Fundamentals of non-parametric statistical inference for integrated quantiles

Nadezhda V. Gribkova [*,1,2], Mengqi Wang [†,3], and Ričardas Zitikis [‡,3]

[1]Saint Petersburg State University, Saint Petersburg, 199034 Russia

[2]Emperor Alexander I Saint Petersburg State Transport University,
Saint Petersburg, 190031 Russia

[3]Western University, London, Ontario N6A 5B7, Canada

March 11, 2025

Abstract. We present a general non-parametric statistical inference theory for integrals of quantiles without assuming any specific sampling design or dependence structure. Technical considerations are accompanied by examples and discussions, including those pertaining to the bias of empirical estimators. To illustrate how the general results can be adapted to specific situations, we derive – at a stroke and under minimal conditions – consistency and asymptotic normality of the empirical tail-value-at-risk, Lorenz and Gini curves at any probability level in the case of the simple random sampling, thus facilitating a comparison of our results with what is already known in the literature. Results, notes and references concerning dependent (i.e., time series) data are also offered. As a by-product, our general results provide new and unified proofs of large-sample properties of a number of classical statistical estimators, such as trimmed means, and give additional insights into the origins of, and the reasons for, various necessary and sufficient conditions.

Key words and phrases: integrated quantiles, expected shortfall, tail value at risk, Lorenz curve, Gini curve, trimmed mean, L-statistic, distortion risk measure, time series, S-mixing, M-mixing.

[*]e-mail: n.gribkova@spbu.ru

[†]e-mail: mwan259@uwo.ca

[‡]Corresponding author; e-mail: rzitikis@uwo.ca

Contents

1   Introduction

Although well understood, developing statistical inference for quantiles, denoted by $F^{-1}(p)$ for various $p \in [0,1]$ and known as values-at-risk in banking and insurance, is a challenging task. This is due to the fact that empirical quantiles are order statistics (e.g., Arnold et al., 2008; David and Nagaraja, 2003), unlike empirical cumulative distribution functions (cdf's) that are sums of (Bernoulli) random variables and can therefore be easily tackled using classical techniques of mathematical statistics and probability theory (e.g., Shorack, 2017). Inevitably, therefore, in the case of quantiles, quite restrictive assumptions on the population cdf F arise. For example, consistency of the empirical quantile requires continuity of the population quantile function at the specified probability level, while asymptotic normality requires conditions on the underlying probability density function (pdf) at the specified quantile (e.g. Serfling, 1980; Shao, 2003).

In many applications that arise in economics, finance, and insurance, quantiles are often integrated. Visualized in Figure 1.1 in the case of a Pareto distribution (specifics in Section 7) are four illustrative examples of such integrals and their combinations. Namely, with p running through the unit interval $(0,1)$, they are the upside tail-value-at-risk

$$\text{TVaR}(p) = \frac{1}{1-p} \int_p^1 F^{-1}(u)\mathrm{d}u, \tag{1.1}$$

the downside tail-value-at-risk

$$\text{TV@R}(p) = \frac{1}{p} \int_0^p F^{-1}(u)\mathrm{d}u, \tag{1.2}$$

the Lorenz curve

$$\text{LC}(p) = \frac{1}{\mu} \int_0^p F^{-1}(u)\mathrm{d}u, \tag{1.3}$$

and the Gini curve

$$\text{GC}(p) = \frac{1}{\mu} \left(\int_{1-p}^1 F^{-1}(u)\mathrm{d}u - \int_0^p F^{-1}(u)\mathrm{d}u \right), \tag{1.4}$$

where

$$\mu = \int_0^1 F^{-1}(u)\mathrm{d}u \tag{1.5}$$

is the mean of the population cdf F. Of course, the mean could serve a fifth illustrative
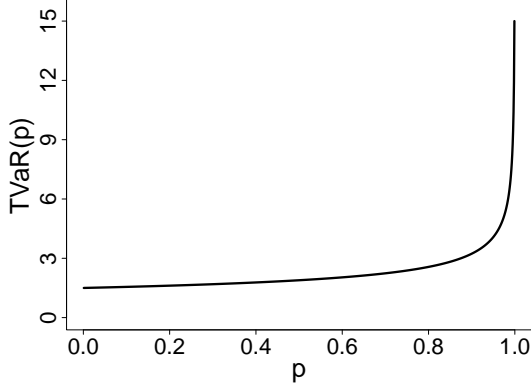
(a) Upside tail-value-at-risk TVaR(p).
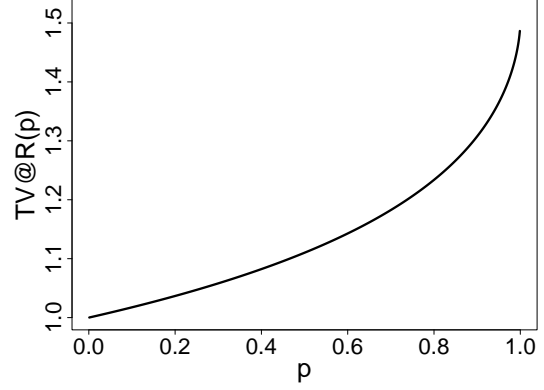
(b) Downside tail-value-at-risk TV@R(p).
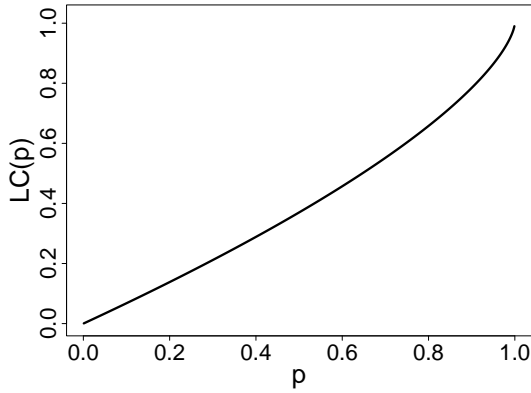
(c) Lorenz curve LC(p).
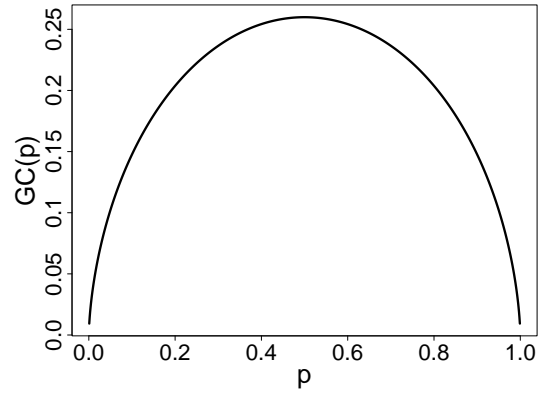
(d) Gini curve GC(p).

Figure 1.1: Four illustrative curves, whose definitions are based on integrated quantiles, depicted here in the case of a Pareto distribution.

example, but it is trivial in the context of the present paper and, therefore, is viewed here only as an auxiliary quantity. We shall give more details on the four measures of risk and economic inequality in Section 7. In Section 9, we shall discuss more complex (i.e., distorted or weighted) integrals of quantiles, known in statistics as L-integrals, and show how they can be reduced to those of the type that we see in the above examples.

Since, generally speaking, integrals are linear functionals of their integrands, which are quantiles in our case, researchers often use the aforementioned asymptotic results for quantiles and then apply the continuous mapping theorem (e.g., Billingsley, 1999) to derive desired results for integrated quantiles (e.g., Csörgő et al., 1986; Bellini et al., 2022; Han et al., 2024). This approach, however, involves the aforementioned assumptions on the population pdf, although we shall soon see that such assumptions are unnecessary for integrated quantilies; even the very existence of pdf is unnecessary.

4

Indeed, when developing an asymptotic theory for the Lorenz curve, whose definition is based on integrated quantiles as pointed out by Gastwirth (1971), Csörgő and Zitikis (1996) observed that neither the original formulation of the problem nor the obtained large-sample asymptotic distribution requires the existence of pdf, let alone assumptions on it. Based on the observation, they set out to find a path that would lead to desired asymptotic results without involving pdf's. They succeeded in achieving this goal by introducing a technical tool that they called the Vervaat process, named after the Dutch mathematician Wim Vervaat, whose pioneering results (Vervaat, 1972a,b) on a combination of the uniform on $[0, 1]$ quantile and empirical processes served an inspiration. We refer to Zitikis (1998) for details and references on the topic.

Note 1.1. The origin of the research path taken by Csörgő and Zitikis (1996) was humble: it was the equation

$$\int_0^1 \big(G^{-1}(u) - F^{-1}(u)\big)\mathrm{d}u = \int_{-\infty}^{\infty} \big(F(x) - G(x)\big)\mathrm{d}x \tag{1.6}$$

that is known to hold (see Appendix A.6.1 for details) for every pair of random variables X and Y with finite first moments, where F is the cdf of X (shorthanded as $X \sim F$) and G is the cdf of Y (shorthanded as $Y \sim G$), and F^{-1} and G^{-1} are the corresponding quantile functions. For example, the quantile function F^{-1} of X is given by the equation

$$F^{-1}(u) = \inf\{x \in \mathbb{R} \ : \ F(x) \geq u\} \tag{1.7}$$

for all $u \in (0, 1]$, while at the point $u = 0$ it is defined as the right-hand limit

$$F^{-1}(0) = \lim_{u \downarrow 0} F^{-1}(u). \tag{1.8}$$

(Mathematicians would call F^{-1} the left-continuous generalized inverse of F.) Since the quantile function is not the ordinary inverse of F, as such may not exist unless the cdf is continuous and strictly increasing, technical difficulties and even overlooks do arise (e.g. Wacker, 2023, and references therein). Note also that $F^{-1}(0)$ and $F^{-1}(1)$ are the two end-points (finite or infinite) of the support of the cdf F, which is the smallest closed-in-$\mathbb{R}$ interval that contains all the points $x \in \mathbb{R}$ such that $F(x) \in (0, 1)$.

A long series of research articles by various authors followed Csörgő and Zitikis (1996),

exploring theoretical and empirical aspects of integrated quantiles *without* involving pdf's. Recently, Wei and Zitikis (2023) suggested an extension of the theory initiated by Csörgő and Zitikis (1996). The current paper takes these developments even further by establishing asymptotic results for integrated quantiles without assuming any specific sampling design or dependence structure between the underlying random variables. In fact, the main results are formulated for generic cdf's F and approximating sequences F_n, $n \in \mathbb{N}$, of cdf's, which may or may not be random, depending on the problem.

2 An overview

We have organized the rest of the paper as follows. In Section 3, we develop a general statistical inference theory for the integral

$$\int_p^1 F^{-1}(u)\mathrm{d}u \tag{2.1}$$

of quantiles over the upper-most layer $(p, 1)$ of probabilities, where $p \in (0, 1)$ is a fixed boundary probability. We call integral (2.1) the upper-layer integral. It plays a particularly prominent role in insurance, where loss random variables are often non-negative, and thus the right-hand tails of their distributions become of particular concern. Risk measures such as the expected shortfall (ES) and its sister risk measure called the "upside" tail-value-at-risk (TVaR) arise (e.g., Denuit et al., 2005), which in turn lead to considerations of more general integrals called distorted expectations (e.g., Dhaene et al., 2012, and references therein). We shall rigorously define and discuss TVaR later in this paper (see Example 7.1 in particular).

In Section 4 we shall offer several simulated experiments that will clarify some of the statistical properties of the empirical upper-layer integral.

Before we delve into the topics of the following sections, we present several notes that are useful for understanding and appreciating the results of this paper, and in particular the assumptions under which the results will be established.

Note 2.1. When $p = 0$, integral (2.1) reduces to the mean μ of the cdf F, given by equation (1.5). Statistical inference for μ under various sampling designs can, of course, be developed without invoking the existence of the pdf of F. When $p = 1$, integral (2.1) vanishes. Hence, by restricting ourselves to $p \in (0, 1)$ we are just excluding the two statistically trivial cases $p = 0$ and $p = 1$.

Note 2.2. There is a deeper reason for separating the boundary case $p = 0$ from $p \in (0, 1)$. First, when $p = 0$, in the case of the simple random sampling (SRS), that is, when we work with n independent and identically distributed (iid) random variables, the empirical estimator of μ is an asymptotically normal estimator under the only requirement that the second moment of the underlying random variable is finite. No assumption related to continuity of the cdf F at any point of the real line is needed. In the case $p \in (0, 1)$, however, for an empirical estimator of the upper-layer integral $\int_p^1 F^{-1}(u)\mathrm{d}u$ to be asymptotically normal, it is necessary to assume that the quantile function F^{-1} is continuous at the point p. The necessity of this condition is demonstrated by Stigler (1973) when the empirical estimator is a trimmed mean based on SRS. This assumption is also needed for our general results as they absorb the SRS-based trimmed mean as a special case. No such condition, however, is required for *consistency* of the empirical estimator of the upper-layer integral $\int_p^1 F^{-1}(u)\mathrm{d}u$. We shall illustrate this phenomenon with a simulated example in Section 4.

Note 2.3. The aforementioned continuity of the quantile function F^{-1} at the point p permeates the entire paper. In our minds, however, we rarely visualize distributions in terms of quantiles – we often think of them in terms of cdf's. For this reason it is beneficial to follow Shao (2003) and reformulate the continuity of F^{-1} at the point $p \in (0, 1)$ in the form of the bounds

$$F\left(F^{-1}(p) - \varepsilon\right) < p < F\left(F^{-1}(p) + \varepsilon\right) \tag{2.2}$$

that need to be satisfied for every $\varepsilon > 0$. Figure 2.1 illustrates the reformulation. Note in the figure that F is strictly increasing at the point $F^{-1}(p)$ for every $p \in (0, 1)$ except when $p = p_2$, in which case there is a flat region adjacent to $x_1 = F^{-1}(p_2)$ and, therefore, the quantile function has a jump of size $x_2 - x_1$ at the point p_2. Given this explanation, for the sake of brevity, throughout the rest of the paper we follow Stigler (1973) and formulate the condition simply as continuity of the quantile function F^{-1} at the point p.

In Section 5, we develop a general statistical inference theory for the integral

$$\int_0^p F^{-1}(u)\mathrm{d}u \tag{2.3}$$

of quantiles over the lower-most layer $(0, p)$ of probabilities, where $p \in (0, 1)$ is a fixed boundary probability. Throughout the current paper we call integral (2.3) the lower-layer integral, although in the literature the integral is often called the generalized (or absolute)
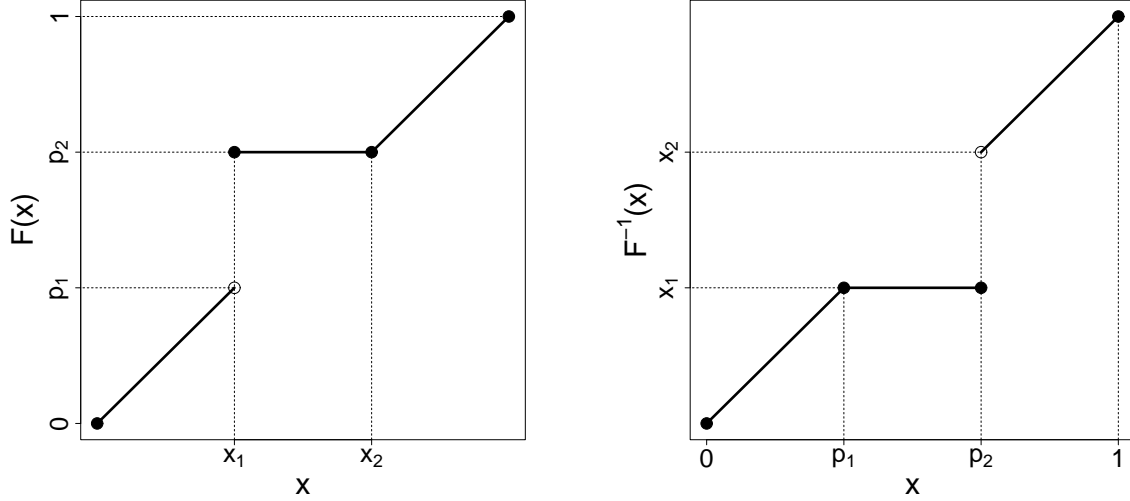
(a) Cdf: The only $p \in (0,1)$ that fails condition (2.2) is $p = p_2$, provided that $x_1 < x_2$. When $x_1 = x_2$ (no gap), every $p \in (0,1)$ satisfies the condition.

(b) Quantile function: The only discontinuity of F^{-1} is at $p = p_2$, provided that $x_1 < x_2$. When $x_1 = x_2$ (no jump), the quantile function is continuous.

Figure 2.1: An illustrative cdf and its quantile function with the boundary points $p_1 \le p_2$ and $x_1 \le x_2$ of the distributional jump and gap, respectively, where $x_1 = F^{-1}(p_2)$.

Lorenz curve (e.g., Davydov and Zitikis, 2002; Helmers and Zitikis, 2005; Davydov et al., 2007, and references therein) to reflect the fact that by dividing the integral by the mean μ of F, we obtain the Lorenz curve (see Example 7.3) that permeates the literature on economic inequality. Integral (2.3) also plays a significant role in finance, because financial losses are usually modelled using negative random variables. Risk measures such as the "downside" tail-value-at-risk (TV@R) arise (e.g., Pflug and Römisch, 2007). Its formula with an accompanying discussion will be provided in Example 7.2.

Note 2.4. Analogous observations to those made in Notes 2.1 and 2.2 apply to integral (2.3) as well: first, we exclude the statistically trivial cases $p = 0$ and $p = 1$, and second, asymptotic normality of the empirical estimator when $p \in (0,1)$, unlike in the case $p = 1$ that gives the mean μ, requires the quantile function F^{-1} to be continuous at the point p, as shown by Stigler (1973) in the case of trimmed mean arising from SRS. No such condition is required for consistency of the estimator of the lower-layer integral. Details are in Section 5.

There are also problems, like those associated with insurance layers (e.g, Wang, 1996), when aggregating middle quantiles is of interest. Looking from a different angle, there are also problems, like those concerning robust parameter estimation (Brazauskas et al., 2009; Zhao et al., 2018), when chopping off, or tempering, a certain percentage of extreme (smallest

and largest) quantiles during moment calculations becomes warranted, thus naturally leading to integrals of moderate (i.e., middle) quantiles. Hence, in Section 6 we develop a general statistical inference theory for the integral

$$\int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u \tag{2.4}$$

of quantiles over the middle layer (p_1, p_2) of probabilities, where $0 < p_1 < p_2 < 1$ are fixed boundary probabilities. We call integral (2.4) the middle-layer integral.

Note 2.5. When $p_1 = 0$ and $p_2 = 1$, integral (2.4) reduces to the mean μ of the cdf F. When $p_1 = 0 < p_2 < 1$, integral (2.4) reduces to integral (2.3), and when $0 < p_1 < p_2 = 1$, integral (2.4) reduces to integral (2.1). Hence, by restricting our study of the middle-layer integral to the case $0 < p_1 < p_2 < 1$ only, we are not losing anything from the statistical point of view.

Note 2.6. Asymptotic normality of the empirical estimator when $0 < p_1 < p_2 < 1$ requires the quantile function F^{-1} to be continuous at the points p_1 and p_2, as seen from Stigler (1973) in the case of trimmed mean arising from SRS. No such condition is required for consistency of the estimator. Details are in Section 6.

When modeling both profits and losses (so-called P&L variables), the underlying population distribution may span the entire real line $\mathbb{R}$, and thus both upper- and lower-layer integrals (2.1) and (2.3) may show up simultaneously. There are also problems that give rise to various combinations of the three integrals introduced above, and one of such instances is the Gini curve (GC), whose formula with an accompanying discussion will be provided in Example 7.4. The general results that we shall develop in the following sections are readily applicable to such combinations as well.

As we have already alluded to, in Section 7 we shall illustrate asymptotic results that can be obtained for the upside and downside tail-values-at-risk, as well as for the Lorenz and Gini curves. To maximally simplify the illustrations, yet to maintain their informative nature, we shall use SRS, that is, we shall work with iid random variables. We stress, however, that this choice of the sampling design is only to facilitate illustrations of our general results, which do not rely on any particular sampling design or dependence structure.

In Section 8 we shall show how our general results can be used in the case of time series data. In particular, we shall show that all that needs to be done is:

- to verify *uniform* convergence, as well as the rate of *uniform* convergence, of approximating empirical cdf's to the population cdf;

- to establish convergence of empirical *integrated* cdf's to the corresponding *integrated* population cdf's;

- to determine limiting distributions.

We shall implement this program in the case of two very broad classes of stationary time series, whose dependence structures are called S- and M-mixing.

Section 9 connects the results of previous sections with research areas such as L-statistics, distortion and spectral risk measures. Section 10 concludes the paper with a brief summary. Longer proofs, accompanying lemmas, and some cursory technicalities are in Appendix A.

3   The upper-layer integral

In this section we develop statistical inference for the upper-layer integral $\int_p^1 F^{-1}(u)\mathrm{d}u$ at any fixed lower probability level

$$p \in (0,1).$$

Of course, we need to assume that F is from the class $\mathcal{F}_1^+$ of those cdf's for which the upper quantiles are integrable, that is, the integral $\int_p^1 F^{-1}(u)\mathrm{d}u$ is finite. Note that this condition is the same as to require that the positive part $X^+ := X \vee 0$ of the random variable $X \sim F$ has a finite first moment $\mathbb{E}(X^+) < \infty$.

The functional

$$\mathcal{F}_1^+ \ni F \mapsto \int_p^1 F^{-1}(u)\mathrm{d}u \in \mathbb{R}$$

is not linear, which makes it challenging to derive statistical inference. It appears, however, that when developing statistical inference, the functional can be approximated by the linear one

$$\mathcal{F}_1^+ \ni G \mapsto \int_{F^{-1}(p)}^\infty \big(1 - G(x)\big)\mathrm{d}x \in \mathbb{R}. \tag{3.1}$$

Indeed, we shall see from the following corollaries that the difference between the empirical upper-layer integral $\int_p^1 F_n^{-1}(u)\mathrm{d}u$ and its population counterpart $\int_p^1 F^{-1}(u)\mathrm{d}u$ gets asymptotically close to the difference between $\int_{F^{-1}(p)}^\infty \big(1 - F_n(x)\big)\mathrm{d}x$ and $\int_{F^{-1}(p)}^\infty \big(1 - F(x)\big)\mathrm{d}x$, where $F_n \in \mathcal{F}_1^+$, $n \in \mathbb{N}$, are cdf's approaching F when the parameter n grows indefinitely.

Note 3.1. If F_n is a deterministic cdf, that is, $F_n : \mathbb{R} \to [0,1]$, then the meaning of statements like $F_n \in \mathcal{F}_1^+$ and $\int_{F^{-1}(p)}^\infty \big(1 - F_n(x)\big) \mathrm{d}x \in \mathbb{R}$ is clear. If, however, F_n is a random cdf, that is, $F_n : \Omega \times \mathbb{R} \to [0,1]$, where Ω denotes the sample space, then $F_n \in \mathcal{F}_1^+$ means that, for almost all $\omega \in \Omega$, the deterministic cdf $F_n(\omega, \cdot) : \mathbb{R} \to [0,1]$ is an element of $\mathcal{F}_1^+$. Likewise, $\int_{F^{-1}(p)}^\infty \big(1 - F_n(x)\big) \mathrm{d}x \in \mathbb{R}$ means $\int_{F^{-1}(p)}^\infty \big(1 - F_n(\omega, x)\big) \mathrm{d}x \in \mathbb{R}$ for almost all $\omega \in \Omega$. Furthermore, to enable ourselves to calculate, e.g., the expected values of integrals like $\int_p^1 F_n^{-1}(u) \mathrm{d}u$ and $\int_{F^{-1}(p)}^\infty \big(1 - F_n(x)\big) \mathrm{d}x$, we need joint measurability of $F_n : \Omega \times \mathbb{R} \to [0,1]$. We do not emphasize these subtle properties in the paper because they are automatically satisfied by empirical cdf's arising from n random variables $X_1, \ldots, X_n$, and all random cdf's in this paper are of such type. Note in this regard, for example, that $\int_0^1 F_n^{-1}(u) \mathrm{d}u \in \mathbb{R}$ is equivalent to $(X_1 + \cdots + X_n)/n \in \mathbb{R}$, and the latter statement is, of course, true almost surely. Nevertheless, it is prudent to keep a watchful eye on all these matters.

The aforementioned corollaries rely on the following fundamental theorem, whose proof is given in Appendix A.2.

Theorem 3.1. *Let F and G be any two cdf's from $\mathcal{F}_1^+$. Then*

$$\int_p^1 \big(G^{-1}(u) - F^{-1}(u)\big) \mathrm{d}u = \int_{F^{-1}(p)}^\infty \big(F(x) - G(x)\big) \mathrm{d}x - \mathrm{Rem}(p; F, G), \qquad (3.2)$$

where the remainder term

$$\mathrm{Rem}(p; F, G) = \int_{G^{-1}(p)}^{F^{-1}(p)} \big(G(x) - p\big) \mathrm{d}x \qquad (3.3)$$

is non-negative, that is, $\mathrm{Rem}(p; F, G) \geq 0$, and satisfies the bounds

$$\mathrm{Rem}(p; F, G) \leq \int_{G^{-1}(p)}^{F^{-1}(p)} \big(G(x) - F(x)\big) \mathrm{d}x \qquad (3.4)$$

$$\leq \big|G^{-1}(p) - F^{-1}(p)\big| \sup_{x \in \mathbb{R}} \big|G(x) - F(x)\big|. \qquad (3.5)$$

Note 3.2. The functional $(F, G) \mapsto \mathrm{Rem}(p; F, G) \in [0, \infty)$ is not symmetric. Note that its symmetrization $\mathrm{Rem}(p; F, G) + \mathrm{Rem}(p; G, F)$ is equal to the right-hand side of bound (3.4), that is,

$$\mathrm{Rem}(p; F, G) + \mathrm{Rem}(p; G, F) = \int_{G^{-1}(p)}^{F^{-1}(p)} \big(G(x) - F(x)\big) \mathrm{d}x. \qquad (3.6)$$

This equation implies bound (3.4) because the remainder term $\mathrm{Rem}(p; G, F)$ is non-negative, just like $\mathrm{Rem}(p; F, G)$ is.

We are now ready to formulate and discuss several corollaries to Theorem 3.1. Their formulations will involve approximating sequences $F_1, F_2, \ldots$ of cdf's, which may or may not be random, depending on the application. For example, we might be dealing with deterministic "contaminated" cdf's F_n given by the equation

$$F_n(x) = (1 - \delta_n)F(x) + \delta_n C(x),$$

where C is a (deterministic) cdf and $(0, 1) \ni \delta_n \downarrow 0$ are contamination proportions. An obvious example of random cdf's would be the empirical cdf

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \leq x\}$$

based on random variables $X_1, \ldots, X_n$, where $\mathbb{1}\{X_i \leq x\}$ is the indicator that takes value 1 when the statement $X_i \leq x$ is correct and 0 otherwise. Nevertheless, since every deterministic case is a special (i.e., degenerate) case of the random one, irrespective of the nature of the approximating cdf's F_n we can always treat them as random functions, that is, defined on $\Omega \times \mathbb{R}$, where Ω is the sample space, given the underlying probability space $\{\Omega, \mathcal{A}, \mathbb{P}\}$. Hence, for example, the requirement $F_n \in \mathcal{F}_1^+$ that we shall see in the following corollary will mean that, almost surely, F_n is an element of the class $\mathcal{F}_1^+$, that is, the integral $\int_p^1 F_n^{-1}(u)\mathrm{d}u$ is finite almost surely, for every value of $p \in (0, 1)$.

Corollary 3.1 (consistency). *Suppose that $F \in \mathcal{F}_1^+$, and let $F_1, F_2, \ldots \in \mathcal{F}_1^+$ be any sequence of cdf's such that*

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{\mathbb{P}} 0 \tag{3.7}$$

when $n \to \infty$. Then

$$\int_p^1 F_n^{-1}(u)\mathrm{d}u - \int_p^1 F^{-1}(u)\mathrm{d}u = \int_{F^{-1}(p)}^{\infty} \big(F(x) - F_n(x)\big)\mathrm{d}x + o_{\mathbb{P}}(1) \tag{3.8}$$

and, therefore, the consistency statement

$$\int_p^1 F_n^{-1}(u)\mathrm{d}u \xrightarrow{\mathbb{P}} \int_p^1 F^{-1}(u)\mathrm{d}u \tag{3.9}$$

holds if and only if

$$\int_{F^{-1}(p)}^{\infty} \big(1 - F_n(x)\big)\mathrm{d}x \xrightarrow{\mathbb{P}} \int_{F^{-1}(p)}^{\infty} \big(1 - F(x)\big)\mathrm{d}x. \tag{3.10}$$

Proof. The corollary follows from Theorem 3.1 with F_n instead of G, because $F^{-1}(p)$ is finite and $F_n^{-1}(p)$ is asymptotically bounded. Hence, condition (3.7) implies $\mathrm{Rem}(p; F, G) = o_{\mathbb{P}}(1)$ due to bound (3.5). This establishes statement (3.8), which implies the equivalence of statements (3.9) and (3.10), and concludes the proof of Corollary 3.1. $\qquad\square$

Example 3.1 (SRS)**.** Let $X_1, \ldots, X_n$ be SRS, that is, iid random variables. The empirical cdf $F_{n,\mathrm{SRS}}$ is defined by the equation

$$F_{n,\mathrm{SRS}}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \le x\},$$

which is the arithmetic mean of n independent copies of the Bernoulli random variable $\mathbb{1}\{X \le x\}$ taking value 1 with probability $F(x)$ and 0 otherwise. In other words, $\mathbb{1}\{X \le x\}$ is the indicator that takes value 1 when the statement $X \le x$ is true and 0 otherwise. Analogously to equation (1.7), the empirical quantile function $F_{n,\mathrm{SRS}}^{-1}$ is defined by

$$F_{n,\mathrm{SRS}}^{-1}(u) = \inf\{x \in \mathbb{R} \; : \; F_{n,\mathrm{SRS}}(x) \ge u\} \tag{3.11}$$

$$= X_{\lceil nu \rceil : n} \tag{3.12}$$

for all $u \in (0, 1]$, where $X_{1:n} \le \cdots \le X_{n:n}$ are the order statistics of the random variables $X_1, \ldots, X_n$, and, for any real $x \in \mathbb{R}$, $\lceil x \rceil$ denotes the smallest integer k such that $x \le k$, that is, $x \mapsto \lceil x \rceil$ is the classical ceiling function. When $u = 0$, we define

$$F_{n,\mathrm{SRS}}^{-1}(0) = X_{1:n} = \min_{1 \le i \le n} X_i.$$

Condition (3.7) with $F_{n,\mathrm{SRS}}$ in place of F_n is satisfied by the Glivenko-Cantelli theorem. Hence, we are left to verify condition (3.10). By linearity of functional (3.1), the integral $\int_{F^{-1}(p)}^{\infty} \big(1 - F_{n,\mathrm{SRS}}(x)\big)\mathrm{d}x$ is the arithmetic mean $n^{-1}\sum_{i=1}^{n} Y_{i,\mathrm{UPPER}}(p)$ of n independent copies

of the random variable

$$Y_{\text{UPPER}}(p) = \int_{F^{-1}(p)}^{\infty} \mathbb{1}\{X > x\}\mathrm{d}x$$
$$= \left(X - F^{-1}(p)\right)^{+}. \tag{3.13}$$

The random variable $Y_{\text{UPPER}}(p)$ has a finite first moment because $F \in \mathcal{F}_1^+$ (see Lemma A.1 for details). Hence, statement (3.10) holds and so, by Corollary 3.1, we have that $\int_p^1 F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u$ is a consistent estimator of $\int_p^1 F^{-1}(u)\mathrm{d}u$. In summary, we conclude that when $F \in \mathcal{F}_1^+$, we have

$$\int_p^1 F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u \xrightarrow{\mathbb{P}} \int_p^1 F^{-1}(u)\mathrm{d}u. \tag{3.14}$$

Note 3.3. In addition to equation (3.12), there are also other ways to express $F_{n,\text{SRS}}^{-1}(u)$ in terms of the order statistics $X_{1:n} \leq \cdots \leq X_{n:n}$. For example, using the integer part $[x]$ of $x \in \mathbb{R}$, in the section on sample quantiles by Shao (2003), we find the equation

$$F_{n,\text{SRS}}^{-1}(u) = c_{nu}X_{[nu]:n} + (1 - c_{nu})X_{([nu]+1):n},$$

where $c_{nu} = 1$ if nu is an integer and $c_{nu} = 0$ otherwise.

Corollary 3.2 (bias). *Suppose that $F \in \mathcal{F}_1^+$, and let $F_1, F_2, \ldots \in \mathcal{F}_1^+$ be any sequence of cdf's that are unbiased estimators of F, that is,*

$$\mathbb{E}\big(F_n(x)\big) = F(x) \tag{3.15}$$

for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$, and such that

$$\mathbb{E}\big(F_n^{-1}(p)\big) \in \mathbb{R} \tag{3.16}$$

for all $n \in \mathbb{N}$. Then $\int_p^1 F_n^{-1}(u)\mathrm{d}u$ is a non-positively biased estimator of $\int_p^1 F^{-1}(u)\mathrm{d}u$.

Proof. Using Theorem 3.1 with F_n instead of G, we have

$$\text{Bias}_n^{\text{UPPER}}(p) := \mathbb{E}\bigg(\int_p^1 F_n^{-1}(u)\mathrm{d}u\bigg) - \int_p^1 F^{-1}(u)\mathrm{d}u$$
$$= -\mathbb{E}\big(\text{Rem}(p; F, F_n)\big)$$
$$\in (-\infty, 0]$$

for every $n \in \mathbb{N}$. Note that $\mathbb{E}\big(\mathrm{Rem}(p; F, F_n)\big)$ is finite because of bound (3.5) and condition (3.16). This establishes Corollary 3.2. $\qquad \square$

Note 3.4. From the proof of Corollary 3.2 we see why and how the bias

$$
\mathrm{Bias}_n^{\mathrm{UPPER}}(p) = \begin{cases} -\mathrm{Rem}(p; F, F_n) & \text{when } F_n \text{ is deterministic,} \\ -\mathbb{E}\big(\mathrm{Rem}(p; F, F_n)\big) & \text{when } F_n \text{ is random,} \end{cases} \tag{3.17}
$$

arises, where

$$
\mathrm{Rem}(p; F, F_n) = \int_{F_n^{-1}(p)}^{F^{-1}(p)} \big(F_n(x) - p\big)\mathrm{d}x.
$$

These formulas can, in turn, be used to further analyze the bias $\mathrm{Bias}_n^{\mathrm{UPPER}}(p)$, assess its magnitude, and possibly find ways to mitigate it in real data-driven applications. For example, if the cdf F has a pdf $f = F'$ that is strictly positive and continuous at the quantile $F^{-1}(p)$, then following the studies of Gribkova and Helmers (2006, 2007) on trimmed means, we can show that in the (random) case of $F_n = F_{n,\mathrm{SRS}}$, the SRS-based bias $\mathrm{Bias}_{n,\mathrm{SRS}}^{\mathrm{UPPER}}(p)$ attains the asymptotic expansion

$$
\mathrm{Bias}_{n,\mathrm{SRS}}^{\mathrm{UPPER}}(p) = -\frac{p(1-p)}{2nf(F^{-1}(p))} + o\left(\frac{1}{n}\right) \tag{3.18}
$$

when $n \to \infty$. We can now construct an empirical estimator for the main term in asymptotic expansion (3.18) by using, for example, the empirical estimator $F_{n,\mathrm{SRS}}^{-1}(p)$ for $F^{-1}(p)$ and a kernel density estimator for f (e.g., Prakasa Rao, 1983). For a discussion of the role of the density-quantile function $p \mapsto f(F^{-1}(p))$ in statistics, and for a comprehensive statistical inference theory for it, including its behaviour near the end points $p = 0$ and $p = 1$, we refer to the pioneering studies of Parzen (1979a,b).

Example 3.2 (SRS). When F_n is $F_{n,\mathrm{SRS}}$, condition (3.15) is obviously satisfied. Condition (3.16) is satisfied when (Stigler, 1974, Proposition 2, p. 679)

there is $\varepsilon > 0$ such that $x^\varepsilon(1 - F(x) + F(-x)) \to 0$ when $x \to \infty$, which is equivalent to saying that $x^\varepsilon \mathbb{P}(|X| \geq x) \to 0$, where $X \sim F$.

For a related study that connects condition (3.16) with the order of finite moments of F, we refer to Gribkova (1995) (see also next Note 3.5). Denote the class of all cdf's satisfying the aforementioned tail-based condition of Stigler (1974) by $\mathcal{T}_\varepsilon$. Hence, Corollary 3.2 tells

us that when $F \in \mathcal{F}_1^+ \cap \mathcal{T}_\varepsilon$ for some $\varepsilon > 0$, the estimator $\int_p^1 F_{n,\mathrm{SRS}}^{-1}(u)\mathrm{d}u$ of the upper-layer integral $\int_p^1 F^{-1}(u)\mathrm{d}u$, although being consistent by Example 3.1, has non-positive bias

$$\mathrm{Bias}_{n,\mathrm{SRS}}^{\mathrm{UPPER}}(p) := -\mathbb{E}\big(\mathrm{Rem}(p; F, F_{n,\mathrm{SRS}})\big) \in (-\infty, 0]. \tag{3.19}$$

Note 3.5. The condition that $F \in \mathcal{F}_1^+ \cap \mathcal{T}_\varepsilon$ for some $\varepsilon > 0$ can equivalently be reformulated as $F \in \mathcal{F}_1^+ \cap \mathcal{T}_\varepsilon^-$ for some $\varepsilon > 0$, where $\mathcal{T}_\varepsilon^-$ consists of all those cdf's that satisfy $|x|^\varepsilon F(x) \to 0$ when $x \to -\infty$. This reformulation is possible due to the fact that $F \in \mathcal{F}_1^+$ implies $x^\varepsilon(1 - F(x)) \to 0$ when $x \to \infty$ for every $\varepsilon \in (0, 1)$, and we need this condition to hold for just one $\varepsilon > 0$. For those researchers who think in terms of moments, a sufficient condition that ensures the existence of $\varepsilon > 0$ such that $F \in \mathcal{F}_1^+ \cap \mathcal{T}_\varepsilon^-$ would be the requirement that

$$\mathbb{E}(X^+) < \infty \quad \text{and} \quad \mathbb{E}((X^-)^\varepsilon) < \infty \tag{3.20}$$

for some $\varepsilon > 0$, where $X^- = (-X) \vee 0$ is the negative part of the random variable $X \sim F$. Note that requirement (3.20) can equivalently be rewritten as $F \in \mathcal{F}_1^+ \cap \mathcal{F}_\varepsilon^-$.

The next corollary plays a pivotal role in deriving the asymptotic distribution of the empirical upper-layer integral. We note at the outset that the corollary crucially relies on the rate of convergence to 0, denoted by $O_\mathbb{P}(1/A_n)$, of the supremum in condition (3.7). For example, we would typically have $A_n = \sqrt{n}$ for iid sequences and, more broadly, for weakly-dependent stationary sequences (X_n), whereas we would have $A_n = n^\theta$ with some $\theta \in (0, 1/2)$ for long-memory random sequences. We shall give a more detailed note on the topic near the end of this section. At the moment we emphasize that next Corollary 3.3 does not depend on any sampling design or dependence structure.

Corollary 3.3 (asymptotic distribution). *Suppose that $F \in \mathcal{F}_1^+$, and let $F_1, F_2, \ldots \in \mathcal{F}_1^+$ be any sequence of cdf's such that*

$$F_n^{-1}(p) \xrightarrow{\mathbb{P}} F^{-1}(p) \tag{3.21}$$

and

$$A_n \sup_{x \in \mathbb{R}} \big| F_n(x) - F(x) \big| = O_\mathbb{P}(1) \tag{3.22}$$

when $n \to \infty$, where $A_n \to \infty$ are some normalizing constants. Then

$$A_n \left(\int_p^1 F_n^{-1}(u)\mathrm{d}u - \int_p^1 F^{-1}(u)\mathrm{d}u \right) = A_n \int_{F^{-1}(p)}^\infty \big(F(x) - F_n(x)\big)\mathrm{d}x + o_\mathbb{P}(1) \tag{3.23}$$

16

and, therefore, the convergence-in-distribution statement

$$A_n \left(\int_p^1 F_n^{-1}(u)\mathrm{d}u - \int_p^1 F^{-1}(u)\mathrm{d}u \right) \xrightarrow{d} \mathcal{L}_{\mathrm{UPPER}}(p) \tag{3.24}$$

holds if and only if

$$A_n \left(\int_{F^{-1}(p)}^{\infty} \left(1 - F_n(x)\right)\mathrm{d}x - \int_{F^{-1}(p)}^{\infty} \left(1 - F(x)\right)\mathrm{d}x \right) \xrightarrow{d} \mathcal{L}_{\mathrm{UPPER}}(p), \tag{3.25}$$

where $\mathcal{L}_{\mathrm{UPPER}}(p)$ is a random variable determined by statement (3.25).

Proof. Using Theorem 3.1 with F_n instead of G, we only need to show that $A_n \mathrm{Rem}(p; F, F_n)$ converges in probability to 0. Since $\mathrm{Rem}(p; F, F_n)$ is non-negative and does not exceed

$$\left| F_n^{-1}(p) - F^{-1}(p) \right| \sup_{x \in \mathbb{R}} \left| F_n(x) - F(x) \right|,$$

the statement $A_n \mathrm{Rem}(p; F, F_n) = o_{\mathbb{P}}(1)$ follows from conditions (3.21) and (3.22). This establishes equation (3.23), which implies the equivalence of statements (3.24) and (3.25), and concludes the proof of Corollary 3.3. $\square$

Note 3.6. The limiting random variable $\mathcal{L}_{\mathrm{UPPER}}(p)$ may or may not be normal. The nature of this random variable is determined by the sequence (X_n): if it is iid or, more broadly, stationary and weakly dependent, then $\mathcal{L}_{\mathrm{UPPER}}(p)$ is normal, but if the stationary sequence (X_n) is long-range dependent, then the limiting random variable $\mathcal{L}_{\mathrm{UPPER}}(p)$, if it exists, is typically non-normal.

Concerning condition (3.21), it is a simple probability exercise to check the validity of the following lemma by imitating related considerations in the section on SRS-based quantiles by Shao (2003).

Lemma 3.1. *When the quantile function F^{-1} is continuous at the point p, condition (3.7) implies condition (3.21) and, therefore, the latter condition can be dropped from Corollary 3.3 because condition (3.22) implies (3.7) and, therefore, it also implies condition (3.21).*

Note 3.7. In the SRS-based context, that is, when F_n is $F_{n,\mathrm{SRS}}$, Shao (2003) derives exponential bounds for the probability $\mathbb{P}(|F_{n,\mathrm{SRS}}^{-1}(p) - F^{-1}(p)| \geq \varepsilon)$ with the aim at establishing strong consistency. We are concerned with (weak) consistency and so do not need exponential bounds. In other words, we just need convergence of $\mathbb{P}(|F_n^{-1}(p) - F^{-1}(p)| \geq \varepsilon)$ to 0 for

every $\varepsilon > 0$ when $n \to \infty$, and for this, condition (3.22) imposed on the generic sequence F_n, $n \in \mathbb{N}$, is sufficient.

In Section 4 we shall give an illustrative numerical study showing that statement (3.21) fails for a cdf F for which the quantile function F^{-1} is discontinuous at a point p. The same example will also show that the empirical upper-layer integral $\int_p^1 F_{n,\mathrm{SRS}}^{-1}(u)\mathrm{d}u$ fails asymptotic normality. This phenomenon is natural given the results of Stigler (1973); see also Gribkova and Helmers (2011) for further notes on the topic.

Example 3.3 (SRS). When F_n is $F_{n,\mathrm{SRS}}$, condition (3.22) is a consequence of the Kolmogorov-Smirnov theorem. Since condition (3.22) is satisfied, condition (3.21) is satisfied as well, provided that, according to Lemma 3.1, the quantile function F^{-1} is continuous at the point p. Furthermore, since the integral $\int_{F^{-1}(p)}^{\infty} \left(1 - F_{n,\mathrm{SRS}}(x)\right)\mathrm{d}x$ is the arithmetic mean of n independent copies of random variable (3.13), we conclude that the integral $\int_{F^{-1}(p)}^{\infty} \left(1 - F_{n,\mathrm{SRS}}(x)\right)\mathrm{d}x$ satisfies the central limit theorem provided that random variable (3.13) has a finite second moment. This is so provided that $F \in \mathcal{F}_2^+$ (see Lemma A.1 for details), where $\mathcal{F}_2^+$ denotes the set of all cdf's such that the integral $\int_p^1 \left(F^{-1}(u)\right)^2 \mathrm{d}u$ is finite. Note that $F \in \mathcal{F}_2^+$ is equivalent to saying that the positive part X^+ of the random variable $X \sim F$ has a finite second moment $\mathbb{E}((X^+)^2) < \infty$. In summary, when $F \in \mathcal{F}_2^+$ and the quantile function F^{-1} is continuous at the point p, we have the asymptotic normality result

$$\sqrt{n}\left(\int_p^1 F_{n,\mathrm{SRS}}^{-1}(u)\mathrm{d}u - \int_p^1 F^{-1}(u)\mathrm{d}u\right) \xrightarrow{d} \mathcal{N}(0, \sigma_{\mathrm{UPPER}}^2(p)), \tag{3.26}$$

where $\sigma_{\mathrm{UPPER}}^2(p)$ is the variance of the random variable $Y_{\mathrm{UPPER}}(p)$. Note that the variance can be expressed in terms of the cdf F by the formula

$$\sigma_{\mathrm{UPPER}}^2(p) = \int_{F^{-1}(p)}^{\infty} \int_{F^{-1}(p)}^{\infty} \left(F(x \wedge y) - F(x)F(y)\right)\mathrm{d}x\mathrm{d}y, \tag{3.27}$$

although we have found that the following alternative representations (recall equation (3.13))

$$\begin{aligned}
\sigma_{\mathrm{UPPER}}^2(p) &= \mathrm{Var}\left(\left(X - F^{-1}(p)\right)^+\right) \\
&= \mathrm{Var}\left(\left(F^{-1}(U) - F^{-1}(p)\right)^+\right)
\end{aligned} \tag{3.28}$$

are sometimes easier to use, where U denotes a uniform on $[0,1]$ random variable.

Note 3.8 (existence of uniform rv's). We often assume for granted that a uniform on $[0,1]$ random variable exists, but whether or not this is the case depends on the underlying probability space $\{\Omega, \mathcal{A}, \mathbb{P}\}$ upon which we are building our statistical experiments and theories. As shown by, for example, Föllmer and Schied (2016, Proposition A.31, p. 547), a necessary and sufficient condition for the existence of a uniform on $[0,1]$ random variable is for the underlying probability space to be atomless. Pertaining to the equality $X = F^{-1}(U)$ $\mathbb{P}$-a.s., we refer to Föllmer and Schied (2016, Lemma A.32, p. 548). The write-up by Ramdas and Wang (2024, Appendix A.1) on atomless probability spaces and their properties is particularly illuminating.

Note 3.9 (dependent data). Quite often in applications, data arrive in the form of time series, and thus researchers have been interested in deriving large sample properties for various tail risk measures based on such data (e.g., Davydov and Zitikis, 2002, 2003; Helmers and Zitikis, 2005; Davydov et al., 2007; Chen, 2008; Linton and Xiao, 2013; Ziegel et al., 2020; Li and Wang, 2023; Moutanabbir and Bouaddi, 2024; Garca-Risueño, 2025, and references therein). Corollaries 3.1 and 3.3, which are not attached to any particular sampling design or dependence structure, are applicable in such scenarios. Indeed, statements like those in condition (3.7) for consistency and condition (3.22) for asymptotic distribution have been verified for various classes of time series, where the empirical cdf plays the role of F_n. To give some flavour of what is known, we first note that in the case of weakly-dependent stationary time series, the normalizing constant A_n is $\sqrt{n}$ and the limiting process of $A_n(F_n - F)$ is normal (centered at 0 and with a variance reflecting the marginal variability as well as the dependence structure of the underlying time series), whereas in the case of long-memory stationary time series, the normalizing constant A_n is of the order n^θ for various $\theta \in (0, 1/2)$ and the limiting process, if exists, is non-normal. For a glimpse of the various normalizing constants and limiting distributions, albeit in the case of the lower-layer integral, we refer to Davydov and Zitikis (2004). The monographs by Dehling et al. (2002), Rio (2017), and Doukhan (2018) offer a wealth of information on empirical processes arising from various classes of times series. In Section 8 we shall provide more detailed user-oriented notes and references.

We next present convenient computational formulas for the SRS-based empirical upper-layer integral $\int_p^1 F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u$.

Computation 3.1 (SRS). When computing, estimators in the form of integrals are not convenient to work with. Hence, using equation (3.12), we shall next express $\int_p^1 F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u$

in terms of the order statistics $X_{1:n} \leq \cdots \leq X_{n:n}$, whose realized values are known in practice. Namely, we have

$$
\int_p^1 F_{n,\mathrm{SRS}}^{-1}(u)\mathrm{d}u = \int_p^{\lceil np \rceil / n} F_{n,\mathrm{SRS}}^{-1}(u)\mathrm{d}u + \sum_{i=\lceil np \rceil+1}^{n} \int_{(i-1)/n}^{i/n} F_{n,\mathrm{SRS}}^{-1}(u)\mathrm{d}u
$$

$$
= \underbrace{\frac{1}{n} \sum_{i=\lceil np \rceil+1}^{n} X_{i:n}}_{\text{main term}} + \underbrace{\left(\frac{\lceil np \rceil}{n} - p \right) X_{\lceil np \rceil:n}}_{\text{asymptotically negligible term}} \tag{3.29}
$$

$$
= \left(1 - \frac{\lceil np \rceil}{n} \right) \underbrace{\frac{1}{n - \lceil np \rceil} \sum_{i=\lceil np \rceil+1}^{n} X_{i:n}}_{\text{left-trimmed mean}} + \underbrace{\left(\frac{\lceil np \rceil}{n} - p \right) X_{\lceil np \rceil:n}}_{\text{asymptotically negligible term}} . \tag{3.30}
$$

Equations (3.29) and (3.30) give computationally-friendly formulas for the empirical upper-layer integral, which is connected to the left-trimmed mean (e.g., Stigler, 1973) via equation (3.30).

4 The upper-layer integral: simulated experiments

In this section we present several experiments that empirically illustrate some of the theoretical considerations in previous Section 3.

Simulation 4.1 (SRS; bias)**.** We shall now empirically illustrate that $\mathrm{Bias}_{n,\mathrm{SRS}}^{\mathrm{UPPER}}(p)$ is non-positive, as stated in Example 3.2. To avoid confounding negativity of bias with skewness of distributions, we work with a symmetric (around its mean) and light-tailed distribution. Namely, let V be the uniform on $[0,2]$ random variable, whose quantile function is $F^{-1}(u) = 2u$ for $0 \leq u \leq 1$. We fix the lower-layer probability level

$$
p = \frac{3}{4}
$$

and choose several values of n (to be specified in a moment) to get an insight into the behaviour of the bias with respect to the sample size. The population upper-layer integral $\int_p^1 F^{-1}(u)\mathrm{d}u$ is equal to $1 - p^2 = 7/16$. We obtain the first SRS by observing n independent copies $V_1, \ldots, V_n$ of V, and then calculate the empirical upper-layer integral $\int_p^1 F_{n,\mathrm{SRS}}^{-1}(u)\mathrm{d}u$

according to formula (3.29). This gives us the first difference

$$\underbrace{\frac{1}{n}\sum_{i=\lceil 3n/4\rceil+1}^{n}V_{i:n}+\left(\frac{\lceil 3n/4\rceil}{n}-\frac{3}{4}\right)V_{\lceil 3n/4\rceil:n}}_{\text{empirical}}-\underbrace{\frac{7}{16}}_{\text{population}}$$

between the empirical and population upper-layer integrals. We denote the difference by $B_n(1)$ and call it the first individual bias. We repeat the procedure $m = 10\,000$ times, keeping the same p and n values, and obtain m values of the individual biases $B_n(1), \ldots, B_n(m)$. We calculate their arithmetic mean B_n^{AVE} and also the median B_n^{MED}. In Table 4.1 we report

n	40	100	200	500	1000
B_n^{AVE}	-0.004855	-0.002113	-0.000944	-0.000311	-0.000154
B_n^{MED}	-0.002482	-0.001203	-0.000463	-0.000207	-0.000099

Table 4.1: The averages and medians of $m = 10\,000$ biases for various sample sizes n.

the values of B_n^{AVE} and B_n^{MED} for different n's. Since m is very large, we expect the averages B_n^{AVE} reported in Table 4.1 to be close to the corresponding theoretical biases $\text{Bias}_{n,\text{SRS}}^{\text{UPPER}}$, which are defined by equation (3.19). Indeed, the table shows that all B_n^{AVE}'s are negative. Furthermore, their absolute values are decreasing when n grows, which is expected given the proven consistency of the empirical upper-layer integral. Nevertheless, as a follow-up research topic, it is worth thinking of modifying the empirical upper-layer integral to reduce its bias. Finally, we note that the reason behind presenting the medians B_n^{MED} in Table 4.1 is that by comparing them with the averages B_n^{AVE}, we can get insights about the skewness of the distribution of individual biases.

Simulation 4.2 (SRS; asymptotic normality holds)**.** Let V be a uniform on $[0,2]$ random variable, and so its cdf is $F(x) = x/2$ for all $0 \le x \le 2$. We set the parameter p value to

$$p = \frac{1}{2}$$

and let the sample size be $n = 100\,000$, which is an even number and thus satisfies $\lceil n/2\rceil = n/2$. We simulate n independent copies $V_1, \ldots, V_n$ of V. Statement (3.26) becomes

$$\Delta_n := \frac{\sqrt{n}}{\sigma_{\text{UPPER}}(1/2)}\left(\int_{1/2}^{1}F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u - \int_{1/2}^{1}F^{-1}(u)\mathrm{d}u\right) \xrightarrow{d} \mathcal{N}(0,1), \qquad (4.1)$$

21

where

$$\int_{1/2}^{1} F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u = \frac{1}{n}\sum_{i=(n/2)+1}^{n} V_{i:n} \quad \big(\text{because } \lceil n/2 \rceil = n/2\big),$$

$$\int_{1/2}^{1} F^{-1}(u)\mathrm{d}u = \frac{3}{4},$$

$$\sigma_{\text{UPPER}}^{2}(1/2) = \frac{5}{48}. \tag{4.2}$$

To arrive at the value of $\sigma_{\text{UPPER}}^{2}(p)$ noted in equation (4.2), we have used representation (3.28) with $F^{-1}(u) = 2u$ for all $0 \le u \le 1$. Simulation results based on $m = 10\,000$ replications of Δ_n are depicted in the form of a relative histogram in Figure 4.1, which supports the
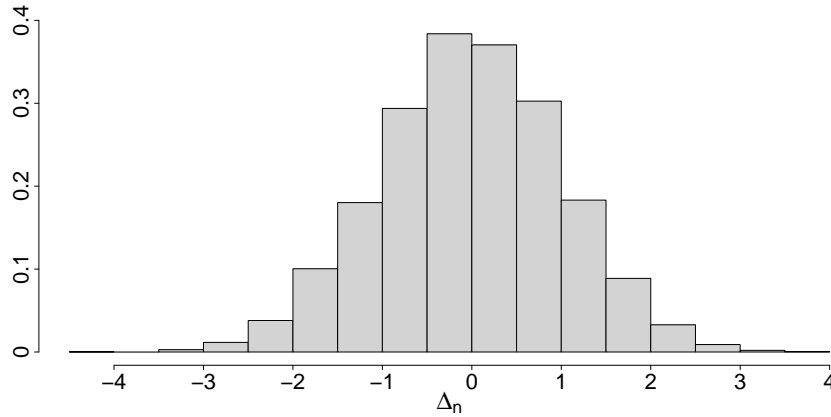
Figure 4.1: The relative histogram of $m = 10\,000$ values of Δ_n in (4.1) when $n = 100\,000$.

limiting standard normal distribution claimed in statement (4.1).

Simulation 4.3 (SRS; asymptotic normality fails)**.** To illustrate the pivotal role of continuity of the quantile function F^{-1} at the point p and thus of condition (3.21) in the statement of asymptotic normality of the empirical upper-layer integral, we next introduce a new random variable, denoted by Z. Its definition is based on a uniform on $[0,2]$ random variable V. Namely, we fix any constant $a \in [0,1]$ and define a "gapped" random variable Z by the formula

$$Z = (1-a)V\mathbb{1}\{V \le 1\} + \big(2a + (1-a)V\big)\mathbb{1}\{V > 1\}.$$

Note that when $a = 0$, the gap vanishes and Z turns into the original uniform random variable V, the case that we have discussed in previous Simulation 4.2. When $a = 1$, the random variable Z turns into a discrete random variable taking values 0 and 2 with probabilities 1/2.

To better understand and appreciate the random variable Z with respect to the parameter $a \in (0, 1)$, we note the following properties of Z when it is viewed as a function of the argument V:

- if $V = 0$, then $Z = 0$;

- if $V = 1$, then $Z = 1 - a$;

- if $V \downarrow 1$, then $Z \downarrow 1 + a$;

- if $V = 2$, then $Z = 2$;

Hence, all the values of Z are in the interval $[0, 2]$, although there are no values in $(1-a, 1+a]$, which represents the gap in the distribution of Z. The cdf $z \mapsto H(z)$ of Z, which is visualized in Figure 4.2, is strictly increasing at every point $z \in (0, 2) \setminus [1-a, 1+a]$ and is constant on

(a) The cdf H. (b) The quantile function H^{-1}.
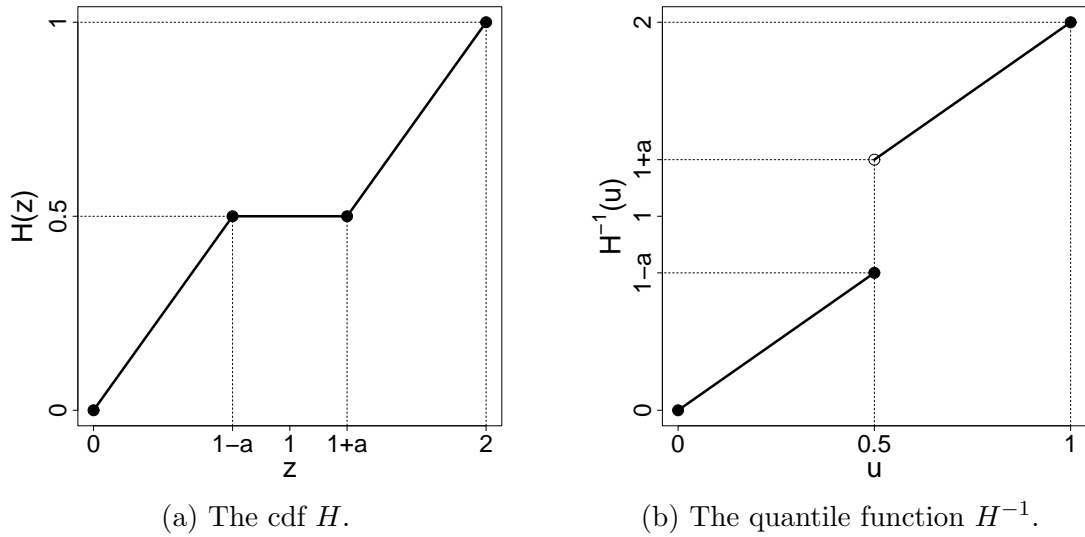
Figure 4.2: The distribution of the gapped random variable Z.

the interval $[1 - a, 1 + a]$. The quantile $H^{-1}(1/2)$, which is the median of Z, is the left-hand endpoint $1 - a$ of the gap of Z. Consequently, the quantile function H^{-1} is not continuous at the point $p = 1/2$. This helps us to understand the reason why in the case

$$p = \frac{1}{2},\tag{4.3}$$

we have the following two non-convergence statements: first, non-consistency

$$H^{-1}_{n,\mathrm{SRS}}(1/2) \overset{\mathbb{P}}{\nrightarrow} H^{-1}(1/2),\tag{4.4}$$

and, second, asymptotic non-normality

$$\Delta_n := \frac{\sqrt{n}}{\sigma_{\text{UPPER}}(1/2)} \left(\int_{1/2}^1 H_{n,\text{SRS}}^{-1}(u)\mathrm{d}u - \int_{1/2}^1 H^{-1}(u)\mathrm{d}u \right) \overset{d}{\nrightarrow} \mathcal{N}(0,1), \qquad (4.5)$$

where (recall equation (3.27))

$$\sigma_{\text{UPPER}}^2(1/2) = \int_{1-a}^2 \int_{1-a}^2 \left(H(x \wedge y) - H(x)H(y) \right) \mathrm{d}x\mathrm{d}y. \qquad (4.6)$$

Although statements (4.4) and (4.5) can be verified theoretically, which we shall discuss in a moment, we find it illuminating to demonstrate the two statements empirically. For this, we set the gap radius to

$$a = \frac{1}{2}.$$

Hence, the distribution of Z has the gap $[1/2, 3/2]$ in its support $[0, 2]$. When simulating we conveniently work with *even* sample sizes n and thus have the equation $\lceil n/2 \rceil = n/2$, which simplifies formulas of the empirical estimators, especially of the upper-layer integral. We proceed by noting that given the nature of the random variable V, which is symmetric around its mean 1, and given the definition of the gapped Z, we expect to see approximately a half of the simulated Z values below $1/2$ and the remaining ones above $3/2$. Indeed, when $n \to \infty$, we have the following two statements (see Appendix A.6.2)

$$\mathbb{P}\left(Z_{\lceil n/2 \rceil:n} > 3/2 \right) \to 1/2, \qquad (4.7)$$

$$\mathbb{P}\left(Z_{\lceil n/2 \rceil:n} < 1/2 \right) \to 1/2. \qquad (4.8)$$

This explains non-consistency statement (4.4), which we have visualized in Figure 4.3 using the formulas

$$H_{n,\text{SRS}}^{-1}(1/2) = Z_{\lceil n/2 \rceil:n} \quad (= Z_{(n/2):n}),$$
$$H^{-1}(1/2) = 1/2$$

for various (even) sample sizes n. As to asymptotic non-normality statement (4.5), we have
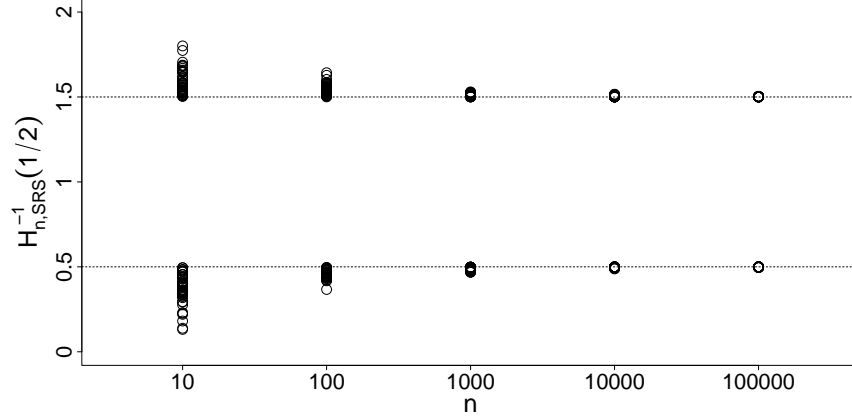
Figure 4.3: Simulated $m = 100$ values of $H^{-1}_{n,\mathrm{SRS}}(1/2)$ for various sample sizes n.

used the formulas

$$\int_{1/2}^{1} H^{-1}_{n,\mathrm{SRS}}(u)\mathrm{d}u = \frac{1}{n} \sum_{i=(n/2)+1}^{n} Z_{i:n} \quad \big(\text{because } \lceil n/2 \rceil = n/2\big),$$

$$\int_{1/2}^{1} H^{-1}(u)\mathrm{d}u = \frac{7}{8},$$

$$\sigma^2_{\mathrm{UPPER}}(1/2) = \frac{77}{192}. \tag{4.9}$$

Figure 4.4 depicts the simulation results based on the sample size $n = 100\,000$ and the
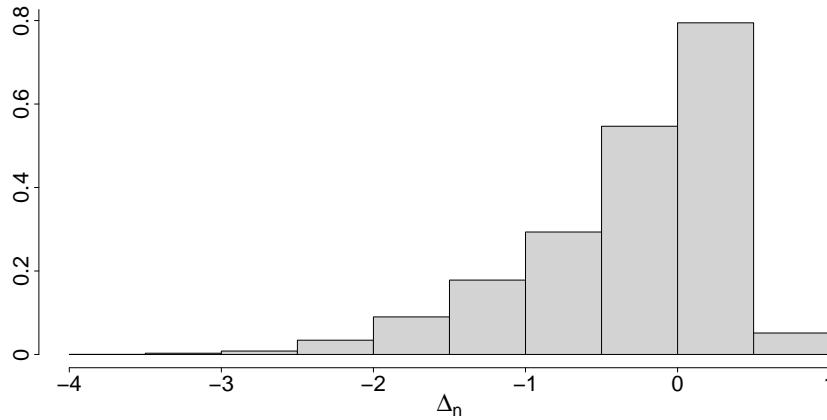
Figure 4.4: The relative histogram of $m = 10\,000$ values of Δ_n when $n = 100\,000$.

number of replications $m = 10\,000$. Both n and m are very large, thus depriving us of any hope that a standard normal density could somehow emerge if even larger values were used. In fact, we know from theoretical studies on trimmed means (Stigler, 1973) that asymptotic

25

normality does fail when (recall Note 2.3) continuous distributions have gaps (i.e., their cdf's have flat regions) adjacent to the quantiles where trimming occurs, as is the case in the present example (see Figure 4.2).

Note 4.1. When calculating the variance $\sigma^2_{\text{UPPER}}(1/2)$ in equation (4.9), we found that the quantile-based expression (recall equation (3.28))

$$\sigma^2_{\text{UPPER}}(1/2) = \text{Var}\left(\left(H^{-1}(U) - H^{-1}(1/2)\right)^+\right)$$
$$= \text{Var}\left(\left(H^{-1}(U) - (1-a)\right)^+\right) \tag{4.10}$$

is a particularly convenient starting point, where U denotes a uniform on $[0,1]$ random variable, with $1/2$ referring to our choice of p made in (4.3). In this way, we also easily arrive at the expression

$$\frac{29a^2 + 14a + 5}{48} \tag{4.11}$$

for $\sigma^2_{\text{UPPER}}(1/2)$ in the case of arbitrary $a \in [0,1]$ (see Appendix A.6.3 for details). Hence, $\sigma^2_{\text{UPPER}}(1/2) = 5/48$ when $a = 0$ (equation (4.2)), $\sigma^2_{\text{UPPER}}(1/2) = 77/192$ when $a = 1/2$ (equation (4.9)), and $\sigma^2_{\text{UPPER}}(1/2) = 1$ when $a = 1$ (in which case Z turns into a discrete random variable taking values 0 and 2 with probabilities $1/2$). Of course, we could have also calculated $\sigma^2_{\text{UPPER}}(1/2)$ using equation (4.6) but this cdf-based route is more involved.

5 The lower-layer integral

In this section we develop statistical inference for the lower-layer integral $\int_0^p F^{-1}(u)\mathrm{d}u$ for any fixed upper probability level

$$p \in (0,1).$$

Naturally, we assume that the cdf F is from the class $\mathcal{F}_1^-$ of those cdf's for which the integral $\int_0^p F^{-1}(u)\mathrm{d}u$ is finite, which is the same as to require that the negative part $X^- = (-X) \vee 0$ of the random variable $X \sim F$ has a finite first moment $\mathbb{E}(X^-) < \infty$.

Note 5.1. When progressing through this section, we shall see that the narrative closely follows that of Section 3. This may prompt us to think that the results could be derived from those of Section 3 by simply replacing X by $-X$. This train of thought, though useful, is fraught with danger because F^{-1} is not the ordinary inverse of F and so subtleties do

arise. It is our desire, therefore, to present the results in such a way that the user would be able to use them directly, without needing to work out the underlying theory.

We continue our analysis of the lower-layer integral by noting that the functional

$$\mathcal{F}_1^- \ni F \mapsto \int_0^p F^{-1}(u)\mathrm{d}u \in \mathbb{R}$$

is not linear. It appears that when developing statistical inference it can be approximated by the linear one

$$\mathcal{F}_1^- \ni G \mapsto \int_{-\infty}^{F^{-1}(p)} G(x)\mathrm{d}x \in \mathbb{R}. \tag{5.1}$$

Indeed, we shall see from the following corollaries that the difference between the empirical lower-layer integral $\int_0^p F_n^{-1}(u)\mathrm{d}u$ and its population counterpart $\int_0^p F^{-1}(u)\mathrm{d}u$ gets asymptotically close to the difference between the integrals $\int_{-\infty}^{F^{-1}(p)} F(x)\mathrm{d}x$ and $\int_{-\infty}^{F^{-1}(p)} F_n(x)\mathrm{d}x$, where $F_n \in \mathcal{F}_1^-$, $n \in \mathbb{N}$, are cdf's approaching F when the parameter n grows indefinitely. The corollaries rely on the following theorem, whose proof is given in Appendix A.2.

Theorem 5.1. *Let F and G be any two cdf's from $\mathcal{F}_1^-$. Then*

$$\int_0^p \big(G^{-1}(u) - F^{-1}(u)\big)\mathrm{d}u = \int_{-\infty}^{F^{-1}(p)} \big(F(x) - G(x)\big)\mathrm{d}x + \mathrm{Rem}(p; F, G), \tag{5.2}$$

where the non-negative remainder term $\mathrm{Rem}(p; F, G)$ is defined by equation (3.3) and satisfies bounds (3.4) and (3.5).

Theorem 5.1 is fundamental for the current section, and it also serves a direct link between the topic of the current paper and the Vervaat process V_n mentioned in Section 1. The following note elucidates the connection in detail.

Note 5.2. The (general) Vervaat process V_n is defined on the unit interval $(0, 1)$ by the equation

$$V_n(p) := \int_0^p \big(F_n^{-1}(u) - F^{-1}(u)\big)\mathrm{d}u + \int_{-\infty}^{F^{-1}(p)} \big(F_n(x) - F(x)\big)\mathrm{d}x$$
$$= \mathrm{Rem}(p; F, F_n),$$

where F_n, which replaces G in equation (5.2), is the empirical cdf based on n random variables $X_1, \ldots, X_n$, each following the cdf F. Originally, the process V_n was introduced

and considered by Csörgő and Zitikis (1996) in the case of iid random variables, with a number of subsequent studies by various authors devoted to asymptotic properties of V_n in the case of independent as well as dependent random variables, such as those generated by time series models. As to the pioneering studies of Vervaat (1972a,b), they concentrated on the process V_n in the case of n independent and uniformly on $[0,1]$ distributed random variables $U_1, \ldots, U_n$, in which case V_n reduces to

$$V_{n,\text{SRS}}^U(p) := \int_0^p \underbrace{\left(E_{n,\text{SRS}}^{-1}(u) - u\right)}_{\text{uniform quantile process}} \, \mathrm{d}u \; + \; \int_0^p \underbrace{\left(E_{n,\text{SRS}}(u) - u\right)}_{\text{uniform empirical process}} \, \mathrm{d}u$$

$$= \int_0^p \left(\underbrace{\left(E_{n,\text{SRS}}^{-1}(u) - u\right) + \left(E_{n,\text{SRS}}(u) - u\right)}_{\text{uniform Bahadur-Kiefer process}} \right) \mathrm{d}u,$$

where $E_{n,\text{SRS}}$ is the empirical cdf based on $U_1, \ldots, U_n$, and $E_{n,\text{SRS}}^{-1}$ is the corresponding empirical quantile function. We infer from Theorem 5.1 that $V_{n,\text{SRS}}^U(p) \geq 0$ for all $p \in [0,1]$, with the boundary values $V_{n,\text{SRS}}^U(0) = 0$ and $V_{n,\text{SRS}}^U(1) = 0$. Sample paths of the uniform Vervaat process $V_{n,\text{SRS}}^U(p)$, $0 \leq p \leq 1$, and those of the uniform empirical process $E_{n,\text{SRS}}(p) - p$, $0 \leq p \leq 1$, are illustrated in Figure 5.1. (We shall discuss computational formulas and

(a) Vervaat process $n\,V_{n,\text{SRS}}^U(p)$.　　(b) Empirical process $n^{1/2}(E_{n,\text{SRS}}(p)-p)$.

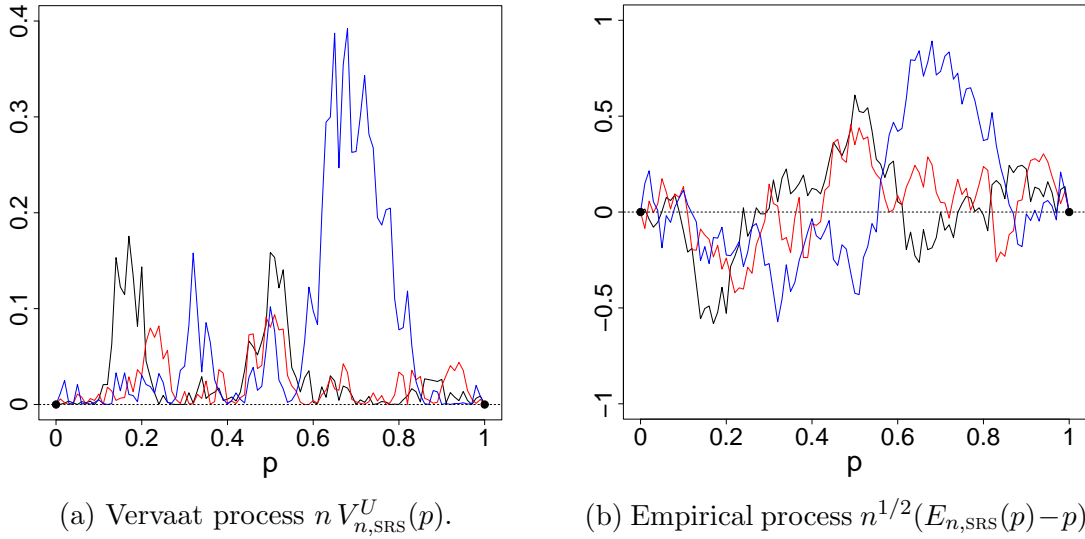Figure 5.1: Sample paths of normalized uniform Vervaat and empirical processes based on three uniform-on-$[0,1]$ SRS's of sizes $n = 100\,000$.

normalizing constants of these two stochastic processes at the end of this section; see Computation 5.2.) Mathematically, $V_{n,\text{SRS}}^U(p)$ is the lower-layer integral of the uniform version of the Bahadur-Kiefer process (Bahadur, 1966; Kiefer, 1967). In the case of general sequences

$X_1, \ldots, X_n \sim F$ of iid random variables, $V_{n,\mathrm{SRS}}(p)$ is the lower-layer integral of the (general) Bahadur-Kiefer process, as seen from the equations

$$
\begin{aligned}
V_{n,\mathrm{SRS}}(p) &= \int_0^p \big(F_{n,\mathrm{SRS}}^{-1}(u) - F^{-1}(u)\big)\mathrm{d}u + \int_{-\infty}^{F^{-1}(p)} \big(F_{n,\mathrm{SRS}}(x) - F(x)\big)\mathrm{d}x \\
&= \int_0^p \Big(\underbrace{\big(F_{n,\mathrm{SRS}}^{-1}(u) - F^{-1}(u)\big) + \frac{1}{f(F^{-1}(u))}\big(F_{n,\mathrm{SRS}}(F^{-1}(u)) - u\big)}_{\text{Bahadur-Kiefer process}} \Big)\mathrm{d}u,
\end{aligned}
$$

but this interpretation necessarily requires the pdf $f = F'$ to exist and be positive at all quantiles $F^{-1}(u)$, $0 < u < p$, as only in this case the (general) Bahadur-Kiefer process is meaningfully defined.

We are now ready to formulate and discuss several corollaries to Theorem 5.1.

Corollary 5.1 (consistency). *Suppose that $F \in \mathcal{F}_1^-$, and let $F_1, F_2, \ldots \in \mathcal{F}_1^-$ be any sequence of cdf's satisfying condition* (3.7). *Then*

$$
\int_0^p F_n^{-1}(u)\mathrm{d}u - \int_0^p F^{-1}(u)\mathrm{d}u = \int_{-\infty}^{F^{-1}(p)} \big(F(x) - F_n(x)\big)\mathrm{d}x + o_{\mathbb{P}}(1) \tag{5.3}
$$

and, therefore, the consistency statement

$$
\int_0^p F_n^{-1}(u)\mathrm{d}u \xrightarrow{\mathbb{P}} \int_0^p F^{-1}(u)\mathrm{d}u \tag{5.4}
$$

holds if and only if

$$
\int_{-\infty}^{F^{-1}(p)} F_n(x)\mathrm{d}x \xrightarrow{\mathbb{P}} \int_{-\infty}^{F^{-1}(p)} F(x)\mathrm{d}x. \tag{5.5}
$$

Proof. The corollary follows from Theorem 5.1 and bound (3.5) with F_n instead of G, because $F^{-1}(p)$ is finite and $F_n^{-1}(p)$ is asymptotically bounded, and therefore condition (3.7) implies $\mathrm{Rem}(p; F, G) = o_{\mathbb{P}}(1)$ due to bound (3.5). This establishes statement (5.3), which implies the equivalence of statements (5.4) and (5.5), and concludes the proof of Corollary 5.1. $\square$

Example 5.1 (SRS). Condition (3.7) with $F_{n,\mathrm{SRS}}$ instead of F_n is satisfied by the Glivenko-Cantelli theorem. Furthermore, by linearity of functional (5.1), the integral $\int_{-\infty}^{F^{-1}(p)} F_n(x)\mathrm{d}x$

is the arithmetic mean $n^{-1} \sum_{i=1}^{n} Y_{i,\text{LOWER}}$ of n independent copies of the random variable

$$
\begin{aligned}
Y_{\text{LOWER}}(p) &= \int_{-\infty}^{F^{-1}(p)} \mathbb{1}\{X \le x\}\mathrm{d}x \\
&= \left(F^{-1}(p) - X\right)^{+}.
\end{aligned}
\tag{5.6}
$$

The random variable $Y_{\text{LOWER}}(p)$ has a finite first moment because $F \in \mathcal{F}_1^{-}$ (see Lemma A.2 for details), and so statement (5.5) with $F_{n,\text{SRS}}$ instead of F_n holds. By Corollary 5.1, therefore, the empirical lower-level integral $\int_0^p F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u$ is a consistent estimator of $\int_0^p F^{-1}(u)\mathrm{d}u$. In summary, when $F \in \mathcal{F}_1^{-}$, we have

$$
\int_0^p F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u \xrightarrow{\mathbb{P}} \int_0^p F^{-1}(u)\mathrm{d}u.
\tag{5.7}
$$

Note 5.3 (dependent data). Although we work with the case $p \in (0,1)$, let us set $p = 1$ for a moment. In this case, statement (5.7) turns into the law of large numbers

$$
\mu_{n,\text{SRS}} = \int_0^1 F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u \xrightarrow{\mathbb{P}} \mu,
\tag{5.8}
$$

which of course holds under SRS. A departure from this sampling design leads to ergodic theory, and in particular to sequences $X_1, X_2, \ldots$ that are stationary and ergodic. Hence, it is only natural that results of the type of statement (5.7) when $p \in (0,1)$ have been established within the class of stationary and ergodic sequences (e.g., Aaronson et al., 1996; Gilat and Helmers, 1997; Davydov and Zitikis, 2002; Helmers and Zitikis, 2005, and references therein). It is also useful to point out that while the latter two references deal with the integral $\int_0^p F^{-1}(u)\mathrm{d}u$, the first two references deal with the more general integral $\int_0^1 F^{-1}(u)w(u)\mathrm{d}u$, which we shall discuss in more detail in Section 9.

Corollary 5.2 (bias). *Suppose that $F \in \mathcal{F}_1^{-}$, and let $F_1, F_2, \ldots \in \mathcal{F}_1^{-}$ be any sequence of cdf's that are unbiased estimators of F, that is, satisfy condition (3.15), and let these cdf's also be such that condition (3.16) is satisfied. Then $\int_0^p F_n^{-1}(u)\mathrm{d}u$ is a non-negatively biased estimator of $\int_0^p F^{-1}(u)\mathrm{d}u$.*

Proof. Using Theorem 5.1 with F_n instead of G, and also recalling that the remainder term

30

$\text{Rem}(p; F, G)$ is non-negative, we have

$$\text{Bias}_n^{\text{LOWER}}(p) := \mathbb{E}\left(\int_0^p F_n^{-1}(u)\mathrm{d}u \right) - \int_0^p F^{-1}(u)\mathrm{d}u$$
$$= \mathbb{E}\big(\text{Rem}(p; F, F_n) \big)$$
$$\in [0, \infty)$$

for every $n \in \mathbb{N}$. Note that $\mathbb{E}\big(\text{Rem}(p; F, F_n) \big)$ is finite because of bound (3.5) and condition (3.16). This establishes Corollary 5.2. $\qquad\square$

Example 5.2 (SRS). When F_n is $F_{n,\text{SRS}}$, condition (3.15) is satisfied. Furthermore, condition (3.16) is satisfied when $F \in \mathcal{T}_\varepsilon$ for some $\varepsilon > 0$ (recall Example 3.2 for details). Hence, we conclude from Corollary 5.2 that when $F \in \mathcal{F}_1^- \cap \mathcal{T}_\varepsilon$, the empirical estimator $\int_0^p F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u$ of the lower-layer integral $\int_0^p F^{-1}(u)\mathrm{d}u$, although being consistent by Example 5.1, has the positive bias

$$\text{Bias}_{n,\text{SRS}}^{\text{LOWER}}(p) := \mathbb{E}\big(\text{Rem}(p; F, F_{n,\text{SRS}}) \big) \in [0, \infty). \tag{5.9}$$

Note 5.4. The condition that $F \in \mathcal{F}_1^- \cap \mathcal{T}_\varepsilon$ for some $\varepsilon > 0$ can equivalently be reformulated as $F \in \mathcal{F}_1^- \cap \mathcal{T}_\varepsilon^+$ for some $\varepsilon > 0$, where $\mathcal{T}_\varepsilon^+$ consists of all those cdf's that satisfy $x^\varepsilon(1 - F(x)) \to 0$ when $x \to \infty$. This reformulation is possible due to the fact that $F \in \mathcal{F}_1^-$ implies $|x|^\varepsilon F(x) \to 0$ when $x \to -\infty$ for every $\varepsilon \in (0, 1)$, and we need this condition to hold for just one $\varepsilon > 0$. For those researchers who think in terms of moments, a sufficient condition that ensures the existence of $\varepsilon > 0$ such that $F \in \mathcal{F}_1^- \cap \mathcal{T}_\varepsilon^+$ would be the requirement that

$$\mathbb{E}(X^-) < \infty \quad \text{and} \quad \mathbb{E}((X^+)^\varepsilon) < \infty \tag{5.10}$$

for some $\varepsilon > 0$, where $X \sim F$. Note that requirement (5.10) can equivalently be rewritten as $F \in \mathcal{F}_1^- \cap \mathcal{F}_\varepsilon^+$.

Note 5.5. Let $F \in \mathcal{F}_1$, where $\mathcal{F}_1 := \mathcal{F}_1^- \cap \mathcal{F}_1^+$. Hence, $F \in \mathcal{F}_1$ is equivalent to saying that $\int_0^1 F^{-1}(u)\mathrm{d}u \in \mathbb{R}$, which is the same as saying that the random variable $X \sim F$ has a finite first moment $\mathbb{E}(X) \in \mathbb{R}$. When in addition to this moment requirement, condition (3.15) is also satisfied, we have the equation

$$\text{Bias}_n^{\text{LOWER}}(p) = -\text{Bias}_n^{\text{UPPER}}(p)$$

because, due to equation (1.6) with F_n instead of G,

$$\text{Bias}_n^{\text{UPPER}}(p) + \text{Bias}_n^{\text{LOWER}}(p) = \mathbb{E}\left(\int_0^1 F_n^{-1}(u)\mathrm{d}u - \int_0^1 F^{-1}(u)\mathrm{d}u \right)$$

$$= \mathbb{E}\left(\int_{-\infty}^{\infty} \big(F(x) - F_n(x)\big)\mathrm{d}x \right)$$

$$= \int_{-\infty}^{\infty} \mathbb{E}\big(F(x) - F_n(x)\big)\mathrm{d}x,$$

where the right-hand side is equal to 0 due to condition (3.15).

Corollary 5.3. *Suppose that $F \in \mathcal{F}_1^-$, and let $F_1, F_2, \ldots \in \mathcal{F}_1^-$ be any sequence of cdf's satisfying condition (3.21) and also condition (3.22) with some normalizing constants $A_n \to \infty$ when $n \to \infty$. Then*

$$A_n \left(\int_0^p F_n^{-1}(u)\mathrm{d}u - \int_0^p F^{-1}(u)\mathrm{d}u \right) = A_n \int_{-\infty}^{F^{-1}(p)} \big(F(x) - F_n(x)\big)\mathrm{d}x + o_{\mathbb{P}}(1) \qquad (5.11)$$

and, therefore, the convergence-in-distribution statement

$$A_n \left(\int_0^p F_n^{-1}(u)\mathrm{d}u - \int_0^p F^{-1}(u)\mathrm{d}u \right) \xrightarrow{d} \mathcal{L}_{\text{LOWER}}(p) \qquad (5.12)$$

holds if and only if

$$A_n \left(\int_{-\infty}^{F^{-1}(p)} F_n(x)\mathrm{d}x - \int_{-\infty}^{F^{-1}(p)} F(x)\mathrm{d}x \right) \xrightarrow{d} -\mathcal{L}_{\text{LOWER}}(p), \qquad (5.13)$$

where $\mathcal{L}_{\text{LOWER}}(p)$ is a random variable determined by statement (5.13).

Proof. Using Theorem 5.1 with F_n instead of G, we only need to show that $A_n\text{Rem}(p; F, F_n)$ converges in probability to 0, which we showed in the proof of Corollary 3.3. This establishes equation (5.11), which implies the equivalence of statements (5.12) and (5.13), and concludes the proof of Corollary 5.3. $\qquad \square$

Example 5.3 (SRS)**.** When F_n is $F_{n,\text{SRS}}$, condition (3.22), which is a requirement in Corollary 5.3, is satisfied by the Kolmogorov-Smirnov theorem. Hence, condition (3.21) is satisfied as well, provided that, according to Lemma 3.1, the quantile function F^{-1} is continuous at the point p. Furthermore, since the integral $\int_{-\infty}^{F^{-1}(p)} F_n(x)\mathrm{d}x$ is the arithmetic mean of n independent copies of random variable (5.6), we conclude that $\int_{-\infty}^{F^{-1}(p)} F_n(x)\mathrm{d}x$ satisfies the

central limit theorem provided that random variable (5.6) has a finite second moment. This is so provided that $F \in \mathcal{F}_2^-$ (see Lemma A.2 for details), where $\mathcal{F}_2^-$ denotes the set of all cdf's such that the integral $\int_0^p \left(F^{-1}(u)\right)^2 \mathrm{d}u$ is finite. Note that $F \in \mathcal{F}_2^-$ is equivalent to saying that the negative part X^- of the random variable $X \sim F$ has a finite second moment $\mathbb{E}((X^-)^2) < \infty$. In summary, when $F \in \mathcal{F}_2^-$ and the quantile function F^{-1} is continuous at the point p, we have the asymptotic normality result

$$\sqrt{n} \left(\int_0^p F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u - \int_0^p F^{-1}(u)\mathrm{d}u \right) \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{LOWER}}^2(p)), \tag{5.14}$$

where $\sigma_{\text{LOWER}}^2(p)$ is the variance of the random variable $Y_{\text{LOWER}}(p)$. The variance can be expressed by the formula

$$\sigma_{\text{LOWER}}^2(p) = \int_{-\infty}^{F^{-1}(p)} \int_{-\infty}^{F^{-1}(p)} \left(F(x \wedge y) - F(x)F(y)\right)\mathrm{d}x\mathrm{d}y. \tag{5.15}$$

Note 5.6 (dependent data). Limiting distributions extending statement (5.14) to classes of dependent random sequences were studied by Davydov and Zitikis (2003), Davydov and Zitikis (2004), and Davydov et al. (2007). The knowledge of empirical processes and their asymptotic behaviour based on such random sequences becomes particulary useful, and we refer to Dehling et al. (2002) for details on the topic. Due to space considerations, we only note here that dependence structures affect asymptotic variances, normalizing constants, and even limiting distributions. For a bird's-eye view of the variety of normalizing constants and limiting distributions in the case of lower-layer integrals (also known as convexifications, absolute Lorenz curves, and by some other names), we refer to Davydov and Zitikis (2004).

Computation 5.1 (SRS). Using equation (3.12) for calculating the quantile $F_{n,\text{SRS}}^{-1}(u)$, we

have

$$\int_0^p F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u = \sum_{i=1}^{\lceil np \rceil} \int_{(i-1)/n}^{i/n} F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u - \int_p^{\lceil np \rceil/n} F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{\lceil np \rceil} X_{i:n}}_{\text{main term}} - \underbrace{\left(\frac{\lceil np \rceil}{n} - p\right) X_{\lceil np \rceil:n}}_{\text{asymptotically negligible term}} \tag{5.16}$$

$$= \frac{\lceil np \rceil}{n} \underbrace{\frac{1}{\lceil np \rceil}\sum_{i=1}^{\lceil np \rceil} X_{i:n}}_{\text{right-trimmed mean}} - \underbrace{\left(\frac{\lceil np \rceil}{n} - p\right) X_{\lceil np \rceil:n}}_{\text{asymptotically negligible term}} \ . \tag{5.17}$$

These expressions give computationally-friendly formulas for the empirical lower-layer integral, which is connected to the right-trimmed mean (e.g., Stigler, 1973) via equation (5.17).

Computation 5.2 (Uniform Vervaat process)**.** Using formulas (5.6) and (5.16) in the case of independent and uniformly on $[0,1]$ distributed random variables $U_1,\ldots,U_n$, we obtain the formulas

$$V_{n,\text{SRS}}^U(p) = \int_0^p E_{n,\text{SRS}}^{-1}(u)\mathrm{d}u + \frac{1}{n}\sum_{i=1}^n (p - U_i)^+ - p^2$$

$$= \frac{1}{n}\sum_{i=1}^{\lceil np \rceil} U_{i:n} - \left(\frac{\lceil np \rceil}{n} - p\right) U_{\lceil np \rceil:n} + \frac{1}{n}\sum_{i=1}^n (p - U_i)^+ - p^2$$

that we used to visualize the normalized Vervaat process

$$n\,V_{n,\text{SRS}}^U(p), \quad 0 \le p \le 1,$$

in Figure 5.1a. The appropriateness of the normalization n for the process V_n^U was established by Vervaat (1972a,b), who showed that $n\,V_{n,\text{SRS}}^U$ converges weakly to a half of the squared Brownian bridge, that is, to $B^2/2$, when $n \to \infty$. On the other hand, the normalized uniform empirical process

$$n^{1/2}(E_{n,\text{SRS}}(p) - p), \quad 0 \le p \le 1,$$

which we have visualized in Figure 5.1b, converges weakly to the Brownian bridge B when $n \to \infty$ (e.g., Billingsley, 1999). Hence, what we essentially see in the two panels of Figure 5.1 are sample paths of a half of the squared Brownian bridge $B^2/2$ (left-hand panel) and of

the classical Brownian bridge B (right-hand panel). For additional insights into these and related results, we refer to Csörgő and Zitikis (1999, 2001), and Csáki et al. (2002).

6 The middle-layer integral

Let $\mathcal{F}$ denote the class of all cdf's. Fix any pair of probability levels p_1 and p_2 such that

$$0 < p_1 < p_2 < 1.$$

The middle-layer integral $\int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u$ is finite for every $F \in \mathcal{F}$. The functional

$$\mathcal{F} \ni F \mapsto \int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u \in \mathbb{R}$$

is not linear, but when developing statistical inference it can be approximated by the linear one

$$\mathcal{F} \ni G \mapsto \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} G(x)\mathrm{d}x \in \mathbb{R}. \tag{6.1}$$

Indeed, we shall see from the following corollaries that the difference between the empirical middle-layer integral $\int_{p_1}^{p_2} F_n^{-1}(u)\mathrm{d}u$ and its population counterpart $\int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u$ gets asymptotically close to the difference between the integrals $\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F(x)\mathrm{d}x$ and $\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F_n(x)\mathrm{d}x$, where F_n, $n \in \mathbb{N}$, are cdf's approaching F when the parameter n grows indefinitely. The corollaries rely on the following theorem, whose proof is given in Appendix A.2.

Theorem 6.1. *Let F and G be any two cdf's. Then*

$$\int_{p_1}^{p_2} \big(G^{-1}(u) - F^{-1}(u)\big)\mathrm{d}u = \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \big(F(x) - G(x)\big)\mathrm{d}x + \mathrm{Rem}(p_2; F, G) - \mathrm{Rem}(p_1; F, G),$$

$$\tag{6.2}$$

where the non-negative remainder terms $\mathrm{Rem}(p_1; F, G)$ and $\mathrm{Rem}(p_2; F, G)$ are defined by equation (3.3) and satisfy bounds (3.4) and (3.5).

Note 6.1. It is tempting to collapse equations (1.6), (3.2), (5.2), and (6.2) into one by relaxing the restriction $0 < p_1 < p_2 < 1$ to $0 \leq p_1 < p_2 \leq 1$. This is indeed possible by augmenting the definition of the remainder term $\mathrm{Rem}(p; F, G)$, which has so far been given only for $p \in (0, 1)$, by setting $\mathrm{Rem}(p; F, G)$ to 0 when $p \in \{0, 1\}$. If we agree with this augmentation, then it is also imperative to replace $F^{-1}(p_1)$ by $-\infty$ when $p_1 = 0$ in the

integral

$$\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \big(F(x) - G(x)\big)\mathrm{d}x \tag{6.3}$$

on the right-hand side of equation (6.2), and to also replace $F^{-1}(p_2)$ by ∞ when $p_2 = 1$. Note that $F^{-1}(0)$ may or may not be equal to $-\infty$, and $F^{-1}(1)$ may or may not be equal to ∞. Hence, although the noted replacements may look artificial, they are absolutely necessary, as next Example 6.1 shows. Consequently, the role of infinities $-\infty$ and ∞ instead of $F^{-1}(p_1)$ and $F^{-1}(p_2)$ in integral (6.3) when $p_1 = 0$ and $p_2 = 1$, respectively, is crucial. In summary, we recommend having four separate equations (1.6), (3.2), (5.2), and (6.2) in the toolbox, instead of having just one general equation with a number of caveats and adjustments – the four separate equations should help to avoid potential overlooks and pitfalls.

Example 6.1. Let $F \sim U_{[0,1]}$ and $G \sim U_{[0,2]}$ be two random variables with uniform on the intervals $[0,1]$ and $[0,2]$ distributions, respectively. That is, $F(x) = x$ when $0 \le x \le 1$ and $G(x) = x/2$ when $0 \le x \le 2$. Consequently, $F^{-1}(t) = t$ and $G^{-1}(t) = 2t$ when $0 \le t \le 1$. From these formulas, we obtain

$$\int_0^1 \big(G^{-1}(t) - F^{-1}(t)\big)\mathrm{d}t = \frac{1}{2},$$
$$\int_{F^{-1}(0)}^{F^{-1}(1)} \big(F(x) - G(x)\big)\mathrm{d}x = \frac{1}{4},$$
$$\int_{-\infty}^{\infty} \big(F(x) - G(x)\big)\mathrm{d}x = \frac{1}{2}.$$

Hence,

$$\int_0^1 \big(G^{-1}(t) - F^{-1}(t)\big)\mathrm{d}t = \int_{-\infty}^{\infty} \big(F(x) - G(x)\big)\mathrm{d}x,$$

which is what equation (1.6) says, but

$$\int_0^1 \big(G^{-1}(t) - F^{-1}(t)\big)\mathrm{d}t \ne \int_{F^{-1}(0)}^{F^{-1}(1)} \big(F(x) - G(x)\big)\mathrm{d}x.$$

This concludes Example 6.1.

We are now ready to formulate and discuss several corollaries to Theorem 6.1.

Corollary 6.1 (consistency). *Let F be any cdf, and let $F_1, F_2, \ldots$ be any sequence of cdf's*

satisfying condition (3.7). *Then*

$$\int_{p_1}^{p_2} F_n^{-1}(u)\mathrm{d}u - \int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u = \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \big(F(x) - F_n(x)\big)\mathrm{d}x + o_{\mathbb{P}}(1) \qquad (6.4)$$

and, therefore, the consistency statement

$$\int_{p_1}^{p_2} F_n^{-1}(u)\mathrm{d}u \overset{\mathbb{P}}{\to} \int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u \qquad (6.5)$$

holds if and only if

$$\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F_n(x)\mathrm{d}x \overset{\mathbb{P}}{\to} \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F(x)\mathrm{d}x. \qquad (6.6)$$

Proof. The corollary follows from Theorem 6.1 and bound (3.5) with F_n instead of G, because for $p \in \{p_1, p_2\} \subset (0,1)$, $F^{-1}(p)$ is finite and $F_n^{-1}(p)$ is asymptotically bounded. Therefore condition (3.7) implies $\mathrm{Rem}(p; F, F_n) = o_{\mathbb{P}}(1)$ for both $p = p_1$ and $p = p_2$, and thus, in turn, implies statement (6.4). This establishes the equivalence of statements (6.5) and (6.6), and concludes the proof of Corollary 6.1. $\qquad \square$

Example 6.2 (SRS)**.** Condition (3.7) with $F_{n,\mathrm{SRS}}$ in place of F_n is satisfied by the Glivenko-Cantelli theorem. Furthermore, by linearity of functional (6.1), the integral $\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F_n(x)\mathrm{d}x$ is the arithmetic mean $n^{-1}\sum_{i=1}^{n} Y_{i,\mathrm{MIDDLE}}(p_1, p_2)$ of n independent copies of the random variable

$$\begin{aligned}
Y_{\mathrm{MIDDLE}}(p_1, p_2) &= \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \mathbb{1}\{X \le x\}\mathrm{d}x \\
&= \big(F^{-1}(p_2) - X\big)^+ - \big(F^{-1}(p_1) - X\big)^+.
\end{aligned} \qquad (6.7)$$

The random variable $Y_{\mathrm{MIDDLE}}(p_1, p_2)$ always has a finite first moment (see Note A.1 for details). Consequently, $\int_{p_1}^{p_2} F_{n,\mathrm{SRS}}^{-1}(u)\mathrm{d}u$ is a consistent estimator of $\int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u$, that is, for every cdf F, we have

$$\int_{p_1}^{p_2} F_{n,\mathrm{SRS}}^{-1}(u)\mathrm{d}u \overset{\mathbb{P}}{\to} \int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u. \qquad (6.8)$$

Corollary 6.2 (bias)**.** *Let F be any cdf, and let $F_1, F_2, \ldots$ be any sequence of cdf's that are unbiased estimators of F, that is, satisfy condition* (3.15)*, and let the cdf's be such that*

condition (3.16) *is satisfied with* $p \in \{p_1, p_2\}$. *Then*

$$\text{Bias}_n^{\text{MIDDLE}}(p_1, p_2) := \mathbb{E}\left(\int_{p_1}^{p_2} F_n^{-1}(u) \mathrm{d}u\right) - \int_{p_1}^{p_2} F^{-1}(u) \mathrm{d}u$$

$$= \mathbb{E}\big(\text{Rem}(p_2; F, F_n)\big) - \mathbb{E}\big(\text{Rem}(p_1; F, F_n)\big)$$

$$\in \mathbb{R}.$$

Proof. Theorem 6.1 with F_n instead of G implies Corollary 6.2. □

Example 6.3 (SRS)**.** When F_n is $F_{n,\text{SRS}}$, condition (3.15) is satisfied. Furthermore, condition (3.16) is satisfied when $F \in \mathcal{T}_\varepsilon$ for some $\varepsilon > 0$ (recall Example 3.2 for details). Hence, we conclude from Corollary 6.2 that when $F \in \mathcal{T}_\varepsilon$, the estimator $\int_{p_1}^{p_2} F_{n,\text{SRS}}^{-1}(u)\mathrm{d}u$ of the middle-layer integral $\int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u$, although being consistent by Example 6.2, has the bias

$$\text{Bias}_{n,\text{SRS}}^{\text{MIDDLE}}(p_1, p_2) = \mathbb{E}\big(\text{Rem}(p_2; F, F_{n,\text{SRS}})\big) - \mathbb{E}\big(\text{Rem}(p_1; F, F_{n,\text{SRS}})\big) \in \mathbb{R},$$

which might be any real number. Note that a sufficient condition for $F \in \mathcal{T}_\varepsilon$ for some $\varepsilon > 0$ is the requirement $F \in \mathcal{F}_\varepsilon$, that is, $\mathbb{E}(|X|^\varepsilon) < \infty$ for some $\varepsilon > 0$.

Remark 6.1. Naturally, we can get more insights into the bias if we are willing to make additional assumptions about F. For example, following the studies of Gribkova and Helmers (2006, 2007) on trimmed means, if we assume that F has a pdf $f = F'$ that is strictly positive and continuous at the quantiles $F^{-1}(p_1)$ and $F^{-1}(p_2)$, then

$$\text{Bias}_{n,\text{SRS}}^{\text{MIDDLE}}(p_1, p_2) = \underbrace{\frac{p_2(1 - p_2)}{2nf(F^{-1}(p_2))}}_{\text{vanishes when } p_2 = 1} - \underbrace{\frac{p_1(1 - p_1)}{2nf(F^{-1}(p_1))}}_{\text{vanishes when } p_1 = 0} + o\left(\frac{1}{n}\right). \tag{6.9}$$

This asymptotic formula helps us to appreciate the fact that in Example 6.3 we were not able to say anything definitive about the sign of the bias $\text{Bias}_{n,\text{SRS}}^{\text{MIDDLE}}(p_1, p_2)$, and it also helps us to see why we could earlier show that $\text{Bias}_{n,\text{SRS}}^{\text{UPPER}}(p) \leq 0$ (statement (3.19)) and $\text{Bias}_{n,\text{SRS}}^{\text{LOWER}}(p) \geq 0$ (statement (5.9)) for any $p \in (0, 1)$. Furthermore, asymptotic formula (6.9) helps us to understand why the Vervaat process needs to be normalized by n to stabilize its asymptotic behaviour and hence to obtain a non-degenerate limit, which is a half of the squared Brownian bridge, that is, $B^2(p)/2$, $0 \leq p \leq 1$. Note in this regard that the expected value of $B^2(p)/2$ is equal to $p(1-p)/2$, thus serving a further clarification of the form of the two leading terms on the right-hand side of equation (6.9).

Corollary 6.3 (asymptotic distribution). *Let F be any cdf, and let $F_1, F_2, \ldots$ be any sequence of cdf's satisfying condition* (3.21) *with $p \in \{p_1, p_2\}$ and condition* (3.22) *with some normalizing constants $A_n \to \infty$ when $n \to \infty$. Then*

$$A_n \left(\int_{p_1}^{p_2} F_n^{-1}(u) \mathrm{d}u - \int_{p_1}^{p_2} F^{-1}(u) \mathrm{d}u \right) = A_n \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \left(F(x) - F_n(x) \right) \mathrm{d}x + o_{\mathbb{P}}(1) \quad (6.10)$$

and, therefore, the convergence-in-distribution statement

$$A_n \left(\int_{p_1}^{p_2} F_n^{-1}(u) \mathrm{d}u - \int_{p_1}^{p_2} F^{-1}(u) \mathrm{d}u \right) \overset{d}{\to} \mathcal{L}_{\mathrm{MIDDLE}}(p_1, p_2) \quad (6.11)$$

holds if and only if

$$A_n \left(\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F_n(x) \mathrm{d}x - \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F(x) \mathrm{d}x \right) \overset{d}{\to} -\mathcal{L}_{\mathrm{MIDDLE}}(p_1, p_2), \quad (6.12)$$

where $\mathcal{L}_{\mathrm{MIDDLE}}(p_1, p_2)$ is a random variable determined by statement (6.12).

Proof. Using Theorem 6.1 with F_n instead of G, we only need to show that $A_n \mathrm{Rem}(p; F, F_n)$ converges in probability to 0, which we showed in the proof of Corollary 3.3. This establishes equation (6.10), which implies the equivalence of statements (6.11) and (6.12), and concludes the proof of Corollary 6.3. $\qquad \square$

Example 6.4 (SRS). When F_n is $F_{n,\mathrm{SRS}}$, condition (3.22) is a consequence of the Kolmogorov-Smirnov theorem. Since condition (3.22) is satisfied, condition (3.21) is satisfied as well, provided that, according to Lemma 3.1, the quantile function F^{-1} is continuous at the points p_1 and p_2. Furthermore, since the integral $\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F_n(x) \mathrm{d}x$ is the arithmetic mean of n independent copies of random variable (6.7), we conclude that the integral $\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F_n(x) \mathrm{d}x$ satisfies the central limit theorem. In summary, therefore, when the quantile function F^{-1} is continuous at the points p_1 and p_2, we have

$$\sqrt{n} \left(\int_{p_1}^{p_2} F_{n,\mathrm{SRS}}^{-1}(u) \mathrm{d}u - \int_{p_1}^{p_2} F^{-1}(u) \mathrm{d}u \right) \overset{d}{\to} \mathcal{N}(0, \sigma_{\mathrm{MIDDLE}}^2(p_1, p_2)), \quad (6.13)$$

where $\sigma_{\mathrm{MIDDLE}}^2(p_1, p_2)$ is the variance of the random variable $Y_{\mathrm{MIDDLE}}(p_1, p_2)$. The variance can

be expressed by the formula (see Note A.1)

$$\sigma^2_{\text{MIDDLE}}(p_1, p_2) = \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \big(F(x \wedge y) - F(x)F(y)\big)\mathrm{d}x\mathrm{d}y. \qquad (6.14)$$

Computation 6.1 (SRS)**.** To obtain a convenient computational formula for the empirical middle-layer integral $\int_{p_1}^{p_2} F^{-1}_{n,\text{SRS}}(u)\mathrm{d}u$, we can start with the equation

$$\int_{p_1}^{p_2} F^{-1}_{n,\text{SRS}}(u)\mathrm{d}u = \int_0^{p_2} F^{-1}_{n,\text{SRS}}(u)\mathrm{d}u - \int_0^{p_1} F^{-1}_{n,\text{SRS}}(u)\mathrm{d}u \qquad (6.15)$$

and then apply, for example, computational formula (5.17) for the two empirical lower-layer integrals on the right-hand side of equation (6.15). (All empirical integrals are finite, and so manipulations like those in equation (6.15) do not pose any technical issues.) In particular, we obtain the following equations

$$\int_{p_1}^{p_2} F^{-1}_{n,\text{SRS}}(u)\mathrm{d}u = \frac{1}{n} \sum_{i=\lceil np_1 \rceil + 1}^{\lceil np_2 \rceil} X_{i:n} - \left(\frac{\lceil np_2 \rceil}{n} - p_2\right) X_{\lceil np_2 \rceil:n} + \left(\frac{\lceil np_1 \rceil}{n} - p_1\right) X_{\lceil np_1 \rceil:n}$$

$$= \frac{\lceil np_2 \rceil - \lceil np_1 \rceil}{n} \underbrace{\frac{1}{\lceil np_2 \rceil - \lceil np_1 \rceil} \sum_{i=\lceil np_1 \rceil + 1}^{\lceil np_2 \rceil} X_{i:n}}_{\text{trimmed mean}}$$

$$\underbrace{- \left(\frac{\lceil np_2 \rceil}{n} - p_2\right) X_{\lceil np_2 \rceil:n} + \left(\frac{\lceil np_1 \rceil}{n} - p_1\right) X_{\lceil np_1 \rceil:n}}_{\text{asymptotically negligible terms}}$$

with the right-hand side connecting the empirical middle-layer integral with the trimmed (on both sides) mean that was considered by Stigler (1973) and played a pivotal role in the development of robust statistics, including the introduction of the method of trimmed moments (MTM) by Brazauskas et al. (2009) and the method of Winsorized moments (MWM) by Zhao et al. (2018).

7 Illustrations: iid random variables

To make the following illustrations less voluminous and easier connectable to what is known in the literature and hence maximally illuminating, we use SRS throughout this section. That is, we work with iid random variables $X_1, \ldots, X_n$ and from them arising empirical cdf

$F_{n,\text{SRS}}$, which is a consistent and unbiased estimator of the underlying population cdf F.

7.1 Upside tail-value-at-risk

Given any $p \in (0, 1)$ and $F \in \mathcal{F}_1^+$, the upside tail-value-at-risk $\text{TVaR}(p)$ is given by equation (1.1). As an illustration, in Figure 1.1a we depicted $p \mapsto \text{TVaR}(p)$ for the Pareto (Type I) distribution, whose quantile function is $p \mapsto x_0(1-p)^{-1/\alpha}$ with the scale $x_0 > 0$ and shape $\alpha > 0$ parameters, which we set to 1 and 3, respectively. The empirical counterpart $\text{TVaR}_{n,\text{SRS}}(p)$ is defined by equation (1.1) with $F_{n,\text{SRS}}$ instead of F. (For a comprehensive review of various estimators of TVaR and its sister risk measure called Expected Shortfall, we refer to Nadarajah et al. (2013).) We have the following three statements concerning consistency, bias, and asymptotic normality of $\text{TVaR}_{n,\text{SRS}}(p)$.

7.1.1 Consistency

When $F \in \mathcal{F}_1^+$, we obtain from Example 3.1 that $\text{TVaR}_{n,\text{SRS}}(p)$ is a consistent estimator of $\text{TVaR}(p)$, that is,

$$\text{TVaR}_{n,\text{SRS}}(p) \xrightarrow{\mathbb{P}} \text{TVaR}(p). \tag{7.1}$$

This result is a special case of Gribkova et al. (2022a) who consider consistency and with it associated fixed-margin confidence intervals for the tail conditional allocation, which generalizes $\text{TVaR}(p)$. For related bootstrap-type considerations and fixed-margin confidence intervals, we refer to Gribkova et al. (2024).

7.1.2 Bias

When $F \in \mathcal{F}_1^+ \cap \mathcal{T}_\varepsilon$ for some $\varepsilon > 0$ (recall Note 3.5), we obtain from Example 3.2 that $\text{TVaR}_{n,\text{SRS}}(p)$ is a non-positively biased estimator of $\text{TVaR}(p)$, that is,

$$\mathbb{E}\big(\text{TVaR}_{n,\text{SRS}}(p)\big) \leq \text{TVaR}(p) \tag{7.2}$$

for every $n \in \mathbb{N}$. This property, although not stated explicitly, was earlier established by Brazauskas et al. (2008, p. 3602). Garca-Risueño (2025) offers an illuminating discussion of the property from the heuristic point of view.

7.1.3 Asymptotic normality

When $F \in \mathcal{F}_2^+$ and the quantile function F^{-1} is continuous at the point p, we obtain from Example 3.3 that

$$\sqrt{n}\left(\mathrm{TVaR}_{n,\mathrm{SRS}}(p) - \mathrm{TVaR}(p)\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathrm{TVaR}}(p)), \tag{7.3}$$

where the asymptotic variance is

$$\sigma^2_{\mathrm{TVaR}}(p) = \frac{1}{(1-p)^2} \int_{F^{-1}(p)}^{\infty} \int_{F^{-1}(p)}^{\infty} \left(F(x \wedge y) - F(x)F(y)\right) \mathrm{d}x \mathrm{d}y. \tag{7.4}$$

This statement is a special case of a more general result derived by Brazauskas et al. (2008), who consider not just pointwise convergence to the limiting distribution but also uniform convergence over all $p \in (0,1)$, which allows one to establish confidence bands for the function $p \mapsto \mathrm{TVaR}(p)$. Statement (7.3) is also a special case of Gribkova et al. (2022b) who consider the limiting distribution of the tail conditional allocation, which generalizes $\mathrm{TVaR}(p)$.

7.2 Downside tail-value-at-risk

Given $p \in (0,1)$ and $F \in \mathcal{F}_1^-$, the downside tail-value-at-risk $\mathrm{TV@R}(p)$ is given by equation (1.2). As an illustration, in Figure 1.1b we depicted $p \mapsto \mathrm{TV@R}(p)$ for the Pareto (Type I) distribution with the scale $x_0 = 1$ and shape $\alpha = 3$ parameters. The empirical counterpart $\mathrm{TV@R}_{n,\mathrm{SRS}}(p)$ is defined by equation (1.2) with $F_{n,\mathrm{SRS}}$ instead of F. We have the following three statements concerning consistency, bias, and asymptotic normality of $\mathrm{TV@R}_{n,\mathrm{SRS}}(p)$.

7.2.1 Consistency

When $F \in \mathcal{F}_1^-$, we obtain from Example 5.1 that $\mathrm{TV@R}_{n,\mathrm{SRS}}(p)$ is a consistent estimator of $\mathrm{TV@R}(p)$, that is,

$$\mathrm{TV@R}_{n,\mathrm{SRS}}(p) \xrightarrow{\mathbb{P}} \mathrm{TV@R}(p). \tag{7.5}$$

7.2.2 Bias

When $F \in \mathcal{F}_1^- \cap \mathcal{T}_\varepsilon$ for some $\varepsilon > 0$ (recall Note 5.4), we obtain from Example 5.2 that $\mathrm{TV@R}_{n,\mathrm{SRS}}(p)$ is a non-negatively biased estimator of $\mathrm{TV@R}(p)$, that is,

$$\mathbb{E}\left(\mathrm{TV@R}_{n,\mathrm{SRS}}(p)\right) \geq \mathrm{TV@R}(p) \tag{7.6}$$

for every $n \in \mathbb{N}$.

7.2.3 Asymptotic normality

When $F \in \mathcal{F}_2^-$ and the quantile function F^{-1} is continuous at the point p, we obtain from Example 5.3 that

$$\sqrt{n}\left(\text{TV@R}_{n,\text{SRS}}(p) - \text{TV@R}(p)\right) \xrightarrow{d} \mathcal{N}(0, \sigma_{\text{TV@R}}^2(p)), \tag{7.7}$$

where the asymptotic variance is

$$\sigma_{\text{TV@R}}^2(p) = \frac{1}{p^2} \int_{-\infty}^{F^{-1}(p)} \int_{-\infty}^{F^{-1}(p)} \left(F(x \wedge y) - F(x)F(y)\right) \mathrm{d}x \mathrm{d}y. \tag{7.8}$$

7.3 Lorenz curve

Let $F \in \mathcal{F}_1^{\mu \neq 0}$, where $\mathcal{F}_1^{\mu \neq 0}$ denotes the set of all cdf's that have finite but non-zero first moments, that is, $\mu = \mathbb{E}(X) \in \mathbb{R} \setminus \{0\}$ for $X \sim F$. Following Gastwirth (1971), the Lorenz curve $\text{LC}(p)$ is defined by equation (1.3) with p running through the unit interval $[0, 1]$. (The Lorenz curve is usually defined and meaningfully interpreted only on the set of non-negative random variables $X \geq 0$.) Since the cases $p = 0$ and $p = 1$ are trivial, we restrict ourselves to $p \in (0, 1)$. As an illustration, in Figure 1.1c we depicted $p \mapsto \text{LC}(p)$ in the case of the Pareto (Type I) distribution with the scale $x_0 = 1$ and shape $\alpha = 3$ parameters. When F is replaced by $F_{n,\text{SRS}}$ on the right-hand side of equation (1.3), we obtain the empirical Lorenz curve $\text{LC}_{n,\text{SRS}}(p)$. We have the following three statements concerning consistency, bias, and asymptotic normality of $\text{LC}_{n,\text{SRS}}(p)$ for any fixed $p \in (0, 1)$.

7.3.1 Consistency

When $F \in \mathcal{F}_1^{\mu \neq 0}$, we obtain from Example 5.1 and consistency of the arithmetic mean $\mu_{n,\text{SRS}}$ that $\text{LC}_{n,\text{SRS}}(p)$ is a consistent estimator of $\text{LC}(p)$, that is,

$$\text{LC}_{n,\text{SRS}}(p) \xrightarrow{\mathbb{P}} \text{LC}(p). \tag{7.9}$$

7.3.2 Bias

The result of Example 5.2 suggests that $\text{LC}_{n,\text{SRS}}(p)$ might be a positively biased estimator of $\text{LC}(p)$, given that the arithmetic mean $\mu_{n,\text{SRS}}$ is an unbiased estimator of μ. This may

or may not be true as demonstrated by Arnold and Villaseñor (2015), who summarize their findings on the subject as follows:

> For several parent distributions it is possible to prove that the sample Lorenz curve is a positively biased estimate of the population Lorenz curve. In this paper, several sufficient conditions for such positive bias are investigated. An example shows that negative bias is not impossible, though apparently not common. (Arnold and Villaseñor, 2015, p. 3)

7.3.3 Asymptotic normality

When $F \in \mathcal{F}_2^{\mu \neq 0}$ and the quantile function F^{-1} is continuous at the point p, we obtain from results of Section 5 and equation (1.6) that (see Lemma A.3 in Appendix A.3 for a proof)

$$\sqrt{n}\left(\mathrm{LC}_{n,\mathrm{SRS}}(p) - \mathrm{LC}(p)\right) \xrightarrow{d} \mathcal{N}(0, \sigma_{\mathrm{LC}}^2(p)), \tag{7.10}$$

where the asymptotic variance $\sigma_{\mathrm{LC}}^2(p)$ is the second moment of the mean-zero random variable

$$Y_{\mathrm{LC}}(p) = \frac{1}{\mu}\int_{-\infty}^{F^{-1}(p)}\left(\mathbb{1}\{X \leq x\} - F(x)\right)\mathrm{d}x + \frac{\mathrm{LC}(p)}{\mu}(X - \mu). \tag{7.11}$$

For confidence bands (i.e., simultaneous confidence intervals over all $p \in (0,1)$) for the Lorenz curve under minimal conditions on the population distribution, we refer to Csörgő et al. (1998). It is also useful to note that the variance $\sigma_{\mathrm{LC}}^2(p)$ can be expressed in a form resembling those for $\sigma_{\mathrm{TVaR}}^2(p)$ and $\sigma_{\mathrm{TV@R}}^2(p)$ in equations (7.4) and (7.8), respectively, but apart from being nice theoretical exercises, we do not see much practical value in doing so: first, this would not lead to a convenient empirical estimator for $\sigma_{\mathrm{LC}}^2(p)$, and second, the variance $\sigma_{\mathrm{LC}}^2(p)$ can be easier estimated via a resampling technique (e.g., Shao and Tu, 1995).

7.4 Gini curve

Let $F \in \mathcal{F}_1^{\mu \neq 0}$, where the class $\mathcal{F}_1^{\mu \neq 0}$ is defined in Example 7.3. The Gini curve $\mathrm{GC}(p)$ is given by equation (1.4), where p runs through the unit interval $[0,1]$. (The Gini curve is usually defined and meaningfully interpreted only on the set of non-negative random variables $X \geq 0$.) Since the cases $p = 0$ and $p = 1$ are trivial, we restrict ourselves to $p \in (0,1)$. As an illustration, in Figure 1.1d we depicted $p \mapsto \mathrm{GC}(p)$ in the case of the Pareto (Type I) distribution with the scale $x_0 = 1$ and shape $\alpha = 3$ parameters. When F is replaced by $F_{n,\mathrm{SRS}}$

on the right-hand side of equation (1.4), we obtain the empirical Gini curve $\mathrm{GC}_{n,\mathrm{SRS}}(p)$. We have the following three statements concerning consistency, bias, and asymptotic normality of $\mathrm{GC}_{n,\mathrm{SRS}}(p)$ for any $p \in (0, 1)$.

7.4.1 Consistency

When $F \in \mathcal{F}_1^{\mu \neq 0}$, we obtain from consistency of the arithmetic mean $\mu_{n,\mathrm{SRS}}$ and Examples 3.1 and 5.1 that $\mathrm{GC}_{n,\mathrm{SRS}}(p)$ is a consistent estimator of $\mathrm{GC}(p)$, that is,

$$\mathrm{GC}_{n,\mathrm{SRS}}(p) \xrightarrow{\mathbb{P}} \mathrm{GC}(p). \tag{7.12}$$

7.4.2 Bias

Even more so than in the case of the Lorenz curve, we cannot say anything definitive about the bias of $\mathrm{GC}_{n,\mathrm{SRS}}(p)$.

7.4.3 Asymptotic normality

When $F \in \mathcal{F}_2^{\mu \neq 0}$ and the quantile function F^{-1} is continuous at the points p and $1 - p$, we obtain from results of Section 5 and equation (1.6) that (see Lemma A.4 in Appendix A.3 for a proof)

$$\sqrt{n} \left(\mathrm{GC}_{n,\mathrm{SRS}}(p) - \mathrm{GC}(p) \right) \xrightarrow{d} \mathcal{N}(0, \sigma_{\mathrm{GC}}^2(p)), \tag{7.13}$$

where the asymptotic variance $\sigma_{\mathrm{GC}}^2(p)$ is the second moment of the mean-zero random variable

$$
\begin{aligned}
Y_{\mathrm{GC}}(p) = \frac{1}{\mu} \int_{-\infty}^{F^{-1}(1-p)} & \left(\mathbb{1}\{X \leq x\} - F(x) \right) \mathrm{d}x \\
& + \frac{1}{\mu} \int_{-\infty}^{F^{-1}(p)} \left(\mathbb{1}\{X \leq x\} - F(x) \right) \mathrm{d}x + \frac{1 - \mathrm{GC}(p)}{\mu} \left(X - \mu \right). \quad (7.14)
\end{aligned}
$$

Estimating the asymptotic variance using a resampling technique might be the most efficient and speediest way toward, e.g., constructing large-sample confidence intervals for $\mathrm{GC}(p)$.

8   Illustrations: stationary time series

Often in applications, data arrive in the form of time series. Depending on the class of time series, and there are many of them, we may see different normalizing constants, different

asymptotic distributions, and even in the case of normal asymptotic distributions, we may see different asymptotic variances. Therefore, to illustrate how our general results work on time series, we need to make a choices, and our's is to work with those time series that are S- and M-mixing. These two notions of mixing have been introduced by Berkes et al. (2009) and Berkes et al. (2011), respectively, and they cover many time series (linear and non-linear). Very importantly, it has also turned out that verifying S- and M-mixing conditions is often easier than verifying classical mixing conditions. Berkes et al. (2009, 2011) provide illuminating discussions of these matters with accompanying examples and references.

8.1 S-mixing and statements (3.23), (5.11) and (6.10)

We follow Berkes et al. (2009) and say that a time series $(X_t)_{t \in \mathbb{Z}}$ is S-mixing if it satisfies the following two conditions:

(A) For any $t \in \mathbb{Z}$ and $m \in \mathbb{N}$, there is a random variable $\Upsilon_t^{(m)}$ such that the property

$$\mathbb{P}\big(|X_t - \Upsilon_t^{(m)}| \geq \gamma_m\big) \leq \delta_m \qquad (8.1)$$

holds for some deterministic sequences $\gamma_m \to 0$ and $\delta_m \to 0$.

(B) For any disjoint intervals $I_1, \ldots, I_r \subset \mathbb{Z}$ of integers and for any positive integers $m_1, \ldots, m_r \in \mathbb{N}$, the vectors $(\Upsilon_t^{(m_1)}, t \in I_1), \ldots, (\Upsilon_t^{(m_r)}, t \in I_r)$ are independent, provided that the separation between the pairs I_i and I_j is greater than $m_i + m_j$ for all $1 \leq i < j \leq r$, that is,

$$\mathrm{dist}(I_i, I_j) := \inf\big\{|a - b|, a \in I_i, b \in I_j\big\}$$
$$> m_i + m_j. \qquad (8.2)$$

We can clearly see why S-mixing has turned out to be such an attractive notion, in particular from the applications point of view: it is based on various portions of the time series $(X_t)_{t \in \mathbb{Z}}$ and not on mathematical constructs such as σ-algebras, as is the case with many classical mixing notions. The only challenge with S-mixing is that one needs to construct random variables $\Upsilon_t^{(m)}$, but Berkes et al. (2009) suggest several recipes for accomplishing this task. To gain intuition on the matter, we next follow one of the recipes and show that the causal autoregressive of order 1 time series is S-mixing.

Example 8.1 (AR(1) is S-mixing). Let $(X_t)_{t\in\mathbb{Z}}$ be a causal AR(1) time series, that is, $X_t = \varphi X_{t-1} + \varepsilon_t$ for all $t \in \mathbb{Z}$, where $|\varphi| < 1$ is a constant and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a mean-zero white noise with finite marginal variances $\sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_t) < \infty$. In addition, we assume that ε_t's are iid random variables, thus making $(X_t)_{t\in\mathbb{Z}}$ strictly stationary. The representation

$$X_t = \sum_{i=0}^{\infty} \varphi^i \varepsilon_{t-i} \tag{8.3}$$

holds for all $t \in \mathbb{Z}$. Denote

$$\Upsilon_t^{(m)} = \sum_{i=0}^{m} \varphi^i \varepsilon_{t-i}. \tag{8.4}$$

Condition (B) of S-mixing is satisfied because the two random vectors

$$(\varepsilon_{s-m_i}, \dots, \varepsilon_s), \quad s \in I_i,$$
$$(\varepsilon_{t-m_j}, \dots, \varepsilon_t), \quad t \in I_j,$$

do not overlap and are therefore independent whenever $\mathrm{dist}(I_i, I_j) > \max\{m_i, m_j\}$, and so definitely when $\mathrm{dist}(I_i, I_j) > m_i + m_j$. To check condition (A), we write

$$\mathbb{P}\big(|X_t - \Upsilon_t^{(m)}| \geq \gamma_m\big) \leq \frac{1}{\gamma_m^2} \mathbb{E}\Bigg(\bigg(\sum_{i=m+1}^{\infty} \varphi^i \varepsilon_{t-i}\bigg)^2\Bigg)$$
$$= \frac{\varphi^{2(m+1)}}{\gamma_m^2} \frac{\sigma_\varepsilon^2}{1-\varphi^2}$$
$$=: \delta_m,$$

where the bound is due to Markov's inequality. For example, if set $\gamma_m = m^{-a}$ for any constant $a > 0$, then we have $\delta_m = O(m^{-A})$ when $m \to \infty$ for any constant $A > 0$. In fact, for γ_m given above, the decay of δ_m is exponential.

The following theorem is a very special case of Berkes et al. (2009, Theorem 2, p. 1303), but it is exactly what we currently need.

Theorem 8.1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a strictly stationary and S-mixing time series whose marginal cdf F is Lipschitz continuous of order $\theta > 0$, that is, there is a constant $c > 0$ such that*

$$\big|F(x) - F(y)\big| \leq c|x - y|^\theta \tag{8.5}$$

for all $x, y \in \mathbb{R}$. Furthermore, assume that condition (A) is satisfied with

$$\gamma_m = \frac{1}{m^{A/\theta}} \quad and \quad \delta_m = O\left(\frac{1}{m^A}\right) \tag{8.6}$$

for some $A > 4$. Then

$$\sqrt{n} \sup_{x \in \mathbb{R}} \left| F_{n,\mathrm{TS}}(x) - F(x) \right| = O_{\mathbb{P}}(1) \tag{8.7}$$

when $n \to \infty$, where $F_{n,\mathrm{TS}}$ is the empirical cdf based on the observable portion $X_1, \ldots, X_n$ of the time series $(X_t)_{t \in \mathbb{Z}}$.

Note 8.1. If the marginal cdf F of $(X_t)_{t \in \mathbb{Z}}$ has a bounded pdf, then condition (8.5) is satisfied with $\theta = 1$, and so $\gamma_m = m^{-A}$ in this case.

We can now formulate the following corollaries to Theorem 8.1.

Corollary 8.1. *Let all the conditions of Theorem 8.1 be satisfied. If $F \in \mathcal{F}_1^+$ and the quantile function F^{-1} is continuous at the point p, then statement (3.23) holds, that is,*

$$\sqrt{n} \left(\int_p^1 F_{n,\mathrm{TS}}^{-1}(u)\mathrm{d}u - \int_p^1 F^{-1}(u)\mathrm{d}u \right)$$
$$= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n h_{\mathrm{UPPER}}(X_i) - \mathbb{E}\left(h_{\mathrm{UPPER}}(X) \right) \right) + o_{\mathbb{P}}(1), \quad (8.8)$$

where

$$h_{\mathrm{UPPER}}(x) = \left(x - F^{-1}(p) \right)^+.$$

The moment condition $F \in \mathcal{F}_1^+$ is sufficient to have $o_{\mathbb{P}}(1)$ in asymptotic equation (8.8), but a stronger moment condition is surely needed to enable the normalized sum on the right-hand side of equation (8.8) to converge to a non-degenerate limit. An analogous note applies to the next corollary.

Corollary 8.2. *Let all the conditions of Theorem 8.1 be satisfied. If $F \in \mathcal{F}_1^-$ and the quantile function F^{-1} is continuous at the point p, then statement (5.11) holds, that is,*

$$\sqrt{n} \left(\int_0^p F_{n,\mathrm{TS}}^{-1}(u)\mathrm{d}u - \int_0^p F^{-1}(u)\mathrm{d}u \right)$$
$$= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n h_{\mathrm{LOWER}}(X_i) - \mathbb{E}\left(h_{\mathrm{LOWER}}(X) \right) \right) + o_{\mathbb{P}}(1), \quad (8.9)$$

where

$$h_{\text{LOWER}}(x) = \left(F^{-1}(p) - x \right)^+.$$

Our final corollary concerns with the middle-layer integral.

Corollary 8.3. *Let all the conditions of Theorem 8.1 be satisfied. If the quantile function F^{-1} is continuous at the points p_1 and p_2, then statement (6.10) holds, that is,*

$$\sqrt{n} \left(\int_{p_1}^{p_2} F_{n,\text{TS}}^{-1}(u)\mathrm{d}u - \int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u \right)$$
$$= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} h_{\text{MIDDLE}}(X_i) - \mathbb{E}\left(h_{\text{MIDDLE}}(X) \right) \right) + o_{\mathbb{P}}(1), \quad (8.10)$$

where

$$h_{\text{MIDDLE}}(X) = \left(F^{-1}(p_2) - x \right)^+ - \left(F^{-1}(p_1) - x \right)^+.$$

Corollary 8.3 does not require any moment assumption because the random variables X_i are transformed by the function h_{MIDDLE} that is bounded. Incidentally, note that the three functions h_{UPPER}, h_{LOWER} and h_{MIDDLE} are Lipschitz continuous of order 1, which will be an important property in next Section 8.2.

8.2 M-mixing and statements (3.24), (5.12) and (6.11)

In view of the results of the previous section, in order to establish asymptotic distributions of the centered and $\sqrt{n}$-normalized integrals of quantiles, i.e., to establish statements (3.24), (5.12) and (6.11), we are left to verify the validity of statements (3.25), (5.13) and (6.12). That is, in view of Corollaries 8.1–8.3, we need to show

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} h_{\text{UPPER}}(X_i) - \mathbb{E}\left(h_{\text{UPPER}}(X) \right) \right) \xrightarrow{d} \mathcal{L}_{\text{UPPER}}(p), \quad (8.11)$$

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} h_{\text{LOWER}}(X_i) - \mathbb{E}\left(h_{\text{LOWER}}(X) \right) \right) \xrightarrow{d} -\mathcal{L}_{\text{LOWER}}(p), \quad (8.12)$$

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} h_{\text{MIDDLE}}(X_i) - \mathbb{E}\left(h_{\text{MIDDLE}}(X) \right) \right) \xrightarrow{d} -\mathcal{L}_{\text{MIDDLE}}(p_1, p_2), \quad (8.13)$$

where the functions h_{UPPER}, h_{LOWER} and h_{MIDDLE} are defined in Corollaries 8.1–8.3. Obviously, the random variables in the sums need to have at least finite second moments, and the forms

of the functions h_{UPPER} and h_{LOWER} translate this requirement into the conditions $F \in \mathcal{F}_2^+$ and $F \in \mathcal{F}_2^-$, respectively. Hence, unlike in Section 8.1 where we worked with indicators and thus employed the probability-based notion of S-mixing, we now need to work with a moment-based notion of mixing. This leads us to the notion of M-mixing introduced by Berkes et al. (2011).

Namely, following Berkes et al. (2011), we say that a time series $(X_t)_{t \in \mathbb{Z}}$ is M_p-mixing for some $p \geq 1$ if it satisfies the following two conditions:

(C) For any $t \in \mathbb{Z}$ and $m \in \mathbb{N}$, there is a random variable $\Psi_t^{(m)}$ such that the bound

$$\left(\mathbb{E}\left(|X_t - \Psi_t^{(m)}|^p \right) \right)^{1/p} \leq \varrho_m \tag{8.14}$$

holds for some deterministic sequence $\varrho_m \to 0$.

(D) For any disjoint intervals $I_1, \ldots, I_r \subset \mathbb{Z}$ of integers and for any positive integers $m_1, \ldots, m_r \in \mathbb{N}$, the vectors $(\Psi_t^{(m_1)}, t \in I_1), \ldots, (\Psi_t^{(m_r)}, t \in I_r)$ are independent provided that

$$\text{dist}(I_i, I_j) > \max\{m_i, m_j\}, \tag{8.15}$$

where the definition of $\text{dist}(I_i, I_j)$ is given in condition (B) of S-mixing.

Note that condition (C) implicitly requires X_t to have a finite p^{th} moment, that is, the cdf of X_t must be in the class $\mathcal{F}_p = \mathcal{F}_p^- \cap \mathcal{F}_p^+$. As to the random variable $\Psi_t^{(m)}$ postulated in the condition, Berkes et al. (2011) offer several recipes for constructing it. The following example illustrates the notion and how to verify it.

Example 8.2 (AR(1) is M_p-mixing). Let $(X_t)_{t \in \mathbb{Z}}$ be the same AR(1) time series as in Example 8.1. In addition, assume that ε_t's have (identical) finite p^{th} moments $\mathbb{E}(|\varepsilon_t|^p) < \infty$ for some $p \geq 2$. Hence, representation (8.3) holds. Denote

$$\Psi_t^{(m)} = \sum_{i=0}^{m} \varphi^i \varepsilon_{t-i},$$

which is the same as $\Upsilon_t^{(m)}$ defined by equation (8.4). (In general, $\Psi_t^{(m)}$ and $\Upsilon_t^{(m)}$ do not need to be the same.) Since $\text{dist}(I_i, I_j) > \max\{m_i, m_j\}$, condition (D) of M_p-mixing is satisfied,

just like condition (B) of S-mixing is. To check condition (C), we write

$$\left(\mathbb{E}\big(|X_t - \Psi_t^{(m)}|^p\big)\right)^{1/p} = \left(\mathbb{E}\bigg(\bigg|\sum_{i=m+1}^{\infty} \varphi^i \varepsilon_{t-i}\bigg|^p\bigg)\right)^{1/p}$$

$$\leq \sum_{i=m+1}^{\infty} \left(\mathbb{E}\big(|\varphi^i \varepsilon_{t-i}|^p\big)\right)^{1/p}$$

$$= \sum_{i=m+1}^{\infty} |\varphi|^i \left(\mathbb{E}\big(|\varepsilon_0|^p\big)\right)^{1/p}$$

$$= \frac{|\varphi|^{m+1}}{1 - |\varphi|} \left(\mathbb{E}\big(|\varepsilon_0|^p\big)\right)^{1/p}$$

$$=: \varrho_m,$$

where the bound is due to Minkowski's inequality. Hence, $\varrho_m = O(m^{-A})$ when $m \to \infty$ for any constant $A > 0$. In fact, the decay of ϱ_m is exponential.

The following theorem is a very special case of Berkes et al. (2011, Theorem 1, p. 2445), but it is exactly what we need.

Theorem 8.2. *Let $(X_t)_{t \in \mathbb{Z}}$ be a strictly stationary and M_p-mixing time series for some $p > 2$ and with*

$$\varrho_m = O\left(\frac{1}{m^A}\right), \tag{8.16}$$

where

$$A > \max\left\{1, \frac{p-2}{2\eta}\left(1 - \frac{1+\eta}{p}\right)\right\} \quad \text{and} \quad \frac{1+\eta}{p} < \frac{1}{2}. \tag{8.17}$$

Then, when $n \to \infty$,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X)\right) \xrightarrow{d} \mathcal{N}(0, \nu^2) \tag{8.18}$$

with the asymptotic variance

$$\nu^2 = \sum_{h=-\infty}^{\infty} \mathrm{Cov}(X_0, X_h).$$

Since the functions h_{UPPER}, h_{LOWER} and h_{MIDDLE} are Lipschitz continuous of order 1, the three transformed time series $(h_{\text{UPPER}}(X_t))_{t \in \mathbb{Z}}$, $(h_{\text{LOWER}}(X_t))_{t \in \mathbb{Z}}$ and $(h_{\text{MIDDLE}}(X_t))_{t \in \mathbb{Z}}$ are M_p-mixing with the same p and ϱ_m as the original time series $(X_t)_{t \in \mathbb{Z}}$. Furthermore, strict stationarity of $(X_t)_{t \in \mathbb{Z}}$ implies strict stationarity of the three transformed time series. Hence, we can now formulate the following corollary.

Corollary 8.4. *Let $F \in \mathcal{F}_p$ for some $p > 2$, and let the conditions of Theorems 8.1 and 8.2 be satisfied.*

(1) *If the quantile function F^{-1} is continuous at the point p, then statement (8.11) and, therefore, statement (3.24) hold with a limiting mean-zero normal random variable $\mathcal{L}_{\mathrm{UPPER}}(p)$ whose variance is*

$$\nu^2_{\mathrm{UPPER}} = \sum_{h=-\infty}^{\infty} \mathrm{Cov}\big(h_{\mathrm{UPPER}}(X_0), h_{\mathrm{UPPER}}(X_h)\big).$$

That is, when $n \to \infty$, we have

$$\sqrt{n}\left(\int_p^1 F_{n,\mathrm{TS}}^{-1}(u)\mathrm{d}u - \int_p^1 F^{-1}(u)\mathrm{d}u\right) \xrightarrow{d} \mathcal{N}\big(0, \nu^2_{\mathrm{UPPER}}\big).$$

(2) *If the quantile function F^{-1} is continuous at the point p, then statement (8.12) and, therefore, statement (5.12) hold with a limiting mean-zero normal random variable $\mathcal{L}_{\mathrm{LOWER}}(p)$ whose variance is*

$$\nu^2_{\mathrm{LOWER}} = \sum_{h=-\infty}^{\infty} \mathrm{Cov}\big(h_{\mathrm{LOWER}}(X_0), h_{\mathrm{LOWER}}(X_h)\big).$$

That is, when $n \to \infty$, we have

$$\sqrt{n}\left(\int_0^p F_{n,\mathrm{TS}}^{-1}(u)\mathrm{d}u - \int_0^p F^{-1}(u)\mathrm{d}u\right) \xrightarrow{d} \mathcal{N}\big(0, \nu^2_{\mathrm{LOWER}}\big).$$

(3) *If the quantile function F^{-1} is continuous at the points p_1 and p_2, then statement (8.13) and, therefore, statement (6.11) hold with a limiting mean-zero normal random variable $\mathcal{L}_{\mathrm{MIDDLE}}(p)$ whose variance is*

$$\nu^2_{\mathrm{MIDDLE}} = \sum_{h=-\infty}^{\infty} \mathrm{Cov}\big(h_{\mathrm{MIDDLE}}(X_0), h_{\mathrm{MIDDLE}}(X_h)\big).$$

That is, when $n \to \infty$, we have

$$\sqrt{n}\left(\int_{p_1}^{p_2} F_{n,\mathrm{TS}}^{-1}(u)\mathrm{d}u - \int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u\right) \xrightarrow{d} \mathcal{N}\big(0, \nu^2_{\mathrm{MIDDLE}}\big).$$

In the three statements of Corollary 8.4, the asymptotic variances are complex to estimate empirically, and so practically useful assessments of the variances should probably be done using resampling techniques (e.g. Lahiri, 2003; Kreiss and Paparoditis, 2011).

9 L-functionals

Given what we have established in the previous sections, we may now wish to see if those results could be extended to the class of so-called L-functionals, whose definition we recall next. Namely, the L-functional $L_w : \mathcal{F} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ is given by the equation

$$L_w(F) = \int_0^1 F^{-1}(u)w(u)\mathrm{d}u,$$

where $w : (0, 1) \to \mathbb{R}$ is a function. For illustrative "actuarial" examples where L-functionals arise, we refer to Wang (1998), and Jones and Zitikis (2003), with the latter paper being perhaps the first one to connect actuarial risk measures with L-functionals for the sake of developing statistical inference for the risk measures. To see the role of L-functionals in the theory of distorted expectations, which are of particular interest when assessing risks in economics, finance and insurance, we refer to Dhaene et al. (2012). Next is a specimen of additional examples that illustrate the w's that arise in statistics, finance, and insurance.

Example 9.1. Let $\mathcal{G} \subset \mathcal{F}$ be the family of normal (i.e., Gaussian) distributions. To assess the scale parameter when location is known (e.g., Serfling, 1980, p. 269), the L-functional $L_w : \mathcal{G} \to \mathbb{R}$ is used with the weight function

$$w(u) = \Phi^{-1}(u),$$

where Φ denotes the standard normal cdf.

Example 9.2. Let $\mathcal{F} \subset \mathcal{F}$ be the family of logistic distributions. To assess the location parameter when scale is known(e.g., Serfling, 1980, p. 269), the L-functional $L_w : \mathcal{F} \to \mathbb{R}$ is used with the weight function

$$w(u) = 6u(1 - u).$$

Example 9.3. The Gini Mean Difference (GMD) is the L-functional $L_w : \mathcal{F} \to \overline{\mathbb{R}}$ with the weight function

$$w(u) = 4u - 2.$$

For properties and manifold applications of the GMD in economics and finance, we refer to Yitzhaki and Schechtman (2013).

Example 9.4. The Tail Gini (Furman et al., 2017, p. 74) is the L-functional $L_w : \mathcal{F} \to \overline{\mathbb{R}}$ with the weight function

$$w(u) = \frac{\mathbb{1}\{p \le u < 1\}}{(1-p)^2}\big(4u - 2(1+p)\big),$$

where $p \in [0, 1)$ is a parameter.

Example 9.5. The Gini Shortfall (Furman et al., 2017, p. 75) is the L-functional $L_w : \mathcal{F} \to \overline{\mathbb{R}}$ with the weight function

$$w(u) = \frac{\mathbb{1}\{p \le u < 1\}}{(1-p)^2}\left(1 - p + 4\lambda\left(u - \frac{1+p}{2}\right)\right),$$

where $p \in [0, 1)$ and $\lambda \in [0, \infty)$ are parameters.

To give a flavour of the technical path that is often taken in the literature (e.g., Serfling, 1980, p. 265) when deriving large-sample statistical properties of L-integrals, we write the equations

$$
\begin{aligned}
L_w(G) - L_w(F) &= \int_0^1 \Big(G^{-1}(u) - F^{-1}(u)\Big)w(u)\mathrm{d}u \\
&= \int_{\mathbb{R}} \Big(K_w\big(F(x)\big) - K_w\big(G(x)\big)\Big)\mathrm{d}x
\end{aligned}
\tag{9.1}
$$

(see Appendix A.6.4 for technical details), where

$$K_w(t) = \int_0^t w(u)\mathrm{d}u,$$

assuming that w is integrable on $(0, 1)$. Note that equation (9.1) collapses into equation (1.6) when $w(u) \equiv 1$. With an estimator F_n instead of G, we can now establish asymptotic properties of the empirical L-estimator $L_w(F_n)$ when $n \to \infty$, but this path requires some form of the Taylor expansion applied on the function K_w, which, in turn, imposes smoothness conditions on the weight function w. Although in some applications such conditions are satisfied, in many other applications they are not.

As a way out of the difficulty, we can successfully use our developed theory for integrated quantiles, and the path toward achieving this goal is analogous to that given by

equation (9.1), although instead of integrating the weight function w as in the definition of K_w, we integrate the quantile function F^{-1}. The following example offers an idea how this goal can be achieved.

Example 9.6 (a proof of concept). For simplicity, let $F \in \mathcal{F}_1^+$, which means that integrals $\int_p^1 F^{-1}(u)\mathrm{d}u$ are finite for every $p \in (0,1)$, and let the weight function w be the difference $w_1 - w_2$ of two non-decreasing, right-continuous, and non-negative functions w_1 and w_2. An example of such a weight function w would be $w(u) = 6u(1-u)$ that appears in Example 9.2. Hence,

$$
\begin{aligned}
L_w(F) &= \int_0^1 F^{-1}(u)w_1(u)\mathrm{d}u - \int_0^1 F^{-1}(u)w_2(u)\mathrm{d}u \\
&= \int_0^\infty \left(\int_{w_1^{-1}(x)}^1 F^{-1}(u)\mathrm{d}u \right)\mathrm{d}x - \int_0^\infty \left(\int_{w_2^{-1}(x)}^1 F^{-1}(u)\mathrm{d}u \right)\mathrm{d}x, \quad (9.2)
\end{aligned}
$$

with the technical details justifying the right-most equation given in Appendix A.6.5. Equation (9.2) paves a path that connects the results of Section 3 with asymptotic properties of the difference $L_w(F_n) - L_w(F)$, where F_n, $n \in \mathbb{N}$, is a sequence of cdf's that approximate F when n increases. Furthermore, via the results of Dhaene et al. (2012), the equation also facilitates the development of statistical inference for distorted expectations.

The following theorem, whose proof is given in Appendix A.4, implements the idea of Example 9.6 under minimal assumptions on F and w, thus enabling to tackle L-functionals that arise from a whole spectrum of applications. To formulate the theorem, we use the notation

$$
L_{w,a,b}(F) = \int_a^b F^{-1}(u)w(u)\mathrm{d}u.
$$

Note that $L_w(F) = L_{w,a,b}(F)$ when $a = 0$ and $b = 1$.

Theorem 9.1. *Let the weight function w be non-decreasing and right-continuous on the interval $(a,b) \subseteq (0,1)$. Then, for every cdf F for which $L_{w,a,b}(F)$ is finite, we have the representation*

$$
L_{w,a,b}(F) = - \int_{-\infty}^0 \mathbb{1}\{w(a) < x\} \left(\int_a^{b \wedge w^{-1}(x)} F^{-1}(u)\mathrm{d}u \right)\mathrm{d}x
$$

$$
+ \int_0^\infty \mathbb{1}\{x \le w(b)\} \left(\int_{a \vee w^{-1}(x)}^b F^{-1}(u)\mathrm{d}u \right)\mathrm{d}x, \quad (9.3)
$$

where w^{-1} is the left-continuous inverse of w.

We see from the right-hand side of equation (9.3) that when $a = 0$ and $b = 1$, the L-integral $L_{w,a,b}(F)$, which is equal to $L_w(F)$, can be expressed in terms of integrated quantiles under the assumption that the weight function w is non-decreasing on the entire unit interval $(0,1)$, but this is a rather restrictive assumption in many applications. For this reason, we shall show next that all w's of practical relevance can be expressed in terms of linear combinations of non-decreasing functions on various subintervals of $(0,1)$. This reduces $L_w(F)$ to a linear combinations of L-integrals $L_{w,a,b}(F)$ with various non-decreasing weight functions w over various intervals $(a,b) \subset (0,1)$. To develop statistical inference for the L-integrals $L_{w,a,b}(F)$, we use the equation

$$L_{w,a,b}(G) - L_{w,a,b}(F) = -\int_{-\infty}^{0} \mathbb{1}\{w(a) < x\}\left(\int_{a}^{b \wedge w^{-1}(x)} \left(G^{-1}(u) - F^{-1}(u)\right)\mathrm{d}u\right)\mathrm{d}x$$
$$+ \int_{0}^{\infty} \mathbb{1}\{x \leq w(b)\}\left(\int_{a \vee w^{-1}(x)}^{b} \left(G^{-1}(u) - F^{-1}(u)\right)\mathrm{d}u\right)\mathrm{d}x$$

that follows from Theorem 9.1. Using appropriate results from Sections 3, 5, and 6, we can now convert integrals of $G^{-1} - F^{-1}$ into integrals of $F - G$ for which asymptotics with F_n instead of G becomes straightforward.

Example 9.7. Consider the weight function

$$w(u) = 6u(1 - u),$$

which we have encountered in Example 9.2. Obviously, we can decompose this weight function as the difference of $6u$ and $6u^2$ on the entire interval $(0,1)$. Alternatively, we can decompose it as the difference of $-6(1-u)^2$ and $-6(1-u)$ on the entire interval $(0,1)$. However, if we adopt the first decomposition, then we need to assume the finiteness of the moment $\mathbb{E}(X^+)$, and if we adopt the second decomposition, then we need to assume the finiteness of the moment $\mathbb{E}(X^-)$. Either way, superfluous conditions arise, which are not required for the finiteness of the L-integral $L_w(F)$ because the original weight function $w(u)$ converges to 0 when $u \downarrow 0$ and $u \uparrow 1$. For this reason, to reduce the problem to non-decreasing functions, we split the interval $[0,1)$ into two parts: $[0, 1/2)$ where the function w is increasing, and $[1/2, 1)$ where the function w is decreasing and thus $-w$ is increasing. Hence, we can express w as the sum of two differences of non-decreasing functions, and this can be done in several

ways. One way is to use the equation

$$w(u) = \mathbb{1}\{0 \le u < 1/2\}\Big(w_{11}(u) - w_{12}(u)\Big) + \mathbb{1}\{1/2 \le u < 1\}\Big(w_{21}(u) - w_{22}(u)\Big) \quad (9.4)$$

with the functions

$$w_{11}(u) = w(u) \quad \text{and} \quad w_{12}(u) = 0,$$
$$w_{21}(u) = 0 \quad \text{and} \quad w_{22}(u) = -w(u).$$

Another way is to use equation (9.4) with the functions

$$w_{11}(u) = 6u \quad \text{and} \quad w_{12}(u) = 6u^2,$$
$$w_{21}(u) = -6(1 - u)^2 \quad \text{and} \quad w_{22}(u) = -6(1 - u).$$

Note that irrespective of the way we use, we always have the bounds

$$|w_{11}(u)| \vee |w_{12}(u)| \le c|w(u)| \quad \text{for all} \quad u \in [0, 1/2),$$
$$|w_{21}(u)| \vee |w_{22}(u)| \le c|w(u)| \quad \text{for all} \quad u \in [1/2, 1),$$

with some positive constant $c < \infty$. Therefore, no superfluous moment conditions arise when developing statistical inference for the L-integrals based on the four functions w_{ij}. This concludes Example 9.7.

Example 9.7 naturally leads to the following condition, which plays a fundamental role in connecting statistical inference for the L-integral $L_w(F)$ with that for integrated quantiles.

(C_w) Let the weight function $w : (0, 1) \to \mathbb{R}$ be such that, for some integer $K \ge 1$, there is a partition $0 = a_0 < a_1 < \cdots < a_K = 1$ of the unit interval $(0, 1)$ and also pairs $(w_{11}, w_{12}), \ldots, (w_{K1}, w_{K2})$ of non-decreasing and right-continuous on $(0, 1)$ functions such that, for a constant $c < \infty$,

$$|w_{11}(u)| \vee |w_{12}(u)| \le c|w(u)| \quad \text{in a neighbourhood of 0,} \quad (9.5)$$
$$|w_{K1}(u)| \vee |w_{K2}(u)| \le c|w(u)| \quad \text{in a neighbourhood of 1,} \quad (9.6)$$

and

$$w(u) = \sum_{k=1}^{K} \mathbb{1}\{a_{k-1} \leq u < a_k\}\Big(w_{k1}(u) - w_{k2}(u)\Big). \tag{9.7}$$

To appreciate condition $(\mathrm{C_w})$, which may look somewhat complex at first sight, we observe that if the function $w : (0,1) \to \mathbb{R}$ is piecewise monotonic, then the assumption is satisfied. All the weight functions that we have encountered in the literature are piecewise monotonic. In the very special case when w is monotonic (i.e., either non-decreasing or non-increasing) on the entire domain $(0,1)$ of its definition, condition $(\mathrm{C_w})$ is satisfied with $K = 1$. The following example illustrates the case $K = 2$. Interestingly, we shall see from following Theorem 9.2, whose proof is constructive and given in Appendix A.5, that if condition $(\mathrm{C_w})$ holds with $K \geq 3$, then the condition can be reduced to the case $K = 2$.

Example 9.8. Consider the weight function

$$w(u) = (1 - u)^\beta$$

defined for all $u \in [0,1)$ and parameterized by $\beta \in \mathbb{R}$. The function can be expressed as the difference of two non-decreasing functions as follows:

$$w(u) = \underbrace{\mathbb{1}\{\beta \leq 0\}(1-u)^\beta}_{w_1(u)} - \underbrace{(-1)\mathbb{1}\{\beta > 0\}(1-u)^\beta}_{w_2(u)}.$$

The two functions w_1 and w_2 are continuous and non-decreasing, and satisfy the bound

$$|w_1(u)| \vee |w_2(u)| \leq |w(u)|$$

for all $u \in (0,1)$. That is, for the weight function w we do not need to split the interval $(0,1)$ into parts, and we can therefore set $K = 1$. This concludes Example 9.8.

Theorem 9.2. *If condition* $(\mathrm{C_w})$ *is satisfied with* $K \geq 3$, *then it is also satisfied with* $K = 2$.

Theorem 9.2, whose proof is given in Appendix A.5, says that there is a constant $a \in [0,1]$ and two pairs (w_{11}, w_{12}) and (w_{21}, w_{22}) of non-decreasing and right-continuous functions such that bounds (9.5) and (9.6) are satisfied and the representation

$$w(u) = \mathbb{1}\{0 \leq u < a\}\Big(w_{11}(u) - w_{12}(u)\Big) + \mathbb{1}\{a \leq u < 1\}\Big(w_{21}(u) - w_{22}(u)\Big) \tag{9.8}$$

holds. Earlier in Example 9.7, we showed that the case $K = 2$ cannot be generally reduced to $K = 1$. In other words, from the theoretical point of view, it is sufficient to consider only the cases $K = 1$ and $K = 2$. From the practical point of view, this minimality property could sometimes be time consuming to realize, but this is not important: we can always work with any $K \geq 1$ that we can easily obtain.

10   Conclusion

We have developed a general statistical inference theory for integrals of quantiles. The required conditions are formulated in terms of general approximating sequences of cdf's F_n, which are not attached to any specific sampling design or dependence structure. To maximally illuminate the developed theory, we have illustrated the obtained results using simple random sampling, due to everyone's familiarity with this sampling design and basic statistical results. In particular, we have discussed consistency, bias, and asymptotic normality of various integrals of quantiles and their combinations that arise in risk and economic-inequality measurements: the upside and downside tail-values-at-risk (TVaR and TV@R, respectively), and the Lorenz and Gini curves. Furthermore, we have illustrated how our general results can be used in the case of time series data. To facilitate the appreciation of theoretical results, we have illustrated some of them numerically. Finally, we have linked the herein developed theory for integrated quantiles with topics such as L-integrals.

References

Aaronson, J., Burton, R., Dehling, H., Gilat, D., Hill, T. and Weiss, B. (1996). Strong laws for L- and U-statistics. *Transactions of the American Mathematical Society*, 348, 2845–2866. http://www.jstor.org/stable/2155273

Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (2008). *A First Course in Order Statistics.* Society for Industrial and Applied Mathematics, Philadelphia. https://doi.org/10.1137/1.9780898719062

Arnold, B.C. and Villaseñor, J.A. (2015). Bias of the sample Lorenz curve. In: *Ordered Data Analysis, Modeling and Health Research Methods* (eds. P.K. Choudhary,

C.H. Nagaraja, and H.K.T. Ng), pp. 3–15. Springer, Cham. https://doi.org/10.1007/978-3-319-25433-3_1

Bahadur, R. R. (1966). A note on quantiles in large samples. *Annals of Mathematical Statistics*, 37, 577–580. http://www.jstor.org/stable/2238725

Bellini, F., Fadina, T., Wang, R. and Wei, Y. (2022). Parametric measures of variability induced by risk measures. *Insurance: Mathematics and Economics*, 106, 270–284.

Berkes, I., Hörmann, S. and Schauer, J. (2009). Asymptotic results for the empirical process of stationary sequences. *Stochastic Processes and their Applications*, 119, 1298–1324. https://doi.org/10.1016/j.spa.2008.06.010

Berkes, I., Hörmann, S. and Schauer, J. (2011). Split invariance principles for stationary processes. *Annals of Probability*, 39, 2441–2473. http://www.jstor.org/stable/23078763

Billingsley, P. (1999). *Convergence of Probability Measures.* (Second Edition.) Wiley, New York. https://doi.org/10.1002/9780470316962

Brazauskas, V., Jones, B.L. and Zitikis, R. (2009). Robust fitting of claim severity distributions and the method of trimmed moments. *Journal of Statistical Planning and Inference*, 139, 2028–2043. https://doi.org/10.1016/j.jspi.2008.09.012

Brazauskas, V., Jones, B.L., Puri, M.L. and Zitikis, R. (2008). Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference*, 138, 3590–3604. https://www.sciencedirect.com/science/article/abs/pii/S0378375808001018

Chen, S.X. (2008). Nonparametric estimation of Expected Shortfall. *Journal of Financial Econometrics*, 6, 87–107. https://doi.org/10.1093/jjfinec/nbm019

Csáki, E., Csörgő, M., Földes, A., Shi, Z. and Zitikis, R. (2002). Pointwise and uniform asymptotics of the Vervaat error process. *Journal of Theoretical Probability*, 15 (4), 845–875. https://doi.org/10.1023/A:1020650502619

Csörgő, M., Csörgő, S. and Horváth, L. (1986). *An Asymptotic Theory for Empirical Reliability and Concentration Processes.* Springer, Heidelberg. https://doi.org/10.1007/978-1-4615-6420-1

Csörgő, M., Gastwirth, J.L. and Zitikis, R. (1998). Asymptotic confidence bands for the Lorenz and Bonferroni curves based on the empirical Lorenz curve. *Journal of Statistical Planning and Inference*, 74, 65–91. https://doi.org/10.1016/S0378-3758(98)00103-7

Csörgő, M. and Zitikis, R. (1996). Strassen's LIL for the Lorenz curve. *Journal of Multivariate Analysis*, 59, 1–12. https://doi.org/10.1006/jmva.1996.0050

Csörgő, M. and Zitikis, R. (1999). On the Vervaat and Vervaat-error processes. *Acta Applicandae Mathematicae* (Special issue on Limit Theorems of Probability Theory in Honour of Vytautas Statulevičius), 58, 91–105. https://doi.org/10.1023/A:1006329502117

Csörgő, M. and Zitikis, R. (2001). The Vervaat process in L_p-spaces. *Journal of Multivariate Analysis*, 78, 103–138. https://doi.org/10.1006/jmva.2000.1941

David, H.A. and Nagaraja, H.N. (2003). *Order Statistics.* (Third Edition.) Wiley, Hoboken. https://doi.org/10.1002/0471722162

Davydov, Y. and Zitikis, R. (2002). Convergence of generalized Lorenz curves based on stationary ergodic random sequences with deterministic noise. *Statistics and Probability Letters*, 59, 329–340. https://doi.org/10.1016/S0167-7152(02)00202-X

Davydov, Y. and Zitikis, R. (2003). Generalized Lorenz curves and convexifications of stochastic processes. *Journal of Applied Probability*, 40, 906–925. https://www.jstor.org/stable/3216050

Davydov, Y. and Zitikis, R. (2004). Convex rearrangements of random elements. In: *Asymptotic Methods in Stochastics: Festschrift for Miklós Csörgő* (eds. L. Horváth and B. Szyszkowicz), Fields Institute Communications, Vol. 44, 141–171. American Mathematical Society, Providence, Rhode Island. https://doi.org/10.1090/fic/044/08

Davydov, Y., Khoshnevisan, D., Shi, Z. and Zitikis, R. (2007). Convex rearrangements, generalized Lorenz curves, and correlated Gaussian data. *Journal of Statistical Planning and Inference* (Special issue on "Nonparametric Statistics and Related Topics: In Honor of M.L. Puri" edited by S. Ralescu, G. Haiman, and F. Ruymgaart), 137, 915–934. https://doi.org/10.1016/j.jspi.2006.06.032

Dehling, H. Mikosch, T. and Sørensen, M. (2002). *Empirical Process Techniques for Dependent Data.* Birkhäuser, Boston. https://doi.org/10.1007/978-1-4612-0099-4

Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2005) *Actuarial Theory for Dependent Risks: Measures, Orders and Models.* Wiley, Chichester. https://doi.org/10.1002/0470016450

Dhaene, J., Kukush, A., Linders, D. and Tang, Q. (2012). Remarks on quantiles and distortion risk measures. *European Actuarial Journal*, 2, 319–328. https://doi.org/10.1007/s13385-012-0058-0

Doukhan, P. (2018). *Stochastic Models for Time Series.* Springer, Berlin. https://doi.org/10.1007/978-3-319-76938-7

Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time.* (Fourth Edition.) Walter de Gruyter, Berlin.

Furman, E., Wang, R., Zitikis, R. (2017). Gini-type measures of risk and variability: Gini shortfall, capital allocations, and heavy-tailed risks. *Journal of Banking and Finance*, 83, 70–84. https://doi.org/10.1016/j.jbankfin.2017.06.013

Garca-Risueño, P. (2025). Historical simulation systematically underestimates the Expected Shortfall. *Journal of Risk and Financial Management*, 18, Article #34. https://doi.org/10.3390/jrfm18010034

Gastwirth, J.L. (1971). A general definition of the Lorenz curve. *Econometrica*, 39, 1037–1039. https://doi.org/10.2307/1909675

Gilat, D. and Helmers, R. (1997). On strong laws for generalized L-statistics with dependent data. *Commentationes Mathematicae Universitatis Carolinae*, 38, 187–192. http://dml.cz/dmlcz/118915

Gribkova, N.V. (1995). Bounds for absolute moments of order statistics. In *Exploring Stochastic Laws* (eds. A.V. Skorokhod and Yu.V. Borovskikh) pp. 129–134. de Gruyter, Berlin. https://doi.org/10.1515/9783112318768-015

Gribkova, N.V. and Helmers, R. (2006). The empirical Edgeworth expansion for a studentized trimmed mean. *Mathematical Methods of Statistics*, 15, 61–87.

Gribkova, N.V. and Helmers, R. (2007). On the Edgeworth expansion and the M out of N bootstrap accuracy for a studentized trimmed mean. *Mathematical Methods of Statistics*, 16, 142–176. https://doi.org/10.3103/S1066530707020056

Gribkova, N.V. and Helmers, R. (2011). On the consistency of the $M \ll N$ bootstrap approximation for a trimmed mean. *Theory of Probability and Its Applications,* 55, 42–53. https://doi.org/10.1137/S0040585X97984607

Gribkova, N., Su, J. and Zitikis, R. (2022a). Empirical tail conditional allocation and its consistency under minimal assumptions. *Annals of the Institute of Statistical Mathematics*, 74, 713–735. https://doi.org/10.1007/s10463-021-00813-3

Gribkova, N., Su, J. and Zitikis, R. (2022b). Inference for the tail conditional allocation: large sample properties, insurance risk assessment, and compound sums of concomitants. *Insurance: Mathematics and Economics*, 107, 199–222. https://doi.org/10.1016/j.insmatheco.2022.08.009

Gribkova, N., Su, J. and Zitikis, R. (2024). Assessing the coverage probabilities of fixed-margin confidence intervals for the tail conditional allocation. *Annals of the Institute of Statistical Mathematics*, 76 (5), 821–850. https://doi.org/10.1007/s10463-024-00904-x

Han, X., Wang, R. and Wu, Q. (2024). Monotonic mean-deviation risk measures. Available on `arXiv:2312.01034`. https://doi.org/10.48550/arXiv.2312.01034

Helmers, R. and Zitikis, R. (2005). Strong laws for generalized absolute Lorenz curves when data are stationary and ergodic sequences. *Proceedings of the American Mathematical Society*, 133, 3703–3712. https://doi.org/10.1090/S0002-9939-05-08096-2

Jones, B.L. and Zitikis, R. (2003). Empirical estimation of risk measures and related quantities. *North American Actuarial Journal*, 7, 44–54.

Kiefer, J. (1967). On Bahadur's representation of sample quantiles. *Annals of Mathematical Statistics*, 38, 1323–1342. http://www.jstor.org/stable/2238951

Kreiss, J.-P. and Paparoditis, E. (2011). Bootstrap methods for dependent data: A review. *Journal of the Korean Statistical Society*, 40, 357–378. https://doi.org/10.1016/j.jkss.2011.08.009

Lahiri, S.N. (2003). *Resampling Methods for Dependent Data.* Springer, New York. https://doi.org/10.1007/978-1-4757-3803-2

Li, H. and Wang, R. (2023). PELVE: Probability Equivalent Level of VaR and ES. *Journal of Econometrics*, 234, 353–370. https://doi.org/10.1016/j.jeconom.2021.12.012

Linton, O. and Xiao, Z. (2013). Estimation of and inference about the Expected Shortfall for time series with infinite variance. *Econometric Theory*, 29, 771–807. https://doi.org/10.1017/S0266466612000692

Moutanabbir, K. and Bouaddi, M. (2024). A new non-parametric estimation of the expected shortfall for dependent financial losses. *Journal of Statistical Planning and Inference*, 232, Article 106151, https://doi.org/10.1016/j.jspi.2024.106151

Nadarajah, S., Zhang, B. and Chan, S. (2013). Estimation methods for Expected Shortfall. *Quantitative Finance*, 14, 271–291. https://doi.org/10.1080/14697688.2013.816767

Parzen, E. (1979a). Nonparametric statistical data modeling. *Journal of the American Statistical Association*, 74, 105–121. https://doi.org/10.2307/2286734

Parzen, E. (1979b). A density-quantile function perspective on robust estimation. In: *Robustness in Statistics* (eds. R.L. Launer and G.N. Wilkinson), pp. 237–258. Academic Press, Orlando. https://doi.org/10.1016/B978-0-12-438150-6.50019-4

Pflug, G.Ch. and Römisch, W. (2007). *Modeling, Measuring and Managing Risk.* Singapore, World Scientific. https://doi.org/10.1142/6478

Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation.* Academic Press, Orlando. https://doi.org/10.1016/C2013-0-11326-8

Ramdas, A. and Wang, R. (2024). *Hypothesis Testing with E-values.* Available on arXiv:2410.23614. https://doi.org/10.48550/arXiv.2410.23614

Rio, E. (2017). *Asymptotic Theory of Weakly Dependent Random Processes.* Springer, Berlin. https://doi.org/10.1007/978-3-662-54323-8

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York. https://doi.org/10.1002/9780470316481

Shao, J. (2003). *Mathematical Statistics.* (Second Edition.) Springer, New York. https://doi.org/10.1007/b97553

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* Springer, New York. https://doi.org/10.1007/978-1-4612-0795-5

Shorack, G.R. (2017). *Probability for Statisticians.* (Second Edition.) Springer, New York. https://doi.org/10.1007/978-3-319-52207-4

Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York. https://doi.org/10.1137/1.9780898719017

Stigler, S.M. (1973). The asymptotic distribution of the trimmed mean. *Annals of Statistics*, 1, 472–477. https://doi.org/10.1214/aos/1176342412

Stigler, S.M. (1974). Linear functions of order statistics with smooth weight functions. *Annals of Statistics*, 2, 676–693. https://doi.org/10.1214/aos/1176342756

Vervaat, W. (1972a). *Success Epochs in Bernouli Trials: with Applications in Number Theory.* Mathematical Centre Tracts 42. Centrum voor Wiskunde en Informatica, Amsterdam. https://ir.cwi.nl/pub/19565/

Vervaat, W. (1972b). Functional central limit theorems for processes with positive drift and their inverses. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 23, 245–253. https://doi.org/10.1007/BF00532510

Wacker, P. (2023). Please, not *another* note about Generalized Inverses. *Texchnical Report* arXiv:2306.06989. https://doi.org/10.48550/arXiv.2306.06989

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin: the Journal of the International Actuarial Association*, 26, 71–92. https://doi.org/10.2143/AST.26.1.563234

Wang, S. (1998). An actuarial index of the right-tail risk, *North American Actuarial Journal*, 2, 88–101.

Wei, Y. and Zitikis, R. (2023). Assessing the difference between integrated quantiles and integrated cumulative distribution functions. *Insurance: Mathematics and Economics*, 111, 163–172. https://doi.org/10.1016/j.insmatheco.2023.04.002

Yitzhaki, S. and Schechtman, E. (2013) *The Gini Methodology: A Primer on a Statistical Methodology.* Springer, New York. https://doi.org/10.1007/978-1-4614-4720-7

Zhao, Q., Brazauskas, V. and Ghorai, J. (2018). Robust and efficient fitting of severity models and the method of Winsorized moments. *ASTIN Bulletin: the Journal of the International Actuarial Association*, 48, 275–309. https://doi.org/10.1017/asb.2017.30

Ziegel, J.F., Krüger, F., Jordan, A. and Fasciati, F. (2020). Robust forecast evaluation of Expected Shortfall. *Journal of Financial Econometrics*, 18, 95–120. https://doi.org/10.1093/jjfinec/nby035

Zitikis, R. (1998). The Vervaat process. In: *Asymptotic Methods in Probability and Statistics – A Volume in Honour of Miklós Csörgő* (ed. B. Szyszkowicz), pp. 667–694. North-Holland, Amsterdam. https://doi.org/10.1016/B978-044450083-0/50044-7

A Technicalities

We have subdivided this appendix into several major parts: Appendix A.1 contains auxiliary lemmas that we shall need in Appendix A.2 for proving Theorems 3.1, 5.1, and 6.1. In Appendix A.3 we shall derive asymptotic expansions for the empirical Lorenz and Gini curves that are needed for establishing their asymptotic normality stated in Examples 7.3 and 7.4. Appendices A.4 and A.5 contain proofs of Theorems 9.1 and 9.2, respectively. In Appendix A.6 we prove several cursory notes made earlier in the paper when introducing and discussing the main results.

A.1 Auxiliary lemmas

In this appendix we shall derive formulas for, and establish finiteness of, the first and second moments of the random variables $Y_{\text{UPPER}}(p)$ and $Y_{\text{LOWER}}(p)$. We shall need these results in Appendix A.2. The use of Fubini's theorem and distributional equations $X \overset{d}{=} F^{-1}(U)$ and $Y \overset{d}{=} G^{-1}(U)$ will permeate our considerations, where U is a uniform on $[0, 1]$ random variable and $\overset{d}{=}$ denotes equality in distribution.

Lemma A.1. *If $F \in \mathcal{F}_1^+$, then the random variable $Y_{\text{UPPER}}(p)$ has a finite first moment, and*

the following equations hold:

$$\mathbb{E}\big(Y_{\text{UPPER}}(p)\big) = \int_{F^{-1}(p)}^{\infty} \big(1 - F(x)\big)\mathrm{d}x \tag{A.1}$$

$$= \int_{p}^{1} F^{-1}(u)\mathrm{d}u - (1-p)F^{-1}(p). \tag{A.2}$$

If $F \in \mathcal{F}_2^+$, then $Y_{\text{UPPER}}(p)$ has a finite second moment, and formula (3.27) holds for the variance of $Y_{\text{UPPER}}(p)$.

Proof. We start with the equations

$$Y_{\text{UPPER}}(p) = \big(X - F^{-1}(p)\big)^+ \tag{A.3}$$

$$\stackrel{d}{=} \big(F^{-1}(U) - F^{-1}(p)\big)^+. \tag{A.4}$$

The first moment of the random variable on the right-hand side of equation (A.4) is finite whenever $F \in \mathcal{F}_1^+$. We can now use Fubini's theorem to interchange the expectation and integration operations on the right-hand side of the equation

$$\mathbb{E}\big(Y_{\text{UPPER}}(p)\big) = \mathbb{E}\bigg(\int_{F^{-1}(p)}^{\infty} \mathbb{1}\{X > x\}\mathrm{d}x\bigg)$$

and in this way arrive at equation (A.1) because $\mathbb{E}\big(\mathbb{1}\{X > x\}\big) = 1 - F(x)$. To prove equation (A.2), we write

$$\begin{aligned}
\int_{F^{-1}(p)}^{\infty} \big(1 - F(x)\big)\mathrm{d}x &= \mathbb{E}\bigg(\int_{F^{-1}(p)}^{\infty} \mathbb{1}\{X > x\}\mathrm{d}x\bigg) \\
&= \mathbb{E}\bigg(\big(F^{-1}(U) - F^{-1}(p)\big)^+\bigg) \\
&= \int_0^1 \big(F^{-1}(s) - F^{-1}(p)\big)^+\mathrm{d}s \\
&= \int_p^1 \big(F^{-1}(s) - F^{-1}(p)\big)\mathrm{d}s \\
&= \int_p^1 F^{-1}(s)\mathrm{d}s - (1-p)F^{-1}(p).
\end{aligned}$$

To prove the second half of the lemma, we first note that in view of equation (A.4), the random variable $Y_{\text{UPPER}}(p)$ has a finite second moment whenever $F \in \mathcal{F}_2^+$. To prove

equation (3.27), we use Fubini's theorem and have

$$
\begin{aligned}
\mathrm{Var}\left(Y_{\mathrm{UPPER}}(p)\right) &= \mathrm{Var}\left(\int_{F^{-1}(p)}^{\infty} \mathbb{1}\{X > x\}\mathrm{d}x\right) \\
&= \mathbb{E}\left(\left(\int_{F^{-1}(p)}^{\infty}\left(\mathbb{1}\{X > x\} - (1 - F(x))\right)\mathrm{d}x\right)^2\right) \\
&= \int_{F^{-1}(p)}^{\infty}\int_{F^{-1}(p)}^{\infty}\mathbb{E}\left(\left(\mathbb{1}\{X \le x\} - F(x)\right)\left(\mathbb{1}\{X \le y\} - F(y)\right)\right)\mathrm{d}x\mathrm{d}y \\
&= \int_{F^{-1}(p)}^{\infty}\int_{F^{-1}(p)}^{\infty}\left(F(x \wedge y) - F(x)F(y)\right)\mathrm{d}x\mathrm{d}y.
\end{aligned}
$$

This completes the proof of Lemma A.1. $\qquad\square$

Lemma A.2. *If $F \in \mathcal{F}_1^-$, then the random variable $Y_{\mathrm{LOWER}}(p)$ has a finite first moment, and the following equations hold:*

$$
\mathbb{E}\left(Y_{\mathrm{LOWER}}(p)\right) = \int_{-\infty}^{F^{-1}(p)} F(x)\mathrm{d}x \tag{A.5}
$$

$$
= pF^{-1}(p) - \int_0^p F^{-1}(u)\mathrm{d}u. \tag{A.6}
$$

If $F \in \mathcal{F}_2^-$, then $Y_{\mathrm{LOWER}}(p)$ has a finite second moment, and formula (5.15) holds for the variance of $Y_{\mathrm{LOWER}}(p)$.

Proof. We start with the equations

$$
Y_{\mathrm{LOWER}}(p) = \left(F^{-1}(p) - X\right)^+ \tag{A.7}
$$

$$
\stackrel{d}{=} \left(F^{-1}(p) - F^{-1}(U)\right)^+. \tag{A.8}
$$

The first moment of the random variable on the right-hand side of equation (A.8) is finite whenever $F \in \mathcal{F}_1^-$. We can now use Fubini's theorem to interchange the expectation and integration operations on the right-hand side of the equation

$$
\mathbb{E}\left(Y_{\mathrm{LOWER}}(p)\right) = \mathbb{E}\left(\int_{-\infty}^{F^{-1}(p)} \mathbb{1}\{X \le x\}\mathrm{d}x\right)
$$

and in this way arrive at equation (A.5) because $\mathbb{E}\left(\mathbb{1}\{X \le x\}\right) = F(x)$. To prove equa-

tion (A.6), we write

$$\int_{-\infty}^{F^{-1}(p)} F(x)\mathrm{d}x = \mathbb{E}\left(\int_{-\infty}^{F^{-1}(p)} \mathbb{1}\{X \leq x\}\mathrm{d}x\right)$$

$$= \mathbb{E}\left(\left(F^{-1}(p) - F^{-1}(U)\right)^+\right)$$

$$= \int_0^1 \left(F^{-1}(p) - F^{-1}(u)\right)^+\mathrm{d}u$$

$$= \int_0^p \left(F^{-1}(p) - F^{-1}(u)\right)\mathrm{d}u$$

$$= pF^{-1}(p) - \int_0^p F^{-1}(u)\mathrm{d}u.$$

To prove the second half of the lemma, we first note that in view of equation (A.8), the random variable $Y_{\mathrm{LOWER}}(p)$ has a finite second moment whenever $F \in \mathcal{F}_2^-$. To prove equation (5.15), we use Fubini's theorem and have

$$\mathrm{Var}\left(Y_{\mathrm{LOWER}}(p)\right) = \mathrm{Var}\left(\int_{-\infty}^{F^{-1}(p)} \mathbb{1}\{X \leq x\}\mathrm{d}x\right)$$

$$= \mathbb{E}\left(\left(\int_{-\infty}^{F^{-1}(p)} \left(\mathbb{1}\{X \leq x\} - F(x)\right)\mathrm{d}x\right)^2\right)$$

$$= \int_{-\infty}^{F^{-1}(p)} \int_{-\infty}^{F^{-1}(p)} \mathbb{E}\left(\left(\mathbb{1}\{X \leq x\} - F(x)\right)\left(\mathbb{1}\{X \leq y\} - F(y)\right)\right)\mathrm{d}x\mathrm{d}y$$

$$= \int_{-\infty}^{F^{-1}(p)} \int_{-\infty}^{F^{-1}(p)} \left(F(x \wedge y) - F(x)F(y)\right)\mathrm{d}x\mathrm{d}y.$$

This completes the proof of Lemma A.2. $\qquad\square$

Note A.1. The asymptotic variance $\sigma^2_{\mathrm{MIDDLE}}(p_1, p_2)$ in statement (6.13) is the variance of the random variable $Y_{\mathrm{MIDDLE}}(p_1, p_2)$ defined by equation (6.7). Since both $F^{-1}(p_1)$ and $F^{-1}(p_2)$ are finite, the random variable $Y_{\mathrm{MIDDLE}}(p_1, p_2)$ is bounded and thus has finite moments of all degrees irrespective of the cdf F. Note the equations

$$Y_{\mathrm{MIDDLE}}(p_1, p_2) = \int_{-\infty}^{F^{-1}(p_2)} \mathbb{1}\{X \leq x\}\mathrm{d}x - \int_{-\infty}^{F^{-1}(p_1)} \mathbb{1}\{X \leq x\}\mathrm{d}x$$

$$= \left(F^{-1}(p_2) - X\right)^+ - \left(F^{-1}(p_1) - X\right)^+.$$

This difference of the two positive parts, which may not have finite variances individually, always has a finite variance (as well as all the moments of higher degrees) irrespective of the

cdf F, because $0 < p_1 < p_2 < 1$. Equation (6.14) follows immediately from

$$
\begin{aligned}
\Big(Y_{\text{MIDDLE}}(p_1, p_2) - \mathbb{E}\big(Y_{\text{MIDDLE}}(p_1, p_2)\big)\Big)^2 & \\
= \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} &\big(\mathbb{1}\{X \le x\} - F(x)\big)\big(\mathbb{1}\{X \le y\} - F(y)\big)\mathrm{d}x\mathrm{d}y
\end{aligned}
$$

and Fubini's theorem. This concludes Note A.1.

A.2 Proofs of Theorems 3.1, 5.1, and 6.1

The three theorems deal with arbitrary cdf's F and G that, at most, are allowed to satisfy only moment conditions, with no further assumptions. It should be noted at the outset that we may find some portions of the following proofs look like simple consequences of integration-by-parts and change-of-variable formulas of calculus, but this would be mathematically inaccurate because, strictly speaking, the quantile function F^{-1} is not an ordinary inverse of the cdf F, unless the latter is strictly increasing and continuous, which is not necessarily the case in the context of the present paper. Indeed, keeping in mind that our general results need to include, as a special case, the empirical cdf $F_{n,\text{SRS}}$, which is a step-wise function and thus has flat regions as well as jumps, good care (recall Wacker (2023)) needs to be taken to maintain the largest possible class of cdf's. Hence, only moment-type assumptions are required, as specified in the formulation of theorems.

Proof of Theorem 3.1. By Lemma A.1, we have

$$
\int_p^1 F^{-1}(u)\mathrm{d}u = (1-p)F^{-1}(p) + \int_{F^{-1}(p)}^\infty \big(1 - F(x)\big)\mathrm{d}x. \tag{A.9}
$$

Since the same equation also holds with G instead of F, taking the difference between the two equations gives

$$
\begin{aligned}
\int_p^1 \big(G^{-1}(u) - F^{-1}(u)\big)\mathrm{d}u = (1-p)\big(G^{-1}(p) - F^{-1}(p)\big) & \\
+ \int_{G^{-1}(p)}^\infty \big(1 - G(x)\big)\mathrm{d}x - \int_{F^{-1}(p)}^\infty &\big(1 - F(x)\big)\mathrm{d}x. \tag{A.10}
\end{aligned}
$$

Next we express the penultimate integral on the right-hand side of equation (A.10) as follows:

$$\int_{G^{-1}(p)}^{\infty} \big(1 - G(x)\big)\mathrm{d}x = \int_{G^{-1}(p)}^{F^{-1}(p)} \big(1 - G(x)\big)\mathrm{d}x + \int_{F^{-1}(p)}^{\infty} \big(1 - G(x)\big)\mathrm{d}x, \tag{A.11}$$

where the first integral on the right-hand side is equal to $-\int_{F^{-1}(p)}^{G^{-1}(p)} \big(1 - G(x)\big)\mathrm{d}x$ when $F^{-1}(p) < G^{-1}(p)$. Combining equations (A.10) and (A.11), we arrive at equation (3.2) with the remainder term

$$\mathrm{Rem}(p; F, G) = (1 - p)\big(F^{-1}(p) - G^{-1}(p)\big) - \int_{G^{-1}(p)}^{F^{-1}(p)} \big(1 - G(x)\big)\mathrm{d}x$$

that can be concisely written as the integral on the right-hand side of equation (3.3).

To show that $\mathrm{Rem}(p; F, G)$ is non-negative, we proceed as follows. If $F^{-1}(p) \geq G^{-1}(p)$, then $x \geq G^{-1}(p)$, which is equivalent to $G(x) \geq p$, and so

$$\mathrm{Rem}(p; F, G) = \int_{G^{-1}(p)}^{F^{-1}(p)} \big(G(x) - p\big)\mathrm{d}x \geq 0.$$

If $F^{-1}(p) < G^{-1}(p)$, then $x < G^{-1}(p)$, which is equivalent to $G(x) < p$, and so

$$\mathrm{Rem}(p; F, G) = -\int_{F^{-1}(p)}^{G^{-1}(p)} \big(G(x) - p\big)\mathrm{d}x \geq 0.$$

Hence, we always have $\mathrm{Rem}(p; F, G) \geq 0$.

Furthermore,

$$\begin{aligned}
\mathrm{Rem}(p; F, G) &= \int_{G^{-1}(p)}^{F^{-1}(p)} \big(G(x) - F(x)\big)\mathrm{d}x + \int_{G^{-1}(p)}^{F^{-1}(p)} \big(F(x) - p\big)\mathrm{d}x \\
&= \int_{G^{-1}(p)}^{F^{-1}(p)} \big(G(x) - F(x)\big)\mathrm{d}x - \mathrm{Rem}(p; G, F) \\
&\leq \int_{G^{-1}(p)}^{F^{-1}(p)} \big(G(x) - F(x)\big)\mathrm{d}x,
\end{aligned}$$

with the last inequality holding because $\mathrm{Rem}(p; G, F) \geq 0$. This establishes upper bound (3.4). Bound (3.5) follows immediately. $\qquad\square$

Proof of Theorem 5.1. By Lemma A.2, we have the equation

$$\int_0^p F^{-1}(u)\mathrm{d}u = pF^{-1}(p) - \int_{-\infty}^{F^{-1}(p)} F(x)\mathrm{d}x. \tag{A.12}$$

Since the same equation also holds with G instead of F, taking the difference between the two equations gives

$$\int_p^1 \left(G^{-1}(u) - F^{-1}(u)\right) \mathrm{d}u = p\left(G^{-1}(p) - F^{-1}(p)\right) - \int_{-\infty}^{G^{-1}(p)} G(x)\mathrm{d}x + \int_{-\infty}^{F^{-1}(p)} F(x)\mathrm{d}x. \quad \text{(A.13)}$$

Next we express the penultimate integral on the right-hand side of equation (A.13) as follows:

$$\int_{-\infty}^{G^{-1}(p)} G(x)\mathrm{d}x = \int_{-\infty}^{F^{-1}(p)} G(x)\mathrm{d}x + \int_{F^{-1}(p)}^{G^{-1}(p)} G(x)\mathrm{d}x, \quad \text{(A.14)}$$

where the second integral on the right-hand side is equal to $-\int_{G^{-1}(p)}^{F^{-1}(p)} G(x)\mathrm{d}x$ when $F^{-1}(p) > G^{-1}(p)$. Combining equations (A.13) and (A.14), we arrive at equation (5.2) with the remainder term

$$\mathrm{Rem}(p; F, G) = p\left(G^{-1}(p) - F^{-1}(p)\right) - \int_{F^{-1}(p)}^{G^{-1}(p)} G(x)\mathrm{d}x$$
$$= -p\left(F^{-1}(p) - G^{-1}(p)\right) + \int_{G^{-1}(p)}^{F^{-1}(p)} G(x)\mathrm{d}x,$$

that can be concisely written as the integral on the right-hand side of equation (3.3). This completes the proof of Theorem 5.1. $\qquad \square$

Proof of Theorem 6.1. We start with the equation

$$\int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u = p_2 F^{-1}(p_2) - p_1 F^{-1}(p_1) - \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F(x)\mathrm{d}x, \quad \text{(A.15)}$$

which can be verified in many different ways, and whose validity does not require any condition on the cdf F. For example, we can prove equation (A.15) as follows.

Since $F^{-1}(p_1)$ and $F^{-1}(p_2)$ are finite, the integral $\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F(x)\mathrm{d}x$ is finite. With X denoting a random variable whose cdf is F, and also noting that X is equal in distribution

to $F^{-1}(U)$, where U is a uniform on the interval $[0,1]$ random variable, we have the equations

$$\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F(x)\mathrm{d}x = \mathbb{E}\left(\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \mathbb{1}\{X \le x\}\mathrm{d}x \right)$$

$$= \mathbb{E}\left(\left(F^{-1}(p_2) - F^{-1}(p_1) \vee X\right)^+ \right)$$

$$= \mathbb{E}\left(\left(F^{-1}(p_2) - F^{-1}(p_1) \vee F^{-1}(U)\right)^+ \right)$$

$$= \int_0^1 \left(F^{-1}(p_2) - F^{-1}(p_1 \vee u)\right)^+ \mathrm{d}u.$$

Splitting the integral $\int_0^1$ on the right-hand side into the sum of $\int_0^{p_1}$, $\int_{p_1}^{p_2}$ and $\int_{p_2}^1$, we obtain

$$\int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F(x)\mathrm{d}x = \int_0^{p_1} \left(F^{-1}(p_2) - F^{-1}(p_1)\right)\mathrm{d}u + \int_{p_1}^{p_2} \left(F^{-1}(p_2) - F^{-1}(u)\right)\mathrm{d}u$$

$$= p_1\left(F^{-1}(p_2) - F^{-1}(p_1)\right) + (p_2 - p_1)F^{-1}(p_2) - \int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u$$

$$= p_2 F^{-1}(p_2) - p_1 F^{-1}(p_1) - \int_{p_1}^{p_2} F^{-1}(u)\mathrm{d}u.$$

Equation (A.15) follows. Of course, the same equation also holds with G instead of F. Hence,

$$\int_{p_1}^{p_2} \left(G^{-1}(u) - F^{-1}(u)\right)\mathrm{d}u = p_2\left(G^{-1}(p_2) - F^{-1}(p_2)\right) - p_1\left(G^{-1}(p_1) - F^{-1}(p_1)\right)$$

$$- \left(\int_{G^{-1}(p_1)}^{G^{-1}(p_2)} G(x)\mathrm{d}x - \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} F(x)\mathrm{d}x \right),$$

which can be rewritten as

$$\int_{p_1}^{p_2} \left(G^{-1}(u) - F^{-1}(u)\right)\mathrm{d}u = p_2\left(G^{-1}(p_2) - F^{-1}(p_2)\right) - p_1\left(G^{-1}(p_1) - F^{-1}(p_1)\right)$$

$$- \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \left(G(x) - F(x)\right)\mathrm{d}x$$

$$- \left(\int_{G^{-1}(p_1)}^{G^{-1}(p_2)} G(x)\mathrm{d}x - \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} G(x)\mathrm{d}x \right).$$

Since

$$\int_{G^{-1}(p_1)}^{G^{-1}(p_2)} G(x)\mathrm{d}x - \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} G(x)\mathrm{d}x = \int_{G^{-1}(p_1)}^{F^{-1}(p_1)} G(x)\mathrm{d}x - \int_{G^{-1}(p_2)}^{F^{-1}(p_2)} G(x)\mathrm{d}x,$$

we have

$$\int_{p_1}^{p_2} \left(G^{-1}(u) - F^{-1}(u) \right) \mathrm{d}u = p_2 \left(G^{-1}(p_2) - F^{-1}(p_2) \right) - p_1 \left(G^{-1}(p_1) - F^{-1}(p_1) \right)$$

$$- \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \left(G(x) - F(x) \right) \mathrm{d}x$$

$$- \int_{G^{-1}(p_1)}^{F^{-1}(p_1)} G(x) \mathrm{d}x + \int_{G^{-1}(p_2)}^{F^{-1}(p_2)} G(x) \mathrm{d}x.$$

Combining the terms on the right-hand side of the latter equation, we obtain

$$\int_{p_1}^{p_2} \left(G^{-1}(u) - F^{-1}(u) \right) \mathrm{d}u = - \int_{F^{-1}(p_1)}^{F^{-1}(p_2)} \left(G(x) - F(x) \right) \mathrm{d}x$$

$$+ \int_{G^{-1}(p_1)}^{F^{-1}(p_1)} \left(p_1 - G(x) \right) \mathrm{d}x - \int_{G^{-1}(p_2)}^{F^{-1}(p_2)} \left(p_2 - G(x) \right) \mathrm{d}x.$$

This establishes equation (6.2) and completes the proof of Theorem 6.1. $\qquad\square$

A.3 Asymptotic results

Lemma A.3. *Let $F \in \mathcal{F}_2^{\mu \neq 0}$, and let the quantile function F^{-1} be continuous at the point $p \in (0,1)$. Then*

$$\sqrt{n} \left(\mathrm{LC}_{n,\mathrm{SRS}}(p) - \mathrm{LC}(p) \right) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_{i,\mathrm{LC}}(p) + o_{\mathbb{P}}(1), \qquad (\mathrm{A.16})$$

where

$$Y_{i,\mathrm{LC}}(p) = \frac{1}{\mu} \int_{-\infty}^{F^{-1}(p)} \left(\mathbb{1}\{X_i \leq x\} - F(x) \right) \mathrm{d}x + \frac{\mathrm{LC}(p)}{\mu} \left(X_i - \mu \right).$$

Proof. We start with the generic equation

$$\frac{\theta_n}{\mu_n} - \frac{\theta}{\mu} = \underbrace{\frac{1}{\mu}(\theta_n - \theta) - \frac{\theta}{\mu^2}(\mu_n - \mu)}_{\text{main term}} + \underbrace{\frac{\theta}{\mu^2 \mu_n}(\mu_n - \mu)^2 - \frac{1}{\mu \mu_n}(\theta_n - \theta)(\mu_n - \mu)}_{\text{remainder term}}, \qquad (\mathrm{A.17})$$

which we next use with the following quantities:

- μ is the mean of F,

- θ is the lower-layer integral defined by equation (2.3),

74

- μ_n is the arithmetic mean $\mu_{n,\mathrm{SRS}}$ given by equation (5.8),

- θ_n is the estimator $\theta_{n,\mathrm{SRS}} = \int_0^p F_{n,\mathrm{SRS}}^{-1}(u)\mathrm{d}u$ of θ.

By the classical central limit theorem, we have

$$\sqrt{n}(\mu_{n,\mathrm{SRS}} - \mu) = O_{\mathbb{P}}(1).$$

Statement (7.7) implies

$$\sqrt{n}(\theta_{n,\mathrm{SRS}} - \theta) = O_{\mathbb{P}}(1).$$

Consequently, the remainder term on the right-hand side of equation (A.17) is of the order $o_{\mathbb{P}}(1/\sqrt{n})$. Hence,

$$
\begin{aligned}
\sqrt{n}\left(\mathrm{LC}_{n,\mathrm{SRS}}(p) - \mathrm{LC}(p)\right) &= \sqrt{n}\left(\frac{\theta_{n,\mathrm{SRS}}}{\mu_{n,\mathrm{SRS}}} - \frac{\theta}{\mu}\right) \\
&= \frac{1}{\mu}\sqrt{n}(\theta_{n,\mathrm{SRS}} - \theta) - \frac{\theta}{\mu^2}\sqrt{n}(\mu_{n,\mathrm{SRS}} - \mu) + o_{\mathbb{P}}(1). \qquad (\mathrm{A.18})
\end{aligned}
$$

By equation (5.11) with $F_{n,\mathrm{SRS}}$ instead of F_n, we have

$$\sqrt{n}(\theta_{n,\mathrm{SRS}} - \theta) = \sqrt{n}\int_{-\infty}^{F^{-1}(p)}\left(F(x) - F_{n,\mathrm{SRS}}(x)\right)\mathrm{d}x + o_{\mathbb{P}}(1). \qquad (\mathrm{A.19})$$

Combining equations (A.18) and (A.19), we obtain

$$\sqrt{n}\left(\mathrm{LC}_{n,\mathrm{SRS}}(p) - \mathrm{LC}(p)\right) = -\frac{1}{\mu}\sqrt{n}\int_{-\infty}^{F^{-1}(p)}\left(F_{n,\mathrm{SRS}}(x) - F(x)\right)\mathrm{d}x - \frac{\theta}{\mu^2}\sqrt{n}(\mu_{n,\mathrm{SRS}} - \mu) + o_{\mathbb{P}}(1),$$

which implies statement (A.16) and concludes the proof of Lemma A.3. $\qquad \square$

Lemma A.4. *Let $F \in \mathcal{F}_2^{\mu \neq 0}$, and let the quantile function F^{-1} be continuous at the points $p \in (0,1)$ and $1 - p$. Then*

$$\sqrt{n}\left(\mathrm{GC}_{n,\mathrm{SRS}}(p) - \mathrm{GC}(p)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Y_{i,\mathrm{LC}}(1-p) + Y_{i,\mathrm{LC}}(p)\right) + o_{\mathbb{P}}(1),$$

where

$$Y_{i,\mathrm{LC}}(u) = \frac{1}{\mu}\int_{-\infty}^{F^{-1}(u)}\left(\mathbb{1}\{X_i \leq x\} - F(x)\right)\mathrm{d}x + \frac{\mathrm{LC}(u)}{\mu}(X_i - \mu).$$

Proof. Note that

$$\mathrm{GC}(p) = 1 - \frac{1}{\mu}\left(\int_0^{1-p} F^{-1}(u)\mathrm{d}u + \int_0^p F^{-1}(u)\mathrm{d}u\right).$$

$$= 1 - \mathrm{LC}(1-p) - \mathrm{LC}(p).$$

Hence,

$$\sqrt{n}\left(\mathrm{GC}_{n,\mathrm{SRS}}(p) - \mathrm{GC}(p)\right) = -\sqrt{n}\left(\mathrm{LC}_{n,\mathrm{SRS}}(1-p) - \mathrm{LC}(1-p)\right) - \sqrt{n}\left(\mathrm{LC}_{n,\mathrm{SRS}}(p) - \mathrm{LC}(p)\right),$$

and with the help of Lemma A.3, we obtain

$$\sqrt{n}\left(\mathrm{GC}_{n,\mathrm{SRS}}(p) - \mathrm{GC}(p)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(Y_{i,\mathrm{LC}}(1-p) + Y_{i,\mathrm{LC}}(p)\right) + o_{\mathbb{P}}(1),$$

which concludes the proof of Lemma A.4. □

A.4   Proof of Theorem 9.1

We begin the proof by splitting the integral $L_{w,a,b}(F)$ into two parts according to whether $w(u) < 0$ or $w(u) \geq 0$. This gives us the equation

$$\int_a^b F^{-1}(u)w(u)\mathrm{d}u = \int_a^b F^{-1}(u)\mathbb{1}\{w(u) < 0\}w(u)\mathrm{d}u + \int_a^b F^{-1}(u)\mathbb{1}\{w(u) \geq 0\}w(u)\mathrm{d}u.$$
(A.20)

We shall next analyze the two integrals on the right-hand side separately, showing that they are equal to the respective integrals on the right-hand side of equation (9.3).

Pertaining to the first integral on the right-hand side of equation (A.20), since w is non-decreasing and right-continuous, we have $w(u) < 0$ if and only if $u < w^{-1}(0)$, and so

$$\int_a^b F^{-1}(u)\mathbb{1}\{w(u) < 0\}w(u)\mathrm{d}u = \int_a^b F^{-1}(u)\mathbb{1}\{u < w^{-1}(0)\}w(u)\mathrm{d}u.$$

Furthermore, since $w(u) < 0$, we have

$$w(u) = -\int_{w(u)}^{0} \mathrm{d}x$$

$$= -\int_{-\infty}^{0} \mathbb{1}\{u < w^{-1}(x)\}\mathrm{d}x,$$

where the last equation holds because we have $w(u) < x$ if and only if $u < w^{-1}(x)$, due to our assumption that w is non-decreasing and right-continuous. Consequently,

$$\int_{a}^{b} F^{-1}(u)\mathbb{1}\{w(u) < 0\}w(u)\mathrm{d}u = -\int_{-\infty}^{0}\left(\int_{a}^{b} F^{-1}(u)\mathbb{1}\{u < w^{-1}(0) \wedge w^{-1}(x)\}\mathrm{d}u\right)\mathrm{d}x$$

$$= -\int_{-\infty}^{0}\left(\int_{a}^{b} F^{-1}(u)\mathbb{1}\{u < w^{-1}(x)\}\mathrm{d}u\right)\mathrm{d}x, \qquad \text{(A.21)}$$

where right-most equation holds because w^{-1} is non-decreasing. The double integral on the right-hand side of equation (A.21) is zero when $w^{-1}(x) \le a$. Hence, continuing with equation (A.21) for only those x that satisfy $a < w^{-1}(x)$, we have

$$\int_{a}^{b} F^{-1}(u)\mathbb{1}\{w(u) < 0\}w(u)\mathrm{d}u = -\int_{-\infty}^{0} \mathbb{1}\{a < w^{-1}(x)\}\left(\int_{a}^{b} F^{-1}(u)\mathbb{1}\{u < w^{-1}(x)\}\mathrm{d}u\right)\mathrm{d}x$$

$$= -\int_{-\infty}^{0} \mathbb{1}\{a < w^{-1}(x)\}\left(\int_{a}^{b \wedge w^{-1}(x)} F^{-1}(u)\mathrm{d}u\right)\mathrm{d}x. \text{ (A.22)}$$

Since w is non-decreasing and right-continuous, we have that $a < w^{-1}(x)$ is equivalent to $w(a) < x$, and so equation (A.22) gives the first integral on the right-hand side of equation (9.3).

Analogously we tackle the second integral on the right-hand side of equation (A.20). First, since $w(u) \ge 0$ if and only if $u \ge w^{-1}(0)$, we have

$$\int_{a}^{b} F^{-1}(u)\mathbb{1}\{w(u) \ge 0\}w(u)\mathrm{d}u = \int_{a}^{b} F^{-1}(u)\mathbb{1}\{u \ge w^{-1}(0)\}w(u)\mathrm{d}u.$$

Furthermore, since $w(u) \geq 0$, we have

$$w(u) = \int_0^{w(u)} \mathrm{d}x$$

$$= \int_0^\infty \mathbb{1}\{u \geq w^{-1}(x)\}\mathrm{d}x,$$

where the last equation holds because $w(u) \geq x$ if and only if $u \geq w^{-1}(x)$. Consequently,

$$\int_a^b F^{-1}(u)\mathbb{1}\{w(u) \geq 0\}w(u)\mathrm{d}u = \int_0^\infty \left(\int_a^b F^{-1}(u)\mathbb{1}\{u \geq w^{-1}(0) \vee w^{-1}(x)\}\mathrm{d}u \right)\mathrm{d}x$$

$$= \int_0^\infty \left(\int_a^b F^{-1}(u)\mathbb{1}\{u \geq w^{-1}(x)\}\mathrm{d}u \right)\mathrm{d}x, \qquad (A.23)$$

where right-most equation holds because w^{-1} is non-decreasing. The double integral on the right-hand side of equation (A.23) is zero when $w^{-1}(x) > b$. Hence, continuing with equation (A.23) for only those x that satisfy $w^{-1}(x) \leq b$, we have

$$\int_a^b F^{-1}(u)\mathbb{1}\{w(u) \geq 0\}w(u)\mathrm{d}u = \int_0^\infty \mathbb{1}\{w^{-1}(x) \leq b\}\left(\int_a^b F^{-1}(u)\mathbb{1}\{u \geq w^{-1}(x)\}\mathrm{d}u \right)\mathrm{d}x$$

$$= \int_0^\infty \mathbb{1}\{w^{-1}(x) \leq b\}\left(\int_{a \vee w^{-1}(x)}^b F^{-1}(u)\mathrm{d}u \right)\mathrm{d}x. \qquad (A.24)$$

Since w is non-decreasing and right-continuous, we have that $w^{-1}(x) \leq b$ is equivalent to $x \leq w(b)$, and so equation (A.24) gives the second integral on the right-hand side of equation (9.3). This concludes the proof of Theorem 9.1.

A.5   Proof of Theorem 9.2.

Let condition (C_w) be satisfied with some $K \geq 3$. Equation (9.7) is equivalent to

$$w(u) = \mathbb{1}\{0 \leq t < a_1\}\Big(w_{11}(u) - w_{12}(u)\Big)$$

$$+ \sum_{k=2}^{K-1} \mathbb{1}\{a_{k-1} \leq t < a_k\}\left(\left\{w_{k1}(u) + \sum_{i=2}^k c_i\right\} - \left\{w_{k2}(u) + \sum_{i=2}^k c_i\right\} \right)$$

$$+ \mathbb{1}\{a_{K-1} \leq t < 1\}\Big(w_{K1}(u) - w_{K2}(u)\Big),$$

where the constants $c_2, \ldots, c_{K-1}$ can be any real numbers, but we choose them so that the inequalities

$$w_{k1}(a_{k-1}) + \sum_{i=2}^{k} c_i \geq w_{(k-1)1}(a_{k-1}-) + \sum_{i=2}^{k-1} c_i \quad \left(\Longleftrightarrow w_{k1}(a_{k-1}) + c_k \geq w_{(k-1)1}(a_{k-1}-) \right)$$

and

$$w_{k2}(a_{k-1}) + \sum_{i=2}^{k} c_i \geq w_{(k-1)2}(a_{k-1}-) + \sum_{i=2}^{k-1} c_i \quad \left(\Longleftrightarrow w_{k2}(a_{k-1}) + c_k \geq w_{(k-1)2}(a_{k-1}-) \right)$$

would hold for every $k = 2, \ldots, K - 1$. The functions

$$w_{11}^*(u) = \mathbb{1}\{0 \leq u < a_1\} w_{11}(u) + \sum_{k=2}^{K-1} \mathbb{1}\{a_{k-1} \leq u < a_k\} \left\{ w_{k1}(u) + \sum_{i=2}^{k} c_i \right\}$$

and

$$w_{12}^*(u) = \mathbb{1}\{0 \leq u < a_1\} w_{12}(u) + \sum_{k=2}^{K-1} \mathbb{1}\{a_{k-1} \leq u < a_k\} \left\{ w_{k2}(u) + \sum_{i=2}^{k} c_i \right\}$$

are non-decreasing and right-continuous on the interval $[0, a_{K-1})$. Furthermore, the functions

$$w_{21}^*(u) = w_{K1}(u) \quad \text{and} \quad w_{22}^*(u) = w_{K2}(u)$$

are non-decreasing and right-continuous on the interval $[a_{K-1}, 1)$. We obviously have the equation

$$w(u) = \mathbb{1}\{0 \leq u < a_{K-1}\} \left(w_{11}^*(u) - w_{12}^*(u) \right) + \mathbb{1}\{a_{K-1} \leq u < 1\} \left(w_{21}^*(u) - w_{22}^*(u) \right),$$

which establishes representation (9.8) and completes the proof of Theorem 9.2.

A.6 Proofs of cursory notes

In this appendix we present proofs of some of the cursory notes that we have made while discussing main results.

A.6.1   Proof of equation (1.6)

Let U denote a uniform on $[0, 1]$ random variable. Since X and Y have the same distributions as $F^{-1}(U)$ and $G^{-1}(U)$, respectively, we therefore have

$$\int_0^1 \big(G^{-1}(u) - F^{-1}(u)\big)\mathrm{d}u = \mathbb{E}\big(G^{-1}(U)\big) - \mathbb{E}\big(F^{-1}(U)\big)$$
$$= \mathbb{E}(Y) - \mathbb{E}(X). \tag{A.25}$$

Furthermore, using Fubini's theorem, we have

$$\int_{-\infty}^{\infty} \big(F(x) - G(x)\big)\mathrm{d}x = \mathbb{E}\left(\int_{-\infty}^{\infty} \big(\mathbb{1}\{X \le x\} - \mathbb{1}\{Y \le x\}\big)\mathrm{d}x\right)$$
$$= \mathbb{E}\left(\int_{-\infty}^{\infty} \big(\mathbb{1}\{Y > x\} - \mathbb{1}\{X > x\}\big)\mathrm{d}x\right)$$
$$= \mathbb{E}(Y - X), \tag{A.26}$$

where the right-most equation holds because the integral inside the expectation is equal to $Y - X$. Since the right-hand sides of equations (A.25) and (A.26) are equal, we have equation (1.6).

A.6.2   On the validity of statements (4.4) and (4.7)-(4.8)

Using the definition of the empirical quantile function $H_{n,\mathrm{SRS}}^{-1}$, statement (4.4) is equivalent to

$$Z_{\lceil n/2 \rceil : n} \overset{\mathbb{P}}{\nrightarrow} \frac{1}{2}.$$

To prove it, we want to find $\epsilon > 0$ such that

$$\lim_{n \to \infty} \mathbb{P}\left(\left|Z_{\lceil n/2 \rceil : n} - \frac{1}{2}\right| > \epsilon\right) > 0. \tag{A.27}$$

We set $\epsilon = 1/2$ and note that statement (A.27) follows if we establish at least one of statements (4.7) and (4.8), but for completeness of the argument, we shall next establish both of them. Using a well-known formula (e.g., David and Nagaraja, 2003, eq. (2.1.3)) for

80

the cdf of order statistics, we have

$$\mathbb{P}\left(Z_{\lceil n/2 \rceil:n} > \frac{3}{2}\right) = 1 - \sum_{j=\lceil n/2 \rceil}^{n} \mathbb{P}\left(Z_{j:n} \leq \frac{3}{2}\right)$$

$$= 1 - \sum_{j=\lceil n/2 \rceil}^{n} \binom{n}{j} \left(F_Z\left(\frac{3}{2}\right)\right)^j \left(1 - F_Z\left(\frac{3}{2}\right)\right)^{n-j}$$

$$= 1 - \sum_{j=\lceil n/2 \rceil}^{n} \binom{n}{j} \left(\frac{1}{2}\right)^j \left(\frac{1}{2}\right)^{n-j}$$

$$= 1 - \left(\frac{1}{2}\right)^n \sum_{j=\lceil n/2 \rceil}^{n} \binom{n}{j}. \tag{A.28}$$

Similarly,

$$\mathbb{P}\left(Z_{\lceil n/2 \rceil:n} < \frac{1}{2}\right) = \sum_{j=\lceil n/2 \rceil}^{n} \mathbb{P}\left(Z_{j:n} < \frac{1}{2}\right)$$

$$= \sum_{j=\lceil n/2 \rceil}^{n} \binom{n}{j} \left(F_Z\left(\frac{1}{2}\right)\right)^j \left(1 - F_Z\left(\frac{1}{2}\right)\right)^{n-j}$$

$$= \sum_{j=\lceil n/2 \rceil}^{n} \binom{n}{j} \left(\frac{1}{2}\right)^j \left(\frac{1}{2}\right)^{n-j}$$

$$= \left(\frac{1}{2}\right)^n \sum_{j=\lceil n/2 \rceil}^{n} \binom{n}{j}. \tag{A.29}$$

We next rearrange the sums of binomials on the right-hand sides of equations (A.28) and (A.29). When n is even, $\lceil n/2 \rceil = n/2$ and 2^n can be expressed in $n+1$ (which is odd) terms

$$2^n = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{\frac{n}{2}-1} + \binom{n}{\frac{n}{2}} + \binom{n}{\frac{n}{2}+1} + \cdots + \binom{n}{n-1} + \binom{n}{n}$$

$$= \binom{n}{\frac{n}{2}} + 2\left(\binom{n}{\frac{n}{2}+1} + \cdots + \binom{n}{n-1} + \binom{n}{n}\right).$$

Hence, we have

$$\sum_{j=\lceil n/2 \rceil}^{n} \binom{n}{j} = \binom{n}{\lceil \frac{n}{2} \rceil} + \binom{n}{\lceil \frac{n}{2} \rceil + 1} + \cdots + \binom{n}{n-1} + \binom{n}{n}$$

$$= \binom{n}{\frac{n}{2}} + \binom{n}{\frac{n}{2} + 1} + \cdots + \binom{n}{n-1} + \binom{n}{n}$$

$$= \frac{1}{2}\left(2^n + \binom{n}{\frac{n}{2}}\right)$$

$$= 2^{n-1} + \frac{1}{2}\binom{n}{\frac{n}{2}}.$$

When n is odd, 2^n can be expressed in $n + 1$ (which is even) terms as follows:

$$2^n = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{\lceil \frac{n}{2} \rceil - 1} + \binom{n}{\lceil \frac{n}{2} \rceil} + \cdots + \binom{n}{n-1} + \binom{n}{n}$$

$$= 2\left(\binom{n}{\lceil \frac{n}{2} \rceil} + \cdots + \binom{n}{n-1} + \binom{n}{n}\right).$$

Hence, we have

$$\sum_{j=\lceil n/2 \rceil}^{n} \binom{n}{j} = \binom{n}{\lceil \frac{n}{2} \rceil} + \binom{n}{\lceil \frac{n}{2} \rceil + 1} + \cdots + \binom{n}{n-1} + \binom{n}{n}$$

$$= \frac{1}{2} 2^n$$

$$= 2^{n-1}.$$

Going back to the original expressions (A.28) and (A.29) of probabilities, we have

$$\mathbb{P}\left(Z_{\lceil n/2 \rceil:n} > \frac{3}{2}\right) = 1 - \left(\frac{1}{2}\right)^n \sum_{j=\lceil n/2 \rceil}^{n} \binom{n}{j}$$

$$= \begin{cases} \frac{1}{2} - \frac{1}{2^{n+1}}\binom{n}{n/2} & \text{when } n \text{ is even;} \\ \frac{1}{2} & \text{when } n \text{ is odd;} \end{cases}$$

and

$$\mathbb{P}\left(Z_{\lceil n/2 \rceil :n} < \frac{1}{2}\right) = \left(\frac{1}{2}\right)^n \sum_{j=\lceil n/2 \rceil}^{n} \binom{n}{j}$$

$$= \begin{cases} \frac{1}{2} + \frac{1}{2^{n+1}}\binom{n}{n/2} & \text{when } n \text{ is even;} \\ \frac{1}{2} & \text{when } n \text{ is odd.} \end{cases}$$

As $n \to \infty$, the term $\frac{1}{2^{n+1}}\binom{n}{n/2}$ vanishes. This establishes statements (4.7) and (4.8), and concludes the proof of statement (4.4).

A.6.3 Derivation of expression (4.11)

Starting with equation (4.10), we need to calculate the first two moments of the random variable $\left(H^{-1}(U) - (1-a)\right)^{+}$. From Figure 4.2 we see that the random variable $H^{-1}(U) - (1-a)$ is (strictly) positive if and only if $U > 1/2$, and under this condition, $H^{-1}(U)$ is equal to $2(1-a)U + 2a$. Hence,

$$\left(H^{-1}(U) - (1-a)\right)^{+} = \left(2(1-a)U + 2a\right)\mathbb{1}\{U > 1/2\}. \tag{A.30}$$

By calculating the first two moments of the random variable on the right-hand side of equation (A.30), we arrive at expression (4.11).

A.6.4 Proof of equation (9.1)

We start with the equations

$$K_w(F(x)) = \int_0^{F(x)} w(u)\mathrm{d}u$$

$$= \int_0^1 \mathbb{1}\{u \le F(x)\}w(u)\mathrm{d}u$$

$$= \int_0^1 \mathbb{1}\{F^{-1}(u) \le x\}w(u)\mathrm{d}u.$$

Analogous equations hold for $K_w\big(G(x)\big)$. Hence,

$$
\begin{aligned}
\int_{\mathbb{R}}\Big(K_w\big(F(x)\big)-K_w\big(G(x)\big)\Big)\mathrm{d}x &= \int_{\mathbb{R}}\bigg(\int_0^1\Big(\mathbb{1}\{F^{-1}(u)\le x\}-\mathbb{1}\{G^{-1}(u)\le x\}\Big)w(u)\mathrm{d}u\bigg)\mathrm{d}x \\
&= \int_0^1\bigg(\int_{\mathbb{R}}\Big(\mathbb{1}\{F^{-1}(u)\le x\}-\mathbb{1}\{G^{-1}(u)\le x\}\Big)\mathrm{d}x\bigg)w(u)\mathrm{d}u \\
&= \int_0^1\Big(G^{-1}(u)-F^{-1}(u)\Big)w(u)\mathrm{d}u,
\end{aligned}
$$

where the right-most equation holds because the inner integral is equal to $G^{-1}(u)-F^{-1}(u)$. This completes the proof of equation (9.1).

A.6.5 Proof of equation (9.2)

Since w_1 is non-decreasing and right-continuous, and such that $w_1(0)\ge 0$, we have

$$
\begin{aligned}
\int_0^1 F^{-1}(u)w_1(u)\mathrm{d}u &= \int_0^1 F^{-1}(u)\bigg(\int_0^\infty \mathbb{1}\{x\le w_1(u)\}\mathrm{d}x\bigg)\mathrm{d}u \\
&= \int_0^1 F^{-1}(u)\bigg(\int_0^\infty \mathbb{1}\{w_1^{-1}(x)\le u\}\mathrm{d}x\bigg)\mathrm{d}u \\
&= \int_0^\infty\bigg(\int_0^1 F^{-1}(u)\mathbb{1}\{w_1^{-1}(x)\le u\}\mathrm{d}u\bigg)\mathrm{d}x \\
&= \int_0^\infty\bigg(\int_{w_1^{-1}(x)}^1 F^{-1}(u)\mathrm{d}u\bigg)\mathrm{d}x.
\end{aligned}
$$

Analogous equations hold for w_2. This establishes equation (9.2).