n^{(1)})_*\}.$$

The latter expression converges to zero since $\mathbb{G}_n^{(1)}$ is assumed asymptotically measurable. The assertion follows from the Portmanteau Theorem (see, e.g., van der Vaart and Wellner, 2000, Theorem 1.3.4 (i) and (vii)). ■

4 Validity of bootstrap-based confidence intervals and tests

Whether the consistency of a resampling scheme is shown at the stochastic process level and then transferred to $\mathbb{R}^d$ or is directly proved at the statistic level, one naturally expects corresponding bootstrap-based confidence intervals and tests to be asymptotically valid. Specifically, the latter amounts to verifying that confidence intervals have the correct asymptotic coverage and that tests maintain their level asymptotically. To formally establish these expected consequences, in this section, we restrict ourselves to the classical situation of a real-valued statistic S_n and thus assume that Condition 2.1 holds for $d = 1$. A result in the desired direction is for example Lemma 23.3 in van der Vaart (1998) and more specialized and deeper results are for instance collected in Horowitz (2001, Sections 3.3 and 3.4). Most results of that type do not however take into account the necessary approximation of the unobservable conditional d.f. of a bootstrap replicate by the empirical d.f. of a sample of bootstrap replicates. The following simple lemma does so and thus allows one to easily verify the asymptotic validity of bootstrap-based confidence intervals and tests constructed from a consistent resampling scheme in the sense of Lemma 2.2. Recall the definitions of F_n and F_n^M in (2.1).

Lemma 4.1. *Suppose that Condition 2.1 is met for $d = 1$ and that one of the equivalent assertions in Lemma 2.2 holds. Then, for any $\alpha \in (0,1)$,*

$$\lim_{n\to\infty} \mathrm{P}\{S_n \ge (F_n)^{-1}(1-\alpha)\} = \alpha \quad \textit{and} \quad \lim_{n,M\to\infty} \mathrm{P}\{S_n \ge (F_n^M)^{-1}(1-\alpha)\} = \alpha,$$

where G^{-1} denotes the generalized inverse of d.f. G, that is $G^{-1}(y) = \inf\{x \in \mathbb{R} : G(x) \geq y\}$, $y \in (0,1]$. The statements with '$\geq$' replaced by '$>$' in the previous display hold as well.

The assertion of this lemma involving conditional quantiles (or a version thereof) is usually provided in textbooks on the bootstrap to validate its use for the construction of confidence intervals and tests (see, e.g., Lemma 23.3 in van der Vaart, 1998). For completeness, we shall prove it below. The assertion involving empirical quantiles is the one to be used in practice as conditional quantiles are not available and must thus be approximated by Monte Carlo. Note in particular that the above formulation is general enough to allow $M = M(n)$ with $M(n) \to \infty$ as $n \to \infty$.

Let us now briefly verify that the asymptotic validity of bootstrap-based confidence intervals and tests is an immediate consequence of the preceding lemma. Start with the former and assume that $S_n = \sqrt{n}(\theta_n - \theta)$, where θ_n is an estimator of some parameter $\theta \in \mathbb{R}$. Then, a natural confidence interval for θ is given by

$$I_{n,M,\alpha} = \Big[\theta_n - n^{-1/2}(F_n^M)^{-1}(1-\alpha/2), \theta_n - n^{-1/2}(F_n^M)^{-1}(\alpha/2)\Big], \qquad \alpha \in (0,1/2).$$

Note in passing that the above confidence interval is related to the so-called *basic bootstrap confidence interval* (see, e.g., Davison and Hinkley, 1997, Chapter 5). A consequence of Lemma 4.1 is then that, if one of the equivalent assertions in Lemma 2.2 holds, $I_{n,M,\alpha}$ is of asymptotic level $1-\alpha$ in the sense that, as $n, M \to \infty$,

$$\mathrm{P}(\theta \in I_{n,M,\alpha}) = \mathrm{P}\{S_n \geq (F_n^M)^{-1}(\alpha/2)\} - \mathrm{P}\{S_n > (F_n^M)^{-1}(1-\alpha/2)\} \to 1-\alpha.$$

Let us now discuss the case of bootstrap-based tests. Assume that S_n is a test statistic for some null hypothesis H_0 such that large values of S_n provide evidence against H_0. It is then natural to reject H_0 at level $\alpha \in (0,1)$ when $S_n > (F_n^M)(1-\alpha)$. Should one of the equivalent assertions in Lemma 2.2 holds under H_0, Lemma 4.1 immediately implies that this test holds it level asymptotically in the sense that, under H_0, $\mathrm{P}\{S_n \geq (F_n^M)^{-1}(1-\alpha)\} \to \alpha$ as $n, M \to \infty$. If the bootstrap replicates are stochastically bounded under the alternative, then the test will also be consistent provided S_n converges to infinity in probability under the alternative.

Under the same setting, another statistic of interest is

$$p_n^M = \frac{1}{M}\sum_{i=1}^{M} \mathbf{1}(S_n^{(i)} > S_n) = 1 - F_n^M(S_n),$$

which may be interpreted as an approximate p-value for the test based on S_n. The theoretical analogue of the latter is

$$p_n = \mathrm{P}(S_n^{(1)} > S_n \mid \boldsymbol{X}_n) = 1 - F_n(S_n).$$

Intuitively, the resampling scheme being valid should imply that, under the null hypothesis, the statistics p_n^M and p_n are approximately standard uniform. The following result formalizes this.

Corollary 4.2. *Suppose that Condition 2.1 is met for $d = 1$ and that one of the equivalent assertions in Lemma 2.2 holds. Then, as $n \to \infty$,*

$$p_n \rightsquigarrow \mathrm{Uniform}(0,1) \qquad \textit{and} \qquad p_n^{M_n} \rightsquigarrow \mathrm{Uniform}(0,1),$$

for any sequence $M_n \to \infty$ as $n \to \infty$.

The proofs of Lemma 4.1 and Corollary 4.2 are given hereafter.

Proof of Lemma 4.1. Consider the assertion involving conditional quantiles. Combine Assertion (b) in Lemma 2.2 with (2.2) to obtain that every subsequence of $d_K(\mathrm{P}^{S_n^{(1)}|\boldsymbol{X}_n}, \mathrm{P}^S)$ has a further subsequence along which this expression converges almost surely to zero as $n \to \infty$. Let $\alpha \in (0,1)$ such that F^{-1} is continuous at $1-\alpha$. As a consequence of Lemma 21.2 in van der Vaart (1998), we obtain that $F_n^{-1}(1-\alpha) \xrightarrow{\text{a.s.}} F^{-1}(1-\alpha)$ along that subsequence. Hence, the random vector $(S_n, F_n^{-1}(1-\alpha))$ converges weakly to $(S, F^{-1}(1-\alpha))$, again along that subsequence. Since $\mathrm{P}\{S = F^{-1}(1-\alpha)\} = 0$ by absolute continuity, the Portmanteau Theorem implies that

$$\mathrm{P}\{S_n \geq F_n^{-1}(1-\alpha)\} \to \mathrm{P}\{S \geq F^{-1}(1-\alpha)\} = \alpha \tag{4.1}$$

along that subsequence. The latter equation holds for all expect at most countably many $\alpha \in (0,1)$. Because the left (resp. right) side of (4.1) is an increasing (resp. increasing continuous) function of α, (4.1) must hold for all $\alpha \in (0,1)$. The first assertion follows since the subsequence we started with was arbitrary. Finally, note that one may replace '$\geq$' by '$>$' in the last display.

Consider the assertion involving empirical quantiles. Let $q_{1-\alpha}$ denote the $(1-\alpha)$-quantile of S. Since $\lim_{n\to\infty} \mathrm{P}(S_n \geq q_{1-\alpha}) = \mathrm{P}(S \geq q_{1-\alpha}) = \alpha$ as a consequence of the Portmanteau Theorem and the absolute continuity of the random variable S, it suffices to show that

$$\begin{aligned}\lim_{n,M\to\infty} &|\mathrm{P}\{S_n \geq (F_n^M)^{-1}(1-\alpha)\} - \mathrm{P}(S_n \geq q_{1-\alpha})| \\ &= \lim_{n,M\to\infty} |\mathrm{P}\{F_n^M(S_n) \geq 1-\alpha\} - \mathrm{P}\{F(S_n) \geq 1-\alpha\}| = 0,\end{aligned}$$

where the equality follows from the fact that F_n^M and F are right-continuous. Using the fact that, for any $a, b, x \in \mathbb{R}$ and $\varepsilon > 0$, $|\mathbf{1}(x \leq a) - \mathbf{1}(x \leq b)| \leq \mathbf{1}(|x-a| \leq \varepsilon) + \mathbf{1}(|a-b| > \varepsilon)$, we can estimate

$$\begin{aligned}&|\mathrm{P}\{F_n^M(S_n) \geq 1-\alpha\} - \mathrm{P}\{F(S_n) \geq 1-\alpha\}| \\ &\qquad \leq \mathrm{P}\{|F(S_n) - 1 + \alpha| \leq \varepsilon\} + \mathrm{P}\{|F(S_n) - F_n^M(S_n)| > \varepsilon\}.\end{aligned}$$

By the continuous mapping theorem and the Portmanteau Theorem, the first term on the right converges to $\mathrm{P}\{|F(S) - 1 + \alpha| \leq \varepsilon\}$ as $n \to \infty$, which can be made arbitrary small by decreasing ε. Combining Assertion (c) from Lemma 2.2 with (2.2) (for $d = 1$) immediately implies that the second term converges to zero as $n, M \to \infty$, hence the first claim.

The claim with '$\geq$' replaced by '$>$' follows the from fact that, by absolute continuity of S and the Portmanteau Theorem, for any $x \in \mathbb{R}$, $\mathrm{P}(S_n < x) \to \mathrm{P}(S < x)$ as $n \to \infty$. The latter convergence can be made uniform by arguments as in Lemma 2.11 in van der Vaart (1998), which, combined with (2.2) (for $d = 1$) implies that $\sup_{x\in\mathbb{R}} \mathrm{P}(S_n = x)$ converges to zero in probability as $n \to \infty$. ■

Proof of Corollary 4.2. The weak convergence $S_n \rightsquigarrow S$ as $n \to \infty$ together with the continuous mapping theorem implies that $1 - F(S_n) \rightsquigarrow 1 - F(S) \sim \text{Uniform}(0,1)$ as $n \to \infty$. Combining Assertion (b) in Lemma 2.2 with (2.2) (for $d = 1$), we additionally immediately obtain that $F_n(S_n) - F(S_n)$ converges to zero in probability as $n \to \infty$, which implies that p_n has the same weak limit as $1 - F(S_n)$ as $n \to \infty$. Similarly, Assertion (c) in Lemma 2.2 combined with (2.2) (for $d = 1$) readily implies that $p_n^{M_n}$ has the same limit distribution as $1 - F(S_n)$ as $n \to \infty$. ■

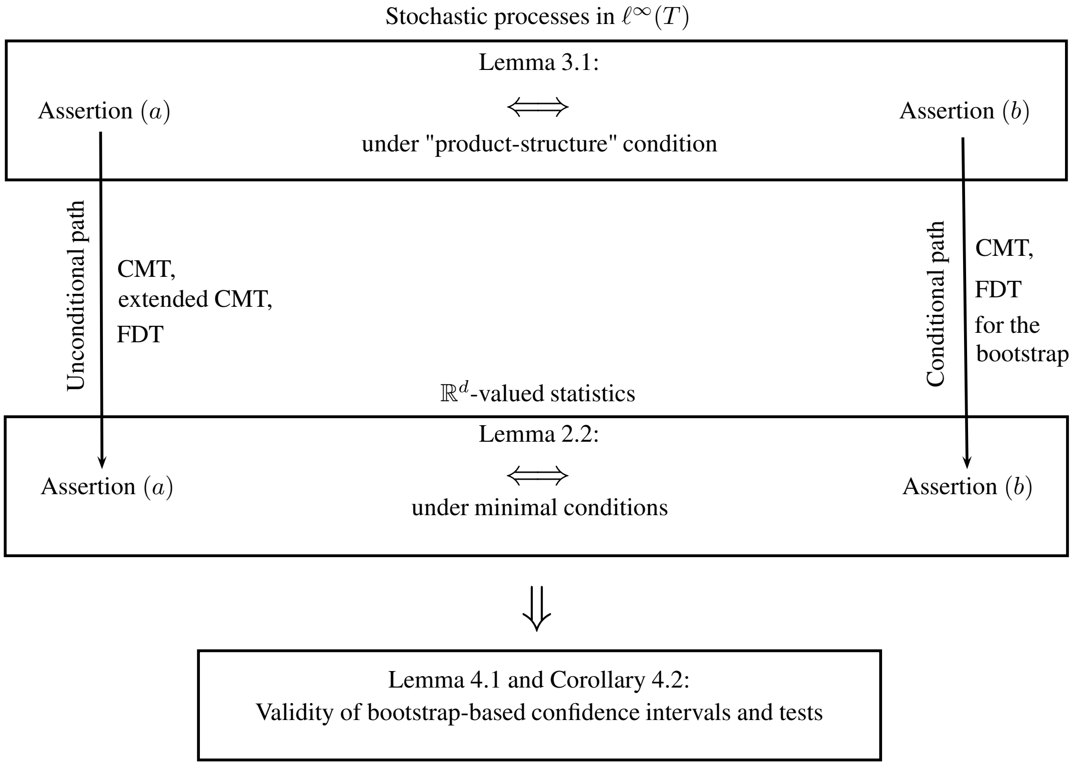

Figure 1: Summary of results; CMT stands for "continuous mapping theorem" and FDT for "functional delta method".

5 Concluding remarks

As a picture often speaks better than words, we summarized the results obtained in this work in the diagram of Figure 1. From the point of view of applications of resampling schemes starting at the stochastic process level, the diagram highlights two paths to proving the asymptotic validity of bootstrap-based confidence intervals and tests: an unconditional path starting at Assertion (a) of Lemma 3.1 and a conditional path starting at Assertion (b) of Lemma 3.1.

We conclude by summarizing the main consequences and features of the results obtained in this note, some of which explicitly appear in the diagram of Figure 1:

- At the stochastic process level, it may be argued that one needs to deal with less subtle mathematical concepts to prove unconditional bootstrap consistency than to show its conditional version. Roughly speaking, the unconditional approach avoids the need to work with the seemingly awkward notion of "conditional law" of a non-measurable function. The fact that the joint weak convergence in Assertion (a) must be proved for all $M \in \mathbb{N}$ is not an obstacle in most cases as going from the result for $M = 1$ to the result for all $M \in \mathbb{N}$ only requires a slightly more involved proof of the weak convergence of the finite dimensional distributions, the asymptotic tightness part of the proof remaining the same.
- Focusing for instance on existing continuous mapping theorems *for the bootstrap* (Kosorok, 2008, Section 10.1.4), it appears that, for transferring Assertion (b) of Lemma 3.1 into Assertion (b) of Lemma 2.2, more assumptions than just continuity of the underlying functional are necessary, thereby suggesting that the unconditional formulation of bootstrap consistency might be slightly more useful. Additionally, Assertion (a) of Lemma 3.1 can

be combined with the extended continuous mapping theorem (van der Vaart and Wellner, 2000, Theorem 1.11.1), a version *for the bootstrap* of which does not seem to exist.

- The equivalence between the unconditional and the conditional formulation of bootstrap consistency at the stochastic process level only holds if the additional randomness in the bootstrap replicates is independent of the data (in fact, this assumption is only needed to make Assertion (b) well-defined). Interestingly enough, such a condition does not seem to be a restriction in practice as it seems satisfied by most if not all resampling schemes. As a consequence, Lemma 3.1 confirms that most bootstrap consistency results obtained under the form of Assertion (a) in the literature (see Section 3 for references) are not any weaker than if the (equivalent) conditional formulation in (b) were proved.

Acknowledgments

The authors would like to thank Jean-David Fermanian for fruitful discussions. This research has been supported by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823) of the German Research Foundation, which is gratefully acknowledged.

References

R. Beran and G.R. Ducharme. *Asymptotic theory for bootstrap methods in statistics.* Les publication CRM, Centre de recherches mathématiques, Université de Montréal, Canada, 1991.

R. J. Beran, L. Le Cam, and P. W. Millar. Convergence of stochastic empirical measures. *J. Multivariate Anal.*, 23(1):159–168, 1987.

B. Berghaus and A. Bücher. Goodness-of-fit tests for multivariate copula-based time series models. *Econometric Theory*, 33(2):292–330, 2017.

Peter J. Bickel and David A. Freedman. Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6):1196–1217, 11 1981.

P. Billingsley. *Convergence of probability Measures.* Wiley, New York, 1999. Second edition.

A. Bücher and I. Kojadinovic. A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli*, 22(2), 2016a.

A. Bücher and I. Kojadinovic. Dependent multiplier bootstraps for non-degenerate u-statistics under mixing conditions with applications. *Journal of Statistical Planning and Inference*, 170:83–105, 2016b.

A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Application.* Cambridge University Press, 1997. ISBN 0-521-57391-2; 0-521-57471-4.

B. Efron. Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, 7(1):1–26, 01 1979.

C. Genest and J. G. Nešlehová. On tests of radial symmetry for bivariate copulas. *Statistical Papers*, 55:1107–1119, 2014.

C. Genest and B. Rémillard. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Annales de l'Institut Henri Poincaré: Probabilités et Statistiques*, 44:1096–1127, 2008.

Evarist Giné and Joel Zinn. Bootstrapping general empirical measures. *Ann. Probab.*, 18(2): 851–869, 1990.

Peter Hall. *The bootstrap and Edgeworth expansion.* Springer Series in Statistics. Springer-Verlag, New York, 1992.

J.L. Horowitz. The bootstrap. In J. J. Heckman and E. E. Leamer, editors, *Handbook of Econometrics*, volume 5, pages 3159–3228. North-Holland, Amsterdam, 2001.

Olav Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

M.R. Kosorok. *Introduction to empirical processes and semiparametric inference.* Springer, New York, 2008.

H.R. Künsch. The jacknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241, 1989.

B. Rémillard and O. Scaillet. Testing for equality between two copulas. *Journal of Multivariate Analysis*, 100(3):377–386, 2009.

J. Segers. Asymptotics of empirical copula processes under nonrestrictive smoothness assumptions. *Bernoulli*, 18:764–782, 2012.

X. Shao. The dependent wild bootstrap. *Journal of the American Statistical Association*, 105 (489):218–235, 2010.

W. Stute, W. Gonzáles Manteiga, and M. Presedo Quindimil. Bootstrap based goodness-of-fit tests. *Metrika*, 40:243–256, 1993.

A.W. van der Vaart. *Asymptotic statistics.* Cambridge University Press, 1998.

A.W. van der Vaart and J.A. Wellner. *Weak convergence and empirical processes.* Springer, New York, 2000. Second edition.