A CENTRAL LIMIT THEOREM AND HYPOTHESES TESTING FOR RISK AVERSE STOCHASTIC PROGRAMS∗

VINCENT GUIGUES†, VOLKER KRÄTSCHMER‡, AND ALEXANDER SHAPIRO§

Abstract. We study statistical properties of the optimal value of the Sample Average Approximation of risk averse stochastic problems. Central Limit Theorem type results are derived for the optimal value when the stochastic program is expressed in terms of a law invariant coherent risk measure having a discrete Kusuoka representation. The obtained results are applied to hypotheses testing problems aiming at comparing the optimal values of several risk averse convex stochastic programs on the basis of samples of the underlying random vectors. We also consider non-asymptotic tests based on confidence intervals on the optimal values of the stochastic programs obtained using the Robust Stochastic Approximation algorithm. Numerical simulations show how to use our developments to choose among different distributions and show the superiority of the asymptotic tests on a class of risk averse stochastic programs.

Key words. Stochastic optimization, Sample Average Approximation, hypotheses testing, coherent risk measures, statistical inference, Central Limit Theorem.

AMS subject classifications. 90C15, 90C90, 90C30

1. Introduction. Consider the following risk averse stochastic program

\[(1) \min_{x \in \mathcal{X}} \{g(x) := \mathcal{R}(G_x)\}.\]

Here \(\mathcal{X}\) is a nonempty compact subset of \(\mathbb{R}^m\), \(G_x\) is a random variable depending on \(x \in \mathcal{X}\) and \(\mathcal{R}\) is a risk measure. We assume that \(G_x\) is given in the form \(G_x(\omega) = G(x, \xi(\omega))\), where \(G : \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}\) and \(\xi : \Omega \to \mathbb{R}^d\) is a random vector defined on a probability space \((\Omega, \mathcal{F}, P)\) whose distribution is supported on set \(\Xi \subset \mathbb{R}^d\). We assume that the functional \(\mathcal{R}\), defined on a space of random variables, is law invariant (we will give precise definitions in Section 2).

Let \(\xi_j = \xi_j(\omega), j = 1, ..., N,\) be an i.i.d sample of the random vector \(\xi\) defined on the same probability space. Then the respective sample estimate of \(g(x)\), denoted \(\hat{g}_N(x)\), is obtained by replacing the “true” distribution of the random vector \(\xi\) with its empirical estimate. Consequently the true optimization problem (1) is approximated by the problem

\[(2) \min_{x \in \mathcal{X}} \hat{g}_N(x),\]

referred to as the Sample Average Approximation (SAA) problem. Note that \(\hat{g}_N(x) = \hat{g}_N(x, \omega)\) is a random function, sometimes we suppress dependence on \(\omega\) in the notation. In particular if \(\mathcal{R}\) is the expectation operator, i.e., \(g(x) = \mathbb{E}[G_x]\), then \(\hat{g}_N(x) = N^{-1} \sum_{j=1}^N G(x, \xi_j)\).

∗Submitted to the editors DATE.

Funding: Research of this author was partly supported by an FGV grant, CNPq grant 307287/2013-0, FAPERJ grants E-26/110.313/2014 and E-26/201.599/2014. Research of third author was partly supported by NSF grant 1633196 and DARPA EQUiPS program, grant SNL 01450769.

†School of Applied Mathematics, Fundação Getulio Vargas, 190 Praia de Botafogo, Rio de Janeiro, Brazil, (vguigues@fgv.br).
‡Faculty of Mathematics, University of Duisburg-Essen, (volker.kraetschmer@uni-due.de).
§School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA, (ashapiro@isye.gatech.edu).
We denote by $\vartheta_v$ and $\hat{\vartheta}_N$ the optimal values of problems (1) and (2), respectively, and study statistical properties of $\hat{\vartheta}_N$. The random sample can be given by collected data or can be generated by Monte Carlo sampling techniques in the goal of solving the true problem by the SAA method. Although conceptually different, both situations lead to the same statistical inference.

The statistical analysis allows us to address the following question of asymptotic tests of hypotheses. Suppose that we are given $V \geq 2$ optimization problems of the form (1) with $\xi$, $G$, and $X$ respectively replaced by $\xi_v$, $G_v$, and $X_v$ for problem $v \in \{1, \ldots, V\}$. On the basis of samples $\xi_1^v, \ldots, \xi_N^v$, of size $N$, of $\xi_v$, $v = 1, \ldots, V$, and denoting by $\hat{\vartheta}_v^v$ the optimal value of problem $v$, we study statistical tests of the null hypotheses

\begin{align*}
(a) \quad & H_0 : \vartheta_1^1 = \vartheta_2^1 = \ldots = \vartheta_V^1, \\
(b) \quad & H_0 : \vartheta_p^v \leq \vartheta_q^v \text{ for } p \text{ fixed and all } 1 \leq q \leq V, \\
(c) \quad & H_0 : \vartheta_1^v \leq \vartheta_2^v \leq \ldots \leq \vartheta_V^v,
\end{align*}

against the corresponding unrestricted alternatives. As a special case, if the feasibility sets of the $V$ optimizations problems are singletons, say $\{x^v_n\}$ for problem $v$, the above tests aim at comparing the risks $R(G_{x_1^v}), \ldots, R(G_{x_V^v})$. These tests are useful when we want to choose among $V$ candidate solutions $x_1^v, \ldots, x_V^v$ of problem (1) for the one with the smallest risk measure value, using risk measure $R$ to rank the distributions $G_{x_v^v}, v = 1, \ldots, V$, e.g., to decide about the preference of a set of assets over another. In this situation, if the risk measure $R$ is polyhedral [5] or extended polyhedral [9] then it can be expressed as the optimal value of a risk-neutral optimization problem and tests on the equality of risk measure values $R(G_{x_1^v}), \ldots, R(G_{x_V^v})$ are of form (3)-(a).

Setting $\theta := (\vartheta_1^1, \ldots, \vartheta_V^V)$, we also consider the following extension of tests (3):

\begin{equation}
H_0 : \theta \in \Theta_0 \text{ against } H_1 : \theta \in \mathbb{R}^V,
\end{equation}

with $\Theta_0 \subset \mathbb{R}^V$ being a linear space or a convex cone. Tests (3) will also be studied in a nonasymptotic setting.

The paper is organized as follows. In section 2 we specify the type of objective functions in the optimization problem (1). We introduce the class of so called law invariant convex risk measures and point out the specific subclass of risk measures which we use. In the following section 3 we study the asymptotics of the SAA estimator for the optimal value of problem 1. Besides consistency its asymptotic distribution is given in Theorem 2. As a by product we may derive a result on asymptotic distributions of sample estimators for the law invariant convex risk measures that we consider, and it will turn out that it improves already known general results. The proof of Theorem 2 is the subject of section 4. This theorem allows us to derive in section 5.1 asymptotic rejection regions for tests (3) and (4). In section 5.2, we derive nonasymptotic rejection regions for tests (3). This analysis is first conducted in a risk-neutral setting (when $R = E$ is the expectation) and is then extended to risk-averse problems. In particular, in this latter case, we obtain nonasymptotic confidence intervals for the optimal value of (1) for a larger class of risk measures than the class considered in [11] where $R = \text{AVaR}$ (the Average Value-at-Risk, see section 2) was considered. Also, when $R = \text{AVaR}$, our bounds are slightly refined versions of the bounds from [11]. Finally, the last section 6 presents numerical simulations that illustrate our results: we show how to use our developments to choose, using tests (3), among different distributions. We also use these tests to compare the optimal value
of several risk averse stochastic programs. It is shown that the Normal (Gaussian) distribution approximates well the distribution of $\hat{\theta}_N$ already for $N = 20$ and problem sizes (dimension of decision variables) up to $m = 10000$, and that the asymptotic tests yield much smaller type II errors than the considered nonasymptotic tests for small to moderate sample size ($N$ up to $10^3$) and problem size ($m$ up to $500$).

We use the following notation throughout the paper. By $F_Z(z) := P(Z \leq z)$ we denote the cumulative distribution function (cdf) of a random variable $Z : \Omega \to \mathbb{R}$. By $F^{-1}(\alpha) = \inf \{ t : F(t) \geq \alpha \}$ we denote the left-side $\alpha$-quantile of the cdf $F$. By $\Omega_F(\alpha)$ we denote the interval of $\alpha$-quantiles of cdf $F$, i.e.,

$$\Omega_F(\alpha) = [a, b], \text{ where } a := F^{-1}(\alpha), \ b := \sup \{ t : F(t) \leq \alpha \}. \tag{5}$$

By $1_A(\cdot)$ we denote the indicator function of set $A$. For $p \in [1, \infty)$ we consider space $Z := L_p(\Omega, \mathcal{F}, P)$ of random variables $Z : \Omega \to \mathbb{R}$ having finite $p$-th order moments. The dual of space $Z$ is the space $Z^* = L_q(\Omega, \mathcal{F}, P)$, where $q \in (1, \infty]$ is such that $1/p + 1/q = 1$. The notation $Z \succeq Z'$ means that $Z(\omega) \geq Z'(\omega)$ for a.e. $\omega \in \Omega$. By $\delta(a)$ we denote the measure of mass one at $a$.

2. Preliminary discussion. Let us turn over to specify the functional (risk measure) $\mathcal{R}$ in the goal of problem (1). It is defined as a mapping $\mathcal{R} : Z \to \mathbb{R}$ on a linear space $Z$ consisting of random variables on $(\Omega, \mathcal{F}, P)$. Specifically we assume that $Z := L_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$. Note that we consider here real valued risk measures, i.e., we do not allow $\mathcal{R}(Z)$ to have an infinite value. It is said that risk measure $\mathcal{R}(Z)$ is law invariant if it depends only on the distribution of $Z$, i.e., if $Z, Z' \in Z$ and $F_Z = F_{Z'}$, then $\mathcal{R}(Z) = \mathcal{R}(Z')$.

In the influential paper of Artzner et al [2] it was suggested that a “good” risk measure should satisfy the following conditions (axioms).

(i) **Monotonicity**: If $Z, Z' \in Z$ and $Z \succeq Z'$, then $\mathcal{R}(Z) \geq \mathcal{R}(Z')$.

(ii) **Subadditivity**: $\mathcal{R}(Z + Z') \leq \mathcal{R}(Z) + \mathcal{R}(Z')$, for all $Z, Z' \in Z$.

(iii) **Translation Equivariance**: If $a \in \mathbb{R}$ and $Z \in Z$, then $\mathcal{R}(Z + a) = \mathcal{R}(Z) + a$.

(iv) **Positive Homogeneity**: If $t \geq 0$ and $Z \in Z$, then $\mathcal{R}(tZ) = t\mathcal{R}(Z)$.

Conditions (ii) and (iv) imply that $\mathcal{R}$ is convex, i.e.,

$$\mathcal{R}(tZ + (1-t)Z') \leq t\mathcal{R}(Z) + (1-t)\mathcal{R}(Z')$$

for all $Z, Z' \in Z$ and all $t \in [0,1]$.

In [2] such risk measures were called coherent and suggested as a mathematical tool to assess the risks of financial positions. Unless stated otherwise we deal in this paper with law invariant coherent risk measures. Systematic accounts of this class of risk measures can be found in the monographs [27, Chapter 6] and [6, Chapter 4].

An important example of law invariant coherent risk measure is the so called **Average Value-at-Risk** (also called Conditional Value-at-Risk, Expected Shortfall and Expected Tail Loss)

$$\text{AVaR}_\alpha(Z) := \frac{1}{1-\alpha} \int_{\alpha}^{1} F^{-1}_Z(t) \, dt, \ \alpha \in [0,1]. \tag{6}$$

It is naturally defined, and is finite valued, on the space $Z = L_1(\Omega, \mathcal{F}, P)$, and has the following useful representation (cf. [21])

$$\text{AVaR}_\alpha(Z) = \inf_{t \in \mathbb{R}} \left\{ t + (1 - \alpha)^{-1} \mathbb{E}[(Z - t)_+] \right\}. \tag{7}$$
Note that \( \text{AVaR}_0(\cdot) = \mathbb{E}[\cdot] \).

The Average Value-at-Risk \( \text{AVaR}_\alpha(Z) \) is an index to describe the tail behavior of the distribution function \( F_Z \) on the interval \( (F_Z^{-1}(\alpha), \infty) \). If we want to take into account different regions of tail behavior we may choose different levels \( 0 = \alpha_0 < \alpha_1 < \cdots < \alpha_k < 1 \), and then weighting the Average Value-at-Risk at the respective levels. That is, consider

\[
\text{AVaR}_\alpha(Z) := \mathbb{E}[Z] - \sum_{i=1}^k w_i \text{AVaR}_{\alpha_i}(Z),
\]

where \( \mathcal{W} \) is a nonempty subset of \( \Delta_{k+1} := \{ w \in \mathbb{R}_{+}^{k+1} : w_0 + \cdots + w_k = 1 \} \). This is a law invariant coherent risk measure defined on the space \( Z = L_1(\Omega, \mathcal{F}, \mathbb{P}) \). Note that \( \mathcal{R} \) is not changed if \( \mathcal{W} \) is replaced by the topological closure of its convex hull. Note also that the set \( \Delta_{k+1} \) and hence the set \( \mathcal{W} \) are bounded. Therefore if \( \mathcal{W} \) is closed, then it is compact. In view of (7) we can write this risk measure in the following minimax form

\[
\mathcal{R}(Z) = \sup_{w \in \mathcal{W}} \inf_{\tau \in \mathbb{R}^k} \mathbb{E}[\phi(Z, w, \tau)],
\]

where

\[
\phi(z, w, \tau) := w_0 z + \sum_{i=1}^k w_i (\tau_i + (1 - \alpha_i)^{-1}[z - \tau_i]_+).
\]

Assuming that the probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) is nonatomic, any law invariant coherent risk measure \( \mathcal{R} : \mathcal{Z} \to \mathbb{R} \) has the following so-called Kusuoka representation (cf. [10])

\[
\mathcal{R}(Z) = \sup_{\mu \in \mathcal{M}} \int_0^1 \text{AVaR}_{\alpha}(Z) d\mu(\alpha),
\]

where \( \mathcal{M} \) is a set of probability measures on the interval \([0,1]\). We can view risk measure (8) as a discretized version of Kusuoka representation where probability measures \( \mu \in \mathcal{M} \) are restricted to have finite support \( \{\alpha_0, ..., \alpha_k\} \).

3. Asymptotics of the optimization problem. Since a law invariant risk measure \( \mathcal{R} \) can be considered as a function of its cdf \( F(\cdot) = F_Z(\cdot) \), we also write \( \mathcal{R}(F) \) to denote the corresponding value \( \mathcal{R}(Z) \). Let \( Z_1, ..., Z_N \) be an i.i.d sample of \( Z \) and \( \hat{F}_N = N^{-1} \sum_{j=1}^N 1_{[Z_j, \infty)} \) be the corresponding empirical estimate of the cdf \( F \). By replacing \( F \) with its empirical estimate \( \hat{F}_N \), we obtain the estimate \( \mathcal{R}(\hat{F}_N) \) to which we refer as the sample or empirical estimate of \( \mathcal{R}(F) \). We assume that for every \( x \in \mathcal{X} \), the random variable \( G_x \) belongs to the space \( \mathcal{Z} \), and hence \( g(x) = \mathcal{R}(G_x) \) is well defined for every \( x \in \mathcal{X} \). Let \( F_x \) be the cdf of random variable \( G_x, x \in \mathcal{X} \), and \( \hat{F}_{x,N} \) be the empirical cdf associated with the sample \( G(x, \xi_1), ..., G(x, \xi_N) \). Then we can write \( g(x) = \mathcal{R}(F_x) \) and \( \hat{g}_N(x) = \mathcal{R}(\hat{F}_{x,N}) \).

We have the following result about the convergence of the optimal value and optimal solutions of the SAA problem (2) to their counterparts of the “true” problem (1) (cf. [24, Theorem 3.3]).

**Theorem 1.** Let \( \mathcal{R} : \mathcal{Z} \to \mathbb{R} \) be a law invariant risk measure satisfying the axioms of monotonicity, convexity and translation equivariance. Suppose that the set
\( \mathcal{X} \) is nonempty and compact and the following conditions hold: (i) the function \( G_\xi(\omega) \) is random lower semicontinuous, i.e., the epigraphical multifunction \( \omega \mapsto \{(x, t) \in \mathbb{R}^{n+1} : G_\xi(\omega) \leq t \} \) is closed valued and measurable, (ii) for every \( \bar{x} \in \mathbb{R}^n \) there is a neighborhood \( \mathcal{V}_\bar{x} \) of \( \bar{x} \) and a function \( h \in \mathcal{Z} \) such that \( G_\xi(\cdot) \geq h(\cdot) \) for all \( x \in \mathcal{V}_\bar{x} \).

Then the optimal value \( \hat{\vartheta}_N \) of problem (2) converges w.p.1 to the optimal value \( \vartheta_* \) of the “true” problem (1), and the distance from an optimal solution \( \hat{x}_N \) of (2) to the set of optimal solutions of (1) converges w.p.1 to zero as \( N \to \infty \).

We derive first order asymptotics of the SAA optimal value for risk measures \( \mathcal{R} \) of the form (8), i.e. having discretized Kusuoka representation. We assume that the set \( \mathcal{X} \) is nonempty convex compact, \( G(x, \xi) \) is convex in \( x \) for all \( \xi \in \Xi \), and \( \mathbb{E}|G_x| < +\infty \) for all \( x \in \mathcal{X} \). It follows that functions \( g(x) \) and \( \hat{g}_N(x) \) are convex and finite valued, and hence the respective optimization problems (1) and (2) are convex.

Since \( \mathcal{R} \) is of the form (8), the optimal value \( \vartheta_* \) of problem (1) can be written as

\[
\vartheta_* = \inf_{x \in \mathcal{X}} \sup_{w \in \mathfrak{W}} \left\{ w_0 \mathbb{E}[G_x] + \sum_{i=1}^k w_i \text{AVaR}_{\alpha_i}(G_x) \right\}.
\]

As it was pointed before we can assume that the set \( \mathfrak{W} \subset \Delta_{k+1} \) is convex and closed. Note that the objective function in the right hand side of (12) is convex in \( x \) and linear in \( w \). Therefore since \( \mathfrak{W} \) and \( \mathcal{X} \) are convex compact, the ‘min’ and ‘max’ operators can be interchanged, i.e.

\[
\vartheta_* = \sup_{w \in \mathfrak{W}} \inf_{x \in \mathcal{X}} \left\{ w_0 \mathbb{E}[G_x] + \sum_{i=1}^k w_i \text{AVaR}_{\alpha_i}(G_x) \right\},
\]

and both problems (12) and (13) have nonempty sets of optimal solutions, denoted respectively as \( \mathfrak{X} \) and \( \mathfrak{W} \). We make the following assumption.

(A) For every \( i \in \{1, \ldots, k\} \) there exists \( w \in \mathfrak{W} \) such that \( w_i \neq 0 \).

This is a natural condition. Otherwise there is \( i \in \{1, \ldots, k\} \) such that \( w_i = 0 \) for all \( w \in \mathfrak{W} \). In that case we can reduce the considered set \( \{\alpha_0, \alpha_1, \ldots, \alpha_k\} \) by removing the corresponding point \( \alpha_i \).

We also can write

\[
\vartheta_* = \inf_{(x, \tau) \in \mathfrak{X} \times \mathbb{R}^k} \sup_{w \in \mathfrak{W}} \mathbb{E}[\phi(G_x, w, \tau)]
\]

\[
= \sup_{w \in \mathfrak{W}} \inf_{(x, \tau) \in \mathfrak{X} \times \mathbb{R}^k} \mathbb{E}[\phi(G_x, w, \tau)],
\]

where the function \( \phi(z, w, \tau) \) is defined in (10). Denote \( \mathcal{Y} := \mathfrak{X} \times \mathbb{R}^k \) and let \( \mathfrak{Y} \subset \mathcal{Y} \) be the set of optimal solutions of problem (14). Note that under assumption (A), the set \( \mathfrak{Y} \) consists of points \((\bar{x}, \bar{\tau})\) such that \( \bar{x} \in \mathfrak{X} \) and \( \bar{\tau}_i \) belongs to the \( \alpha_i \)-quantile interval of the cdf of \( G_x \), \( i = 1, \ldots, k \). It follows that the set \( \mathfrak{Y} \) is nonempty, convex and compact. The set of optimal solutions of problem (15) is \( \mathfrak{Y} \), the same as the one of problem (13). The minimax problem (14)–(15) is convex in \((x, \tau) \in \mathcal{Y} \) and concave (linear) in \( w \in \mathbb{R}^k \). The set of saddle points of this minimax problem is \( \mathfrak{W} \times \mathfrak{Y} \). The SAA problem for (14) writes

\[
\hat{\vartheta}_N = \inf_{(x, \tau) \in \mathfrak{X} \times \mathbb{R}^k} \sup_{w \in \mathfrak{W}} \frac{1}{N} \sum_{j=1}^N \phi(G(x, \xi_j), w, \tau).
\]
The following theorem is the main result of this section. Its proof is presented in Section 4. The main tools in derivation of this result is the minimax representations (14)–(16) and a minimax functional Central Limit Theorem (cf. [23]).

**Theorem 2.** Suppose that: (i) \( R \) is of the form (8) with the set \( \mathfrak{W} \subset \Delta_{k+1} \) being convex and closed, (ii) the set \( \mathcal{X} \) is nonempty convex and compact and \( G(x, \xi) \) is convex in \( x \), (iii) condition (A) holds, (iv) \( \mathbb{E}[G_x^2] \) is finite for some \( x^* \in \mathcal{X} \), (v) there is a measurable function \( C(\xi) \) such that \( \mathbb{E}[C(\xi)^2] \) is finite and

\[
|G(x, \xi) - G(x', \xi)| \leq C(\xi)||x - x'||, \forall x, x' \in \mathcal{X}, \forall \xi \in \Xi.
\]

Then

\[
\hat{\theta}_N = \inf_{(x, \tau) \in \mathcal{Y}} \sup_{w \in \mathfrak{W}} \left\{ \frac{w}{N} \sum_{i=1}^{N} G(x, \xi) + \sum_{i=1}^{k} w_i \left( \tau_i + \frac{1}{N(1-\alpha_i)} \sum_{j=1}^{N} [G(x, \xi) - \tau_i]_+ \right) \right\} + o_p(N^{-1/2}),
\]

and

\[
N^{1/2}(\hat{\theta}_N - \vartheta) \xrightarrow{D} \sup_{w \in \mathfrak{W}} \inf_{(x, \tau, w) \in \mathcal{Y}} \mathbb{Y}(x, \tau, w),
\]

where \( \mathbb{Y}(w, \tau) \) is a Gaussian process with mean zero and covariances

\[
\mathbb{E}[\mathbb{Y}(x, \tau, w)\mathbb{Y}(x', \tau', w')] = \mathbb{C}(w_0 G_x + \sum_{i=1}^{k} \frac{w_i}{1-\alpha_i} (G_x - \tau_i)_+, w'_0 G_{x'} + \sum_{i=1}^{k} \frac{w'_i}{1-\alpha_i} (G_{x'} - \tau'_i)_+).
\]

Moreover, if the sets \( \mathfrak{W} = \{\hat{w}\} \) and \( \mathcal{Y} = \{(\hat{x}, \hat{\tau})\} \) are singletons, then \( N^{1/2}(\hat{\theta}_N - \vartheta) \) converges in distribution to normal \( \mathcal{N}(0, \nu^2) \) with variance

\[
\nu^2 := \text{Var}[\phi(G_x, \hat{w}, \hat{\tau})] = \text{Var}\left\{ w_0 G_x + \sum_{i=1}^{k} \frac{w_i}{1-\alpha_i} (G_x - \hat{\tau}_i)_+ \right\}.
\]

**Remark 1.** It is assumed in the above theorem that the set \( \mathcal{X} \) is compact. Actually it is possbile to push the proof through with relaxing this assumption to that the respective set \( \mathcal{X} \) of optimal solutions is nonempty and compact.

**Remark 2.** For further calculation of the covariance structure (20) we may invoke Hoeffding’s covariance formula (e.g. Lemma 5.24 in [13]) to obtain for \( t, s \in \mathbb{R} \),

\[
\mathbb{C}(G_x - t)_+, (G_{x'} - s)_+) = \int_{t}^{\infty} \int_{s}^{\infty} (F_{x,x'}(u, v) - F_x(u)F_{x'}(v)) du \, dv,
\]

where \( F_{x,x'} \) denotes the joint distribution function of \( G_x \) and \( G_{x'} \) and \( F_x \) and \( F_{x'} \) denote their marginal distribution functions respectively.

Let us discuss now estimation of the variance \( \nu^2 \) given in (21). Let \( (\hat{x}_N, \hat{\tau}_N, \hat{w}_N) \) be a saddle point of the SAA problem (16). Suppose that the sets \( \mathfrak{W} = \{\hat{w}\} \) and \( \mathcal{Y} = \{(\hat{x}, \hat{\tau})\} \) are singletons. Since the sets \( \mathcal{Y} \) and \( \mathfrak{W} \) are convex and the function \( \phi(G(x, \xi), w, \tau) \) is convex in \( (x, \tau) \) and concave (linear) in \( w \), it follows that \( (\hat{x}_N, \hat{\tau}_N) \) converges w.p.1 to \( (\hat{x}, \hat{\tau}) \) and \( \hat{w}_N \) converges w.p.1 to \( \hat{w} \) as \( N \to \infty \) (e.g., [27, Theorem

6
Moreover, if the sets $Y$ follows.

Consider the sets having cdf $F$ and hence $F$ associated with cdf $F$.

In particular Theorem 2 provides an interesting asymptotic result concerning empirical estimates of risk measures. We again assume that $\mathcal{R}$ has representation as in (8) with $\mathfrak{W}$ being convex and compact. Let $Z$ be an integrable random variable having cdf $F$, and $\hat{F}_N$ be the empirical cdf based on an iid sample $Z_1, ..., Z_N \sim F$, and hence

$$\mathcal{R}(\hat{F}_N) = \sup_{w \in \mathfrak{W}} \inf_{\tau \in \mathbb{R}^k} \frac{1}{N} \sum_{j=1}^{N} \phi(Z_j, w, \tau).$$

Consider the sets

$$\mathfrak{W}(F) := \arg \max_{w \in \mathfrak{W}} \left\{ w_0 \mathbb{E}[Z] + \sum_{i=1}^{k} w_i \text{AVaR}_{\alpha_i}(Z) \right\} = \arg \max_{w \in \mathfrak{W}} \left\{ \inf_{\tau \in \mathbb{R}^k} \mathbb{E}[\phi(Z, w, \tau)] \right\},$$

$$\mathfrak{T}(F) := \Omega_F(\alpha_1) \times \cdots \times \Omega_F(\alpha_k)$$

associated with cdf $F$ of random variable $Z$. Note that under assumption (A), the set $\mathfrak{T}(F)$ gives the set of minimizers of $\mathbb{E}[\phi(Z, w, \tau)]$ for any $w \in \mathfrak{W}$. We have that $\mathfrak{W}(F) \times \mathfrak{T}(F)$ is the set of saddle points of the respective minimax problem associated with $\mathcal{R}$. Then an application of Theorem 2 to the sample estimate $\mathcal{R}(\hat{F}_N)$ reads as follows.

**Corollary 3.** Suppose that $\mathcal{R}$ is of the form (8) with $\mathfrak{W}$ being convex and closed, condition (A) holds and $\mathbb{E}_F[Z^2] < +\infty$. Then

$$\mathcal{R}(\hat{F}_N) = \sup_{w \in \mathfrak{W}(F)} \inf_{\tau \in \mathfrak{T}(F)} \left\{ \frac{1}{N} \sum_{j=1}^{N} Z_j + \sum_{i=1}^{k} w_i \left( \tau_i + \frac{1}{N(1-\alpha_i)} \sum_{j=1}^{N} \left[ Z_j - \tau_i \right]_{+} \right) \right\} + o_p(N^{-1/2})$$

and

$$N^{1/2} \left[ \mathcal{R}(\hat{F}_N) - \mathcal{R}(F) \right] \overset{D}{\rightarrow} \sup_{w \in \mathfrak{W}(F)} \inf_{\tau \in \mathfrak{T}(F)} \mathcal{Y}(w, \tau),$$

where $\mathcal{Y}(w, \tau)$ is a Gaussian process with mean zero and covariances

$$\text{Cov}_F \left( w_0 Z + \sum_{i=1}^{k} \frac{w_i}{1-\alpha_i} [Z - \tau_i]_+ , w'_0 Z + \sum_{i=1}^{k} \frac{w'_i}{1-\alpha_i} [Z - \tau'_i]_+ \right).$$

Moreover, if the sets $\mathfrak{W}(F) = \{ \bar{w} \}$ and $\mathfrak{T}(F) = \{ \bar{\tau} \}$ are singletons, then $N^{1/2} \left[ \mathcal{R}(\hat{F}_N) - \mathcal{R}(F) \right]$ converges in distribution to normal $N(0, \nu^2)$ with variance

$$\nu^2 = \text{Var}_F \left\{ \bar{w}_0 Z + \sum_{i=1}^{k} \frac{\bar{w}_i}{1-\alpha_i} [Z - \bar{\tau}_i]_+ \right\}. $
Remark 3. Corollary 3 provides an alternative representation of the asymptotic distribution of the estimator \( \mathcal{R}(\hat{F}_N) \) in comparison with the already known ones from [18] and [3]. The results there are formulated for general law invariant coherent risk measures, however with additional assumptions about tail behavior of the distribution \( F \). In particular, in [18] \( F \) is required to have a polynomial tail, more precisely
\[
\sup_{\alpha \in [0,1]} (F^{-1}(\alpha) \alpha^{d_1} (1 - \alpha^{d_2})) < \infty \quad \text{for some } d_1, d_2 \in (0, 1/2)
\]
(cf. [18, Theorem 3.7]). For law invariant coherent risk measures on \( L_1(\Omega, \mathcal{F}, \mathbb{P}) \) this condition has been relaxed in [3] by
\[
\int_{-\infty}^{\infty} \sqrt{F(u)[1 - F(u)]} \, du < \infty
\]
(cf. [3, Theorem 3.1]). It is well known that condition (28) is fulfilled if the random variable \( Z \) has absolute moments of order \( q \) for some \( q > 2 \), that property (28) implies that \( Z \) has absolute moments of order 2 (e.g. [12, p. 10]). Moreover, if \( Z \) has absolute moments of second order, it does not satisfy (28) necessarily. Hence in case of risk measures with representation of the form (8) Corollary 3 improves existing results, as it only assumes the existence of the second order moments.

Remark 4. Theorem 2 and Corollary 3 give quite a complete description of the asymptotics in case risk measure \( \mathcal{R} \) has the discrete Kusuoka representation (8). It would be natural to try to extend this analysis to the general case of Kusuoka representation (11) by writing the corresponding risk measure in the respective minimax form. It turned out to be surprisingly difficult to handle such general setting in a rigorous way. The following examples demonstrate that asymptotics of empirical estimates of law invariant coherent risk measures could behave in a quite weird way; some specific conditions are required in order for the empirical estimates to have asymptotically normal distributions.

Example 1 (Absolute semideviation risk measure). Consider risk measure
\[
\mathcal{R}_c(F) := \mathbb{E}_F[Z] + c \mathbb{E}_F[Z - \mathbb{E}_F(Z)]_+ , \quad c \in (0, 1].
\]
We assume that cdf \( F \) has finite first order moment. This risk measure has the following representation (cf., [25])
\[
\mathcal{R}_c(F) = \sup_{\gamma \in [0,1]} \{ (1 - c\gamma) \mathbb{E}_F(Z) + c\gamma \mathbb{E}_F \mathbb{V} \} \mathbb{A} \mathbb{R}_{1-\gamma}(F)
\]
\[
= \sup_{\gamma \in [0,1]} \inf_{t \in \mathbb{R}} \mathbb{E}_F \{ (1 - c\gamma) Z + c\gamma t + c[Z - t]_+ \}
\]
\[
= \inf_{t \in \mathbb{R}} \sup_{\gamma \in [0,1]} \mathbb{E}_F \{ (1 - c\gamma) Z + c\gamma t + c[Z - t]_+ \}.
\]
Representation (30) is the (minimal) Kusuoka representation (11) of \( \mathcal{R}_c \) with the corresponding set \( \mathfrak{M} = \cup_{\gamma \in [0,1]} \{ (1 - c\gamma) \mathbb{A} \} + c\gamma \mathbb{V} \{ (1 - \gamma) \} \}. Since
\[
\sup_{\gamma \in [0,1]} \mathbb{E}_F \{ (1 - c\gamma) Z + c\gamma t + c[Z - t]_+ \} = \mathbb{E}_F[Z] + c \max \{ \mathbb{E}_F[Z - t]_+, \mathbb{E}_F[t - Z]_+ \},
\]
it follows that problem (32) has unique optimal solution \( t^* = m \), where \( m := \mathbb{E}_F[Z] \), i.e., \( \mathcal{R}_c(F) = \mathbb{E}_F[Z] + c \max \{ \mathbb{E}_F[Z - \mathbb{E}_F[Z]]_+, \mathbb{E}_F[\mathbb{E}_F[Z] - Z]_+ \} \), which is consistent with definition (29) of \( \mathcal{R}_c \), because \( \mathbb{E}_F[Z - \mathbb{E}_F[Z]]_+ = \mathbb{E}_F[\mathbb{E}_F[Z] - Z]_+ \).
Now the set of minimizers of \( \gamma t + \mathbb{E}[Z - t] \), over \( t \in \mathbb{R} \), is defined by the equation \( F(t) = 1 - \gamma \). It follows that the set of saddle points of the minimax representation (31) is \([\bar{\gamma}, \gamma] \times \{ m \}\), where

\[
\bar{\gamma} := 1 - \text{Pr}(Z \leq m), \quad \gamma := 1 - \text{Pr}(Z < m)
\]

(cf., [27, Section 6.6.2]). In other words here the set of maximizers of measures \( \mu \in \mathfrak{M} \) in the Kusuoka representation is

\[
\mathfrak{M}(F) = \cap_{\gamma \in [\bar{\gamma}, \gamma]} \{ (1 - c\gamma)\delta(0) + c\gamma\delta(1 - \gamma) \},
\]

and the respective set \( \mathfrak{F}(F) = \{ \bar{\tau}(\alpha) \} \) is the singleton with \( \bar{\tau}(\alpha) = \mathbb{E}_F[Z] \) for all \( \alpha \in [0, 1] \).

The minimax representation (31) leads to the following asymptotics. Suppose that \( \mathbb{E}_F[Z^2] < +\infty \). Then by a finite dimensional minimax asymptotics theorem (cf., [23])

\[
\mathcal{R}_c(\hat{F}_N) = \sup_{\gamma \in [\bar{\gamma}, \gamma]} \left\{ c\gamma m + (1 - c\gamma)\bar{Z} + cN^{-1} \sum_{j=1}^{N} [Z_j - m]_+ \right\} + o_p(N^{-1/2}),
\]

where \( \bar{Z} := N^{-1} \sum_{j=1}^{N} Z_j \). We have here that a condition which is required for asymptotic normality of the corresponding empirical estimate is that \( \gamma = \tau \), i.e., that \( F(\cdot) \) should be continuous at \( m = \mathbb{E}_F[Z] \). If the cdf \( F(\cdot) \) is continuous at \( m = \mathbb{E}_F[Z] \), then \( N^{1/2}[\mathcal{R}_c(\hat{F}_N) - \mathcal{R}_c(F)] \) converges in distribution to normal \( \mathcal{N}(0, \nu^2) \) with variance

\[
\nu^2 = \text{Var}_F\{ (1 - c\gamma^*) Z + c[Z - m]_+ \},
\]

where \( \gamma^* := 1 - F(m) = \tilde{F}(m) \).

**Example 2** (Mean-semideviation risk measure). Consider the following risk measure

\[
\mathcal{R}_c(F) := \mathbb{E}_F[Z] + c \left( \mathbb{E}_F[Z - \mathbb{E}_F(Z)]^2 \right)^{1/2}, \quad c \in (0, 1].
\]

Asymptotics of empirical estimates of such risk measures were discussed in [4]. If \( F(\cdot) \) is continuous at \( m := \mathbb{E}_F[Z] \), then \( \mathcal{R}_c(\cdot) \) is Gâteaux differentiable at \( F \) and the corresponding influence function is

\[
IF(z) = z + c(2\theta)^{-1} \left( [z - m]_+^2 - \theta^2 + 2\kappa(1 - F(m))(z - m) \right),
\]

where \( \theta := (\mathbb{E}_F[Z - \mathbb{E}_F[Z]]^2)^{1/2} \) and \( \kappa := \mathbb{E}_F[Z - m]_+ \) (see, e.g., [27, p.345] for a more detailed discussion of this example). This indicates that continuity of \( F(\cdot) \) at \( m \) is a necessary condition for \( \mathcal{R}_c(\cdot) \) to be Gâteaux differentiable at \( F \). Here again continuity of \( F(\cdot) \) at \( m \) is a required condition for \( \mathcal{R}(\hat{F}_N) \) to be asymptotically normal.

**4. Proof of Theorem 2.** Throughout this section, we shall use notation and assumptions from Theorem 2. Moreover, let us define for \( x \in \mathcal{X}, \tau \in \mathbb{R}^k \) and \( w \in \mathfrak{M} \) the function

\[
f_{x, \tau, w} : \mathbb{R}^d \to \mathbb{R}, \quad z \mapsto w_0 G(x, z) + \sum_{i=1}^{k} w_i \left( \tau_i + \frac{1}{1 - \alpha_i} [G(x, z) - \tau_i]_+ \right).
\]
The idea to show Theorem 2 is to apply asymptotic results from empirical process theory to the class of the functions $f_{x,\tau,w}$, and then to invoke a minimax Delta Theorem. For preparation to make use of mentioned results from empirical process theory we shall verify first that the functions $f_{x,\tau,w}$ satisfy pointwise some certain Lipschitz continuity w.r.t. their parameters.

**Lemma 4.** For any $n \in \mathbb{N}$, there is a Borel-measurable function $C_n : \mathbb{R}^d \to \mathbb{R}$ such that $\mathbb{E}[C_n(z)^2] < \infty$ holds, and

$$|f_{x,\tau,w}(z) - f_{x,\tau,w}(\bar{z})| \leq C_n(z)(\|x - \bar{x}\|_{m,2} + \|\tau - \bar{\tau}\|_{k,2} + \|w - \bar{w}\|_{k+1,2})$$

is valid for any $z \in \mathbb{R}^d$, $x, \tau, \bar{x}, \tau \in [-n,n]^k$ as well as $w, \bar{w} \in \mathbb{W}$. Here $\|\cdot\|_{m,2}, \|\cdot\|_{k,2}$ and $\|\cdot\|_{k+1,2}$ denote respectively the Euclidean norms on $\mathbb{R}^m, \mathbb{R}^k$ and $\mathbb{R}^{k+1}$.

**Proof.** Let $x, \tau, \bar{x}, \tau \in [-n,n]^k$ and $w, \bar{w} \in \mathbb{W}$. Furthermore, let $x^* \in \mathcal{X}$ as in assumption (iv) of Theorem 2. Then using the triangle inequality several times we may observe for $z \in \mathbb{R}^d$

$$|f_{x,\tau,w}(z) - f_{x,\tau,w}(\bar{z})|$$

$$\leq u_0 |G(x,z) - G(\bar{x},z)| + [(w_0 - \bar{w}_0)(G(\bar{x},z) - G(x^*,z))] + |w_0 - \bar{w}_0||G(x^*,z)|$$

$$+ \sum_{i=1}^{k} w_i \left[(\tau_i - \bar{\tau}_i) + (1 - \alpha_i)^{-1}[(G(x,z) - \tau_i) + (G(\bar{x},z) - \bar{\tau}_i)] \right]$$

$$+ \sum_{i=1}^{k} \frac{|w_i - \bar{w}_i|}{1 - \alpha_i} \cdot \left[|(G(x^*,z) - \tau_i) + (G(\bar{x},z) - \bar{\tau}_i)| + |(1 - \alpha_i)\tau_i + (G(x^*,z) - \bar{\tau}_i)| \right]$$

$$\leq u_0 |G(x,z) - G(\bar{x},z)| + [(w_0 - \bar{w}_0)(G(\bar{x},z) - G(x^*,z))] + |w_0 - \bar{w}_0||G(x^*,z)|$$

$$+ \sum_{i=1}^{k} \frac{w_i}{1 - \alpha_i} \cdot [(2 - \alpha_i)|\tau_i - \bar{\tau}_i| + |G(x,z) - G(\bar{x},z)|]$$

$$+ \sum_{i=1}^{k} \frac{|w_i - \bar{w}_i|}{1 - \alpha_i} \cdot \left[|G(\bar{x},z) - G(x^*,z)| + (2 - \alpha_i)|\tau_i| + |G(x^*,z)| \right].$$

Then invoking Cauchy Schwarz inequality we obtain

$$|f_{x,\tau,w}(z) - f_{x,\tau,w}(\bar{z})|$$

$$\leq \max_{i=1,...,k} (1 - \alpha_i)^{-1} \cdot \|w\|_{k+1,2}(\sqrt{k+1}|G(x,z) - G(\bar{x},z)| + 2\|\tau - \bar{\tau}\|_{k,2})$$

$$+ \max_{i=1,...,k} (1 - \alpha_i)^{-1} \cdot \|w - \bar{w}\|_{k+1,2}(\sqrt{k+1}(|G(x^*,z) - G(\bar{x},z)| + |G(x^*,z)|) + 2\|\tau\|_{k,2}).$$

By assumption the inequalities

$$|G(x^*,z) - G(\bar{x},z)| \leq C(z)||x^* - \bar{x}||_{m,2} \leq C(z) \text{ diam}(\mathcal{X})$$

and $|G(x,z) - G(\bar{x},z)| \leq C(z)||x - \bar{x}||_{m,2}$ hold, where $C$ denotes the nonnegative Borel-measurable function $C$ as in assumption (v) of Theorem 2, and diam$(\mathcal{X})$ stands
for the diameter of the compact set $\mathcal{X}$ w.r.t. the Euclidean norm on $\mathbb{R}^n$. Hence

$$|f_{x, \tau, w}(z) - f_{x, \tau, \bar{w}}(z)|$$

$$\leq \max_{i=1, \ldots, k} (1 - \alpha_i)^{-1} \cdot \|w\|_{k+1, 2} \left( \sqrt{k + 1} C(z) + 2\|\tau\|_{k, 2} \right)$$

$$+ \max_{i=1, \ldots, k} (1 - \alpha_i)^{-1} \cdot \|w - \bar{w}\|_{k+1, 2} \left( \sqrt{k + 1} C(z) \text{ diam}(\mathcal{X}) + |G(x^*, z)| + 2\|\tau\|_{k, 2} \right)$$

$$\leq \max_{i=1, \ldots, k} (1 - \alpha_i)^{-1} \cdot \left( \sqrt{k + 1} C(z) + 2\|\tau\|_{k, 2} \right)$$

$$+ \max_{i=1, \ldots, k} (1 - \alpha_i)^{-1} \cdot \|w - \bar{w}\|_{k+1, 2} \left( \sqrt{k + 1} C(z) \text{ diam}(\mathcal{X}) + |G(x^*, z)| + 2\sqrt{k} n \right).$$

Now the function $C_n : \mathbb{R}^d \to \mathbb{R}$, defined by

$$C_n(z) := \max_{i=1, \ldots, k} (1 - \alpha_i)^{-1} \cdot \left( \sqrt{k + 1} C(z), 2\sqrt{k} n + \sqrt{k + 1} C(z) \text{ diam}(\mathcal{X}) + |G(x^*, z)| \right),$$

is as required due to assumptions (iv), (v) of Theorem 2.

In the next step we want to show that from an asymptotic viewpoint we may replace the estimator $\hat{\vartheta}_N$ with the estimator

$$\hat{\vartheta} := \inf_{(x, \tau) \in \mathcal{K}} \sup_{w \in \mathbb{R}} \left\{ \frac{w_0}{N} \sum_{j=1}^{N} G(x, \xi_j) + \sum_{i=1}^{k} w_i \left( \tau_i + \frac{1}{N(1 - \alpha_i)} \sum_{j=1}^{N} [G(x, \xi_j) - \tau_j]_{+} \right) \right\}$$

for some compact subset $\mathcal{K}$ of $\mathcal{X} \times \mathbb{R}^k$. This estimator is more convenient as it allows to apply the minimax functional Central Limit Theorem from [23].

**Lemma 5.** Let $n_0 \in \mathbb{N}$ such that $\mathcal{Y} \subseteq \mathcal{X} \times (-n_0, n_0)^k$. Then

$$\hat{\vartheta}_N = \inf_{(x, \tau) \in \mathcal{K}} \sup_{w \in \mathbb{R}} \left\{ \frac{w_0}{N} \sum_{j=1}^{N} G(x, \xi_j) + \sum_{i=1}^{k} w_i \left( \tau_i + \frac{1}{N(1 - \alpha_i)} \sum_{j=1}^{N} [G(x, \xi_j) - \tau_j]_{+} \right) \right\} + o_P(N^{-1/2}),$$

where $\mathcal{K} := \mathcal{X} \times [-n_0, n_0]^k$.

**Proof.** Let $\phi$ denote the mapping as defined in (10). In particular

$$(37) \quad f_{x, \tau, w}(z) = \phi(G(x, z), w, \tau) \quad \text{for} \ (x, \tau, w, z) \in \mathcal{X} \times \mathbb{R}^k \times \mathcal{W} \times \mathbb{R}^d.$$  

According to Lemma 4, we may draw on [30, Example 19.7 and Theorem 19.4] to find for every $n \in \mathbb{N}$ some $A_n \in \mathcal{F}$ with $\mathbb{P}(A_n) = 1$ such that

$$(38) \quad \sup_{(x, \tau) \in \mathcal{X} \times [-n, n]^k} \sup_{w \in \mathcal{W}} \left| \frac{1}{N} \sum_{j=1}^{N} \phi(G(x, \xi_j(\omega)), w, \tau) - \mathbb{E} \left[ \phi(G(x, \xi), w, \tau) \right] \right| \rightarrow 0$$

for $\omega \in A_n$. Then $A := \bigcap_{n=1}^{\infty} A_n$ satisfies $\mathbb{P}(A) = 1$. Moreover, in view of Lemma 4 along with (38)

$$h_N : \mathcal{X} \times \mathbb{R}^k \times \Omega \to \mathbb{R}, \ (x, \tau, \omega) \mapsto \sup_{w \in \mathcal{W}} \frac{1}{N} \sum_{j=1}^{N} \phi(G(x, \xi_j(\omega)), w, \tau)$$

11
defines a sequence \((h_N)_{N \in \mathbb{N}}\) of mappings such that for any \(\omega \in A\), the mapping \(h_N(\cdot, \omega)\) is lower-semicontinuous and the sequence \((h_N(\cdot, \omega))_{N \in \mathbb{N}}\) converges uniformly on compact subsets to the function \(\sup_{w \in \mathbb{W}} E[\phi(G(\cdot, \xi), w, \cdot)]\) with

\[
\inf_{(x, \tau) \in K} h_N(x, \tau, \omega) \to \inf_{(x, \tau) \in K} \sup_{w \in \mathbb{W}} E[\phi(G(\cdot, \xi), w, \cdot)] = \partial_0^* \quad \text{for } N \to \infty.
\]

Since \(K\) is compact, we may find for every \(\omega \in A\) a sequence \((\hat{x}_N(\omega), \hat{\tau}_N(\omega))\) such that \((\hat{x}_N(\omega), \hat{\tau}_N(\omega))\) minimizes \(h_N(\cdot, \omega)\) for any \(N \in \mathbb{N}\). By compactness of \(K\) for any \(\omega \in A\), the sequence \((\hat{x}_N(\omega), \hat{\tau}_N(\omega))\) has cluster points which all belong to the set \(\mathcal{S}\) of minimizer of the function \(\sup_{w \in \mathbb{W}} E[\phi(G(\cdot, \xi), w, \cdot)]\) on \(K\) because \((h_N(\cdot, \omega))_{N \in \mathbb{N}}\) converges uniformly on \(K\) to \(\sup_{w \in \mathbb{W}} E[\phi(G(\cdot, \xi), w, \cdot)]\) (cf. [22, Theorem 7.31]). In particular, the distance of \((\hat{x}_N(\omega), \hat{\tau}_N(\omega))\) to \(\mathcal{S}\) tends to zero as \(N \to \infty\) for every \(\omega \in A\). Note \(\mathcal{S} = \overline{Y}\) so that for every \(\omega \in A\) there is some \(N(\omega) \in \mathbb{N}\) such that

\[
(39) \quad (\hat{x}_N(\omega), \hat{\tau}_N(\omega)) \in X \times (-n_0, n_0)^k \quad \text{for arbitrary } N \in \mathbb{N} \text{ with } N \geq N(\omega).
\]

In view of assumption (ii) of Theorem 2, the mapping \(\phi(G(\cdot, z), w, \cdot)\) is convex for every \(z \in \mathbb{R}^d\) and any \(w \in \mathbb{W}\). This implies that \(h_N(\cdot, \omega)\) is convex for \(N \in \mathbb{N}\) and \(\omega \in A\), and thus

\[
\min_{\lambda \in (0, 1)} h_N(\lambda(x, \tau) + (1 - \lambda)(\hat{x}_N(\omega), \hat{\tau}_N(\omega)), \omega) \leq \min \left\{ h_N(x, \tau, \omega), h_N(\hat{x}_N(\omega), \hat{\tau}_N(\omega), \omega) \right\}
\]

holds for \((x, \tau) \in X \times \mathbb{R}^k\). Then by (39), we obtain for any \(\omega \in A\) and every \(N \in \mathbb{N}\) with \(N \geq N(\omega)\)

\[
\hat{\partial}_N(\omega) = \inf_{(x, \tau) \in \mathbb{R}^k} h_N(x, \tau, \omega) = \inf_{(x, \tau) \in K} h_N(x, \tau, \omega),
\]

and then

\[
\sqrt{N} \left[ \inf_{(x, \tau) \in X \times \mathbb{R}^k} \sup_{w \in \mathbb{W}} \left\{ \frac{1}{N} \sum_{j=1}^{N} \phi(G(x, \xi_j(\omega)), w, \tau) \right\} - \inf_{(x, \tau) \in K} \sup_{w \in \mathbb{W}} \left\{ \frac{1}{N} \sum_{j=1}^{N} \phi(G(x, \xi_j(\omega)), w, \tau) \right\} \right] = 0
\]

for \(N \in \mathbb{N}\) with \(N \geq N(\omega)\). Hence

\[
\sqrt{N} \left[ \inf_{(x, \tau) \in X \times \mathbb{R}^k} \sup_{w \in \mathbb{W}} \left\{ \frac{1}{N} \sum_{j=1}^{N} \phi(G(x, \xi_j(\omega)), w, \tau) \right\} - \inf_{(x, \tau) \in K} \sup_{w \in \mathbb{W}} \left\{ \frac{1}{N} \sum_{j=1}^{N} \phi(G(x, \xi_j(\omega)), w, \tau) \right\} \right] \to 0 \quad \mathbb{P} - \text{a.s.,}
\]

implying

\[
\sqrt{N} \left[ \inf_{(x, \tau) \in X \times \mathbb{R}^k} \sup_{w \in \mathbb{W}} \left\{ \frac{1}{N} \sum_{j=1}^{N} \phi(G(x, \xi_j(\omega)), w, \tau) \right\} - \inf_{(x, \tau) \in K} \sup_{w \in \mathbb{W}} \left\{ \frac{1}{N} \sum_{j=1}^{N} \phi(G(x, \xi_j(\omega)), w, \tau) \right\} \right] \to 0
\]

in probability. This completes the proof. \(\square\)

Now we are ready to prove Theorem 2.

**Proof of Theorem 2**
Proof. Let φ denote the function defined in (10). By Lemma 5 we may find some \( n_0 \in \mathbb{N} \) such that \( Y \subseteq X \times [-n_0, n_0]^k \),

\[
\vartheta_\ast = \inf_{(x, \tau) \in X \times [-n_0, n_0]^k} \sup_{w \in \mathcal{W}} \mathbb{E}[\phi(G(x, \xi), w, \tau)]
\]

and

\[
\hat{\vartheta}_N = \inf_{(x, \tau) \in X \times [-n_0, n_0]^k} \sup_{w \in \mathcal{W}} \frac{1}{N} \sum_{j=1}^{N} \phi(G(x, \xi_j), w, \tau) + o_p(N^{-1/2}).
\]

Set \( \mathcal{K} := X \times [-n_0, n_0]^k \) and

\[
\overline{\vartheta}_N = \inf_{(x, \tau) \in \mathcal{K}} \sup_{w \in \mathcal{W}} \frac{1}{N} \sum_{j=1}^{N} \phi(G(x, \xi_j), w, \tau).
\]

The idea now is to apply Theorem 2.1 from [23], a minimax Delta Theorem, to \((\overline{\vartheta}_N)_{N \in \mathbb{N}}\) and \( \vartheta_\ast \). For this purpose consider the stochastic process \((V^N_{x, \tau, w}(x, \tau, w))_{(x, \tau, w) \in \mathcal{K} \times \mathcal{W}}\), defined by

\[
V^N_{x, \tau, w} = \frac{1}{N} \sum_{j=1}^{N} \phi(G(x, \xi_j), w, \tau) \quad \text{for} \ (x, \tau, w) \in \mathcal{K} \times \mathcal{W}.
\]

Using Lemma 4 and recalling (37) it may be viewed as a Borel random element \( V^N \) of the space \( \mathcal{C}(\mathcal{K} \times \mathcal{W}) \) of continuous real-valued mappings on \( \mathcal{K} \times \mathcal{W} \) which is endowed with the uniform metric. In the same way the mapping

\[
V : \mathcal{K} \times \mathcal{W} \rightarrow \mathbb{R}, \ (x, \tau, w) \mapsto \mathbb{E}[\phi(G(x, \xi), w, \tau)]
\]

may be verified as a member of \( \mathcal{C}(\mathcal{K} \times \mathcal{W}) \). Drawing on Lemma 4 again, we may apply Example 19.7 from [30] to conclude that the sequence \( N^{-1/2}(V^N - V)_{N \in \mathbb{N}} \) converges in law to some centered Gaussian random element \( Y \) of \( \mathcal{C}(\mathcal{K} \times \mathcal{W}) \) with covariances

\[
\mathbb{E}[Y(x, \tau, w) \cdot Y(x', \tau', w')] = \text{Cov}(\phi(G(x, \xi), w, \tau), \phi(G(x', \xi), w', \tau')).
\]

By assumption (ii) of Theorem 2, the mapping \( \phi(G(\cdot, z), w, \cdot) \) is convex for every \( z \in \mathbb{R}^d \) and any \( w \in \mathcal{W} \). Hence the stochastic process \((V^N_{x, \tau, w}(x, \tau, w))_{(x, \tau, w) \in \mathcal{K}}\) has convex paths and \( V(\cdot, \cdot, w) \) is convex for any \( w \in \mathcal{W} \). Moreover, the mapping \( \phi(G(x, z), \cdot, \tau) \) is concave for every \( (x, \tau) \in \mathcal{K} \) which implies that the stochastic process \((V^N_{x, \tau, w}, w \in \mathcal{W})\) has concave paths and \( V(x, \tau, \cdot) \) is concave for arbitrary \( (x, \tau) \in \mathcal{K} \). Now the statement of Theorem 2 follows immediately from [23, Theorem 2.1] along with (40), (41) and (42).

5. Hypotheses testing. Using the results of the previous sections, we now propose asymptotic rejection regions for tests (3) and (4) (in Section 5.1) on the basis of samples \( \xi^{N,v} = (\xi_1^v, \ldots, \xi_N^v) \) of \( \xi^v \) for \( v = 1, \ldots, V \). We will also study tests (3) in a nonasymptotic framework (in Section 5.2) deriving nonasymptotic confidence intervals on the optimal value of (1). We will denote by \( 0 < \beta < 1 \) the maximal probability of type I error.
5.1. Asymptotic tests. Tests (3) and (4). Let us consider \( V > 1 \), optimization problems of the form (1) with \( \xi \), \( g(x) \), and \( \mathcal{X} \) respectively replaced by \( \xi^v \), \( g_v(x) = \mathcal{R}(G_v(x)) \), and \( \mathcal{X}_v \) for problem \( v \). In the above definition of \( g_v \), \( G_v \) satisfies \( G_v(\omega) = G_v(x, \xi^v(\omega)) \). For \( v = 1, \ldots, V \), let \( (\xi_v^1, \ldots, \xi_v^N) \) be a sample from the distribution of \( \xi^v \), let \( \hat{\vartheta}_v^* \) be the optimal value of problem \( v \) and let \( z_v^* = (x_v^*, \tau_v^*, w_v^*) \) be an optimal solution of the problem, written under form (14), in variables \( z = (x, \tau, w) \). Let \( \hat{\vartheta}_v^* \) be the SAA estimator of the optimal value for problem \( v = 1, \ldots, V \). Defining the function \( H_v(z, \xi^v) = \phi(G_v(x, \xi^v), w, \tau) \) in variables \( z = (x, \tau, w) \) with \( \phi \) given by (10), we also denote by \( \hat{\vartheta}_v^* \) the empirical estimator of the variance \( \text{Var}[H_v(z_v^*, \xi^v)] \) based on the sample for problem \( v \). We assume that the samples are i.i.d. and that \( \xi_{N,1}^1, \ldots, \xi_{N,V}^N \) are independent. Under the assumptions of Theorem 2 for \( N \) large we can approximate the distribution of \( N^{1/2}(\hat{\vartheta}_v^* - \vartheta^*_{v}) / \sigma_v \) by the standard normal \( \mathcal{N}(0, 1) \).

Let us first consider the statistical tests (3)-(a) and (3)-(b) with \( V = 2 \):

\[
\begin{align*}
H_0 : & \quad \vartheta_1 = \vartheta_2 \text{ against } H_1 : \vartheta_1 \neq \vartheta_2 \\
H_0 : & \quad \vartheta_1 \leq \vartheta_2 \text{ against } H_1 : \vartheta_1 > \vartheta_2 .
\end{align*}
\]

For \( N \) large, we approximate the distribution of \( N^{1/2}(\hat{\vartheta}_v^* - \vartheta^*_{v}) / \sigma_v \) by the standard normal \( \mathcal{N}(0, 1) \) and we obtain the rejection regions

\[
\begin{align*}
(\xi_{N,1}, \xi_{N,2}) & : |\hat{\vartheta}_N - \vartheta_1| > \sqrt{\frac{(\vartheta_1^2 - \vartheta_2^2)^2}{N} + \frac{(\vartheta_1^2 - \vartheta_2^2)^2}{N} \Phi^{-1}(1 - \beta)} \\
(\xi_{N,1}, \xi_{N,2}) & : \hat{\vartheta}_N > \vartheta_2 + \sqrt{\frac{(\vartheta_1^2 - \vartheta_2^2)^2}{N} + \frac{(\vartheta_1^2 - \vartheta_2^2)^2}{N} \Phi^{-1}(1 - \beta)}
\end{align*}
\]

for test (3)-(a) with \( V = 2 \). For test (3)-(b) with \( V = 2 \):

\[
(\xi_{N,1}, \xi_{N,2}) : \hat{\vartheta}_N > \vartheta_2 + \sqrt{\frac{(\vartheta_1^2 - \vartheta_2^2)^2}{N} + \frac{(\vartheta_1^2 - \vartheta_2^2)^2}{N} \Phi^{-1}(1 - \beta)}
\]

Let us now consider test (4):

\[
H_0 : \theta \in \Theta_0 \text{ against } H_1 : \theta \in \mathbb{R}^V
\]

for \( \theta = (\vartheta_1^1, \ldots, \vartheta_V^1)^T \) with \( \Theta_0 \) a linear space or a closed convex cone.

Let \( \Theta_0 \) be the subspace

\[
\Theta_0 = \{ \theta \in \mathbb{R}^V : A\theta = 0 \}
\]

where \( A \) is a \( k_0 \times V \) matrix of full rank \( k_0 \). Note that test (3)-(a) can be written in this form with \( A \) a \( (V - 1) \times V \) matrix of rank \( V - 1 \). We have for \( \theta \) the estimator \( \hat{\vartheta}_N = (\hat{\vartheta}_N^1, \ldots, \hat{\vartheta}_N^V)^T \). Fixing \( N \) large, since \( \xi_{N,1}^1, \ldots, \xi_{N,V}^N \) are independent, using the fact that \( N^{1/2}(\hat{\vartheta}_N - \vartheta^*_{v}) / \sigma_v \to \mathcal{N}(0, 1) \), the distribution of \( \hat{\vartheta}_N \) can be approximated by the Gaussian \( \mathcal{N}(\theta, \Sigma) \) distribution with \( \Sigma \) the diagonal matrix \( \Sigma = (1/N) \text{diag} \left( \text{Var}(H_1(z_v^1, \xi^1)), \ldots, \text{Var}(H_V(z_v^V, \xi^V)) \right) \). The log-likelihood ratio statistic for test (4) is \( \Lambda = \sup_{\theta \in \Theta_0, \Sigma > 0} \frac{L(\theta, \Sigma)}{L(\hat{\theta}_N, \hat{\Sigma})} \) where \( L(\theta, \Sigma) \) is the likelihood function for a Gaussian multivariate model. For a sample \( (\tilde{\vartheta}_1, \ldots, \tilde{\vartheta}_M) \) of \( \hat{\vartheta}_N \), introducing the

\[\text{This sample is obtained from independent samples } \xi_{N,m,v} \text{ of size } N \text{ of } \xi^v \text{ for } m = 1, \ldots, M, v = 1, \ldots, V. \text{ More precisely, } \theta \text{ component of } \tilde{\theta}_m \text{ is the optimal value of the SAA of problem } v \text{ obtained taking sample } \xi_{N,m,v} \text{ of } \xi^v.\]
estimators
\[ \hat{\theta} = \frac{1}{M} \sum_{i=1}^{M} \hat{\theta}_i \quad \text{and} \quad \hat{\Sigma} = \frac{1}{M-1} \sum_{i=1}^{M} (\hat{\theta}_i - \hat{\theta})(\hat{\theta}_i - \hat{\theta})^T \]
of respectively \( \theta \) and \( \Sigma \), we have
\[ (45) \quad -2 \ln \Lambda = V \ln \left(1 + \frac{T^2}{M-1}\right) \quad \text{where} \quad T^2 = V \min_{\theta \in \Theta_0} (\hat{\theta} - \theta)^T \hat{\Sigma}^{-1}(\hat{\theta} - \theta) \]
and when \( \Theta_0 \) is of the form (44), under \( H_0 \), we have that Hotelling’s \( T^2 \) squared statistic approximately has distribution \( \frac{k_0(M-1)}{M-k_0} F_{k_0,M-k_0} \) (see, e.g., [14]), where \( F_{p,q} \) is the Fisher-Snedecor distribution with degrees of freedom \( p \) and \( q \). For asymptotic test (4) at confidence level \( \beta \) with \( \Theta_0 \) given by (44), we then reject \( H_0 \) if \( T^2 \geq \frac{k_0(M-1)}{M-k_0} F_{1,1}^{-1}(1-\beta) \) where \( F_{p,q}^{-1}(\beta) \) is the \( \beta \)-quantile of the Fisher-Snedecor distribution.

Now take for \( \Theta_0 \) the convex cone \( \Theta_0 = \{ \theta \in \mathbb{R}^V : A\theta \leq 0 \} \) where \( A \) is a \( k_0 \times V \) matrix of full rank \( k_0 \) (tests (3)-(b), (c) are special cases) and assume that \( M \geq V+1 \). Since the corresponding null hypothesis is \( \theta \) belongs to a one-sided cone, on the basis of the sample \((\hat{\theta}_1, \ldots, \hat{\theta}_M) \) of \( \hat{\theta}_N \), we can use [16] and we reject \( H_0 \) for large values of the statistic
\[ U(\Theta_0) = \|\hat{\theta}\|_S^2 - \|\Pi_S(\hat{\theta}|\Theta_0)\|_S^2 = \|\hat{\theta} - \Pi_S(\hat{\theta}|\Theta_0)\|_S^2 \]
where \( S = \frac{M-1}{M} \hat{\Sigma} \), \( \|x\|_S = \sqrt{x^T S^{-1} x} \), and \( \Pi_S(x|A) \) is any point in \( A \) minimizing \( \|y-x\|_S \) among all \( y \in A \). For a type I error of at most \( 0 < \beta < 1 \), knowing that [16]
\[ (46) \quad \sup_{u \in \Theta_0, S > 0} \mathbb{P}(U(\Theta_0) \geq u) \leq \text{Err}(u) := \frac{1}{2} \left[ \mathbb{P}(G_{V\cup M-V-1} \geq u) + \mathbb{P}(G_{V-M-V} \geq u) \right], \]
where \( G_{m,n} = (m/n) F_{m,n} \), we reject \( H_0 \) if \( U(\Theta_0) \geq u_\beta \) where \( u_\beta \) satisfies \( \beta = \text{Err}(u_\beta) \) with \( \text{Err}(\cdot) \) given by (46).

5.2. Nonasymptotic tests.

5.2.1. Risk-neutral case. Let us consider \( V \geq 2 \) optimization problems of the form (1) with \( \mathcal{R} := \mathbb{E} \) the expectation. In this situation, several papers have derived nonasymptotic confidence intervals on the optimal value of (1): [17] using Talagrand inequality ([28], [29]), [26], [8] using large-deviation type results, and [15], [11], [7] using Robust Stochastic Approximation (RSA) [19], [20], Stochastic Mirror Descent (SMD) [15] and variants of SMD. In all cases, the confidence interval depends on a sample \( \xi^N = (\xi_1, \ldots, \xi_N) \) of \( \xi \) and of parameters. For instance, the confidence interval \([\text{Low}(\Theta_2, \Theta_3, N), \text{Up}(\Theta_1, N)]\) with confidence level \( 1 - \beta \) from [7] obtained using RSA depends on parameters \( \Theta_1 = 2 \sqrt{\ln(2/\beta)} \), \( \Theta_3 = 2 \sqrt{\ln(4/\beta)} \), \( \Theta_2 \) satisfying \( e^{-\Theta_2^2} + e^{-\Theta_2^2/4} = \frac{2}{\beta} \), and \( L, M_1, M_2, D(\mathcal{X}) \) with \( D(\mathcal{X}) \) the maximal Euclidean distance in \( \mathcal{X} \) to \( x_1 \) (the initial point of the RSA algorithm), \( L \) a uniform upper bound on \( \mathcal{X} \) on the \( \| \cdot \|_2 \)-norm of some selection (say, selection \( g'(x) \in \partial g(x) \) at \( x \)) of subgradients of \( g \), and \( M_1, M_2 < +\infty \) such that for all \( x \in \mathcal{X} \) it holds
\[ (47) \quad \begin{align*}
(a) \quad & \mathbb{E} \left[ (g(x, \xi) - g(x))^2 \right] \leq M_1^2, \\
(b) \quad & \mathbb{E} \left[ \|g'_x(x, \xi) - \mathbb{E}[g'_x(x, \xi)]\|_2^2 \right] \leq M_2^2,
\end{align*} \]
for some selection \( g'_x(x, \xi) \) belonging to the subdifferential \( \partial_x g(x, \xi) \).
With this notation, on the basis of a sample $\xi^N = (\xi_1, \ldots, \xi_N)$ of size $N$ of $\xi$ and of the trajectory $x_1, \ldots, x_N$ of the RSA algorithm, setting

$$a(\Theta, N) = \frac{\Theta M_1}{\sqrt{N}} \quad \text{and} \quad b(\Theta, \mathcal{X}, N) = \frac{K_1(\mathcal{X}) + \Theta(K_2(\mathcal{X}) - M_1)}{\sqrt{N}},$$

where the constants $K_1(\mathcal{X})$ and $K_2(\mathcal{X})$ are given by

$$K_1(\mathcal{X}) = \frac{D(\mathcal{X})(M_2^2 + 2L^2)}{\sqrt{2(M_2^2 + L^2)}} \quad \text{and} \quad K_2(\mathcal{X}) = \frac{D(\mathcal{X})M_2^3}{\sqrt{2(M_2^2 + L^2)}} + 2D(\mathcal{X})M_2 + M_1,$$

the lower bound $\text{Low}(\Theta_2, \Theta_3, N)$ is

$$\text{Low}(\Theta_2, \Theta_3, N) = \frac{1}{N} \sum_{t=1}^N G(x_t, \xi_t) - b(\Theta_2, \mathcal{X}, N) - a(\Theta_3, N),$$

and the upper bound $\text{Up}(\Theta_1, N)$ is

$$\text{Up}(\Theta_1, N) = \frac{1}{N} \sum_{t=1}^N G(x_t, \xi_t) + a(\Theta_1, N).$$

More precisely, we have $\mathbb{P}(\vartheta^* < \text{Low}(\Theta_2, \Theta_3, N)) \leq \beta/2$ and $\mathbb{P}(\vartheta^* > \text{Up}(\Theta_1, N) \leq \beta/2$.

**Test (3)-(a).** Using these bounds $\text{Low}$ and $\text{Up}$ or one of the aforementioned cited procedures, we can determine for optimization problem $v \in \{1, \ldots, V\}$ (stochastic) lower and upper bounds on $\vartheta^*$ that we will denote by $\text{Low}_v$ and $\text{Up}_v$ respectively for short, such that $\mathcal{P}(\vartheta^*_v < \text{Low}_v) \leq \frac{\beta}{2V}$ and $\mathcal{P}(\vartheta^*_v > \text{Up}_v) \leq \frac{\beta}{2V}$.

We define for test (3)-(a) the rejection region $\mathcal{W}_{(3)-(a)}$ to be the set of samples such that the realizations of the confidence intervals $[\text{Low}_v, \text{Up}_v], v = 1, \ldots, V$, on the optimal values have no intersection, i.e.,

$$\mathcal{W}_{(3)-(a)} = \{ (\xi^n_1, \ldots, \xi^n_N, V) : \bigcap_{v=1}^V [\text{Low}_v, \text{Up}_v] = \emptyset \} = \{ (\xi^n_1, \ldots, \xi^n_N, V) : \max_{v=1, \ldots, V} \text{Low}_v > \min_{v=1, \ldots, V} \text{Up}_v \}.$$

If $H_0$ holds, denoting $\vartheta^*_1 = \vartheta^*_2 = \ldots = \vartheta^*_V$, we have

$$\mathbb{P} \left( \max_{v=1, \ldots, V} \text{Low}_v > \min_{v=1, \ldots, V} \text{Up}_v \right) = \mathbb{P} \left( \max_{v=1, \ldots, V} [\text{Low}_v - \vartheta^*_v] + \max_{v=1, \ldots, V} [\vartheta^*_v - \text{Up}_v] > 0 \right) \leq \sum_{v=1}^V \mathbb{P} \left( \text{Low}_v - \vartheta^*_v > 0 \right) + \mathbb{P} \left( \vartheta^*_v - \text{Up}_v > 0 \right) \leq \beta$$

and $\mathcal{W}_{(3)-(a)}$ is a rejection region for (3)-(a) yielding a probability of type I error of at most $\beta$. Moreover, as stated in the following lemma, if $H_0$ does not hold and if two optimal values are sufficiently distant then the probability to accept $H_0$ will be small:

**Lemma 6.** Consider test (3)-(a) with rejection region $\mathcal{W}_{(3)-(a)}$. If for some $p, q \in \{1, \ldots, V\}$ with $p \neq q$, we have almost surely $\vartheta^*_p > \vartheta^*_q + \text{Up}_p - \text{Low}_p + \text{Up}_q - \text{Low}_q$ then the probability to accept $H_0$ is not larger than $\frac{\beta}{V}$. 

16
Proof. We first check that
\begin{equation}
\begin{cases}
\vartheta_*^p > \vartheta_*^q + U_p q - L_q + U_p L_q \quad (a) \\
L_q \leq \vartheta_*^q \quad (b) \\
\vartheta_*^p \leq U_p \quad (c)
\end{cases}
\Rightarrow U_p q < L_q p.
\end{equation}
Indeed, if (51)-(a), (b), and (c) hold then
\[ U_p q = L_q q + U_p q - L_q q \leq \vartheta_*^q + U_p q - L_q q < \vartheta_*^q + U_p q - U_p \leq L_q p. \]
Assume now that \( \vartheta_*^p > \vartheta_*^q + U_p q - L_q q + U_p q - L_q q \). Since \( U_p q < L_q p \) implies that \( H_0 \) is rejected, we get
\[ \mathbb{P}(\text{reject } H_0) \geq \mathbb{P}(U_p q < L_q q) \geq \mathbb{P}\left( L_q q \leq \vartheta_*^q \right) \cap \left\{ \vartheta_*^p \leq U_p \right\}, \]
which achieves the proof of the lemma.

Similarly, for tests (3)-(b) and (3)-(c), we define respectively the rejection regions \( \mathcal{W}_{(3)-(b)} \) and \( \mathcal{W}_{(3)-(c)} \) given by
\[ \mathcal{W}_{(3)-(b)} = \left\{ \xi \in \mathbb{R}^N : \exists 1 \leq q \neq p \leq V \text{ such that } L_q > U_p \right\}, \]
\[ \mathcal{W}_{(3)-(c)} = \left\{ \xi \in \mathbb{R}^N : \exists v \in \{1, \ldots, V-1\} \text{ such that } L_v > U_{v+1} \right\}, \]
yielding a probability of type I error of at most \( \beta \) provided \( [L_v, U_v] \) is a confidence interval with confidence level at least \( 1 - \beta/2( V - 1 ) \) for problem \( v \):
\begin{equation}
\mathbb{P}(\vartheta_*^v < L_v) \leq \beta/2( V - 1 ) \quad \text{and} \quad \mathbb{P}(\vartheta_*^v > U_v) \leq \beta/2( V - 1 ).
\end{equation}

Similarly to Lemma 6, we can bound from above the probability of type I error for test (3)-(b) if \( \vartheta_*^p > \vartheta_*^q + U_p q - L_q q + U_p q - L_q q \) almost surely and for test (3)-(c) if \( \vartheta_*^p > \vartheta_*^{p+1} + U_p q - L_q q + U_{p+1} q - L_{q+1} q \) almost surely.

Remark 5. Though \( L_q \) and \( U_q \) are stochastic, for bounds (49) and (50), the difference \( U_p - L_q = a(\Theta_1, N) + b(\Theta_2, N) + c(\Theta_3, N) \) is deterministic and inequalities \( \vartheta_*^p > \vartheta_*^q + U_p q - L_q q + U_p q - L_q q \) in Lemma 6 and \( \vartheta_*^p > \vartheta_*^{p+1} + U_p q - L_q q + U_{p+1} q - L_{q+1} q \) are deterministic too.

5.2.2. Risk averse case. Consider \( K \geq 2 \) optimization problems of the form (1). For such problems, nonasymptotic confidence intervals \( [L_q, U_q] \) on the optimal value \( \theta_* \) were derived in [7] and [11] using RSA and SMD, taking for \( \mathcal{R} \) an extended polyhedral risk measure (introduced in [9]) in [7] and \( \mathcal{R} = \text{AVaR}_\alpha \) and \( G(x, \xi) = \xi^\top x \) in [11]. With such confidence intervals at hand, we can use the developments of the previous section for testing hypotheses (3). However, the analysis in [7] assumes boundedness of the feasible set of the optimization problem defining the risk measure; an assumption that can be enforced for risk measure \( \mathcal{R} \) given by (53). We provide in this situation formulas for the constants \( L, M_1, \) and \( M_2 \) defined in the previous section, necessary to compute the bounds from [7]. These constants are slightly refined versions of the constants given in Section 4.2 of [11] for the special case \( \mathcal{R} = \text{AVaR}_\alpha \) and \( G(x, \xi) = \xi^\top x \).
We assume here that the set $\Xi$ is compact, $G(\cdot, \cdot)$ is continuous, for every $x \in \mathcal{X}$ the distribution of $G_x$ is continuous, and that the set $\mathcal{W} = \{w\}$ is a singleton i.e.,

$$R(Z) = w_0 \mathbb{E}[Z] + \sum_{i=1}^{k} w_i \text{AVaR}_{\alpha_i}(Z)$$

for some $w \in \Delta_{k+1}$. Consequently problem (1) can be written as

$$\theta_* = \inf_{(x, \tau) \in \mathcal{X} \times \mathbb{R}^k} \{ \mathbb{E}[\phi(G_{x, \tau})] = \mathbb{E}[H(x, \tau, \xi)] \},$$

where $\phi(G_{x, \tau})$ is defined in (10), with vector $w$ omitted, and

$$H(x, \tau, \xi) := w_0 G(x, \xi) + \sum_{i=1}^{k} w_i \left( \tau_i + \frac{1}{1-\alpha_i} [G(x, \xi) - \tau_i]_+ \right).$$

For a given $x \in \mathcal{X}$ the minimum in (54) is attained at $\tau_i = F^{-1}_x(\alpha_i), i = 1, \ldots, k,$ where $F_x$ is the cdf of $G_x$. Therefore, using the lower and upper bounds from [11] for the quantile of a continuous distribution with finite mean and variance, we can restrict $\tau$ to compact set $T = [\bar{\tau}, \bar{\tau}] \subset \mathbb{R}^k$ where

$$\tau_i = \min_{x \in \mathcal{X}} \mathbb{E}[G_x] - \sqrt{\frac{1-\alpha_i}{\alpha_i}} \sqrt{\max_{x \in \mathcal{X}} \text{Var}(G_x)},$$

$$\bar{\tau}_i = \max_{x \in \mathcal{X}} \mathbb{E}[G_x] + \sqrt{\frac{\alpha_i}{1-\alpha_i}} \sqrt{\max_{x \in \mathcal{X}} \text{Var}(G_x)},$$

for $i = 1, \ldots, k$. This implies that we can take $D(\mathcal{X} \times T) = \sqrt{D(\mathcal{X})^2 + \|\bar{\tau} - \bar{\tau}\|^2}$.

**Computation of $M_0$.** Setting

$$M_0 := \max_{(x, \xi) \in \mathcal{X} \times \Xi} G(x, \xi) \quad \text{and} \quad m_0 := \min_{(x, \xi) \in \mathcal{X} \times \Xi} G(x, \xi),$$

we have for $(x, \tau) \in \mathcal{X} \times T$ that $|G_{x} - \mathbb{E}[G_x]| \leq M_0 - m_0$ and $|G_{x} - \tau|_+ - \mathbb{E}[G_x - \tau|_+] \leq M_0 - \tau$ which implies that almost surely

$$|\phi(G_{x}, \tau) - \mathbb{E}[\phi(G_{x}, \tau)]| \leq M_1 := w_0 (M_0 - m_0) + \sum_{i=1}^{k} w_i \frac{1}{1-\alpha_i} (M_0 - \tau_i).$$

**Computation of $M_2$ and $L$.** We have $H_{x, \tau}^2(x, \tau, \xi) = [H_{x}^2(x, \tau, \xi); H_{x}^2(x, \tau, \xi)]$ with

$$H_{x}^2(x, \tau, \xi) = w_0 G_{x}^2(x, \xi) + \sum_{i=1}^{k} \frac{w_i}{1-\alpha_i} G_{x}^2(x, \xi) 1_{G(x, \xi) \geq \tau_i},$$

$$H_{x}^2(x, \tau, \xi) = (w_i (1 - \frac{1}{1-\alpha_i}) 1_{G(x, \xi) \geq \tau_i})_{i=1, \ldots, k}.$$

We assume that for every $x \in \mathcal{X}$, the stochastic subgradients $G_x(x, \xi)$ are almost surely bounded and we denote by $\mathcal{M}$ and $\overline{\mathcal{M}}$ vectors such that almost surely $\underline{m} \leq G_x(x, \xi) \leq \overline{M}$. Then for $(x, \tau) \in \mathcal{X} \times T$, setting $b_i = \max(w_0 \overline{M}_i, (w_0 + \sum_{j=1}^{k} \frac{w_j}{1-\alpha_j}) \overline{M}_i)$ and $a_i = \min(w_0 \underline{m}_i, (w_0 + \sum_{j=1}^{k} \frac{w_j}{1-\alpha_j}) \underline{m}_i)$, we have

$$||\mathbb{E}[H_{x, \tau}^2(x, \tau, \xi)]||_2^2 \leq L^2 := \sum_{i=1}^{m} \max(a_i^2, b_i^2) + \sum_{i=1}^{k} w_i^2 \max \left[ \frac{\alpha_i^2}{1-\alpha_i}, \frac{\alpha_i^2}{1-\alpha_i} \right],$$

$$\mathbb{E}[H_{x, \tau}^2(x, \tau, \xi) - \mathbb{E}[H_{x, \tau}^2(x, \tau, \xi)]]_2^2 \leq M_2^2 := \sum_{i=1}^{m} (a_i - b_i)^2 + \sum_{i=1}^{k} \frac{w_i^2}{1-\alpha_i}. $$

In some cases, the above formulas for $\tau, \tau, L, M_1$, and $M_2$ can be simplified:
Example 3. Let \( k = 1 \) in (53) and \( G(x, \xi) = \xi^T x \) where \( \xi \) is a random vector with mean \( \mu \) and covariance matrix \( \Sigma \). In this case \( \min_{x \in X} E[G_x] \) and \( \max_{x \in X} E[G_x] \) are convex optimization problems with linear objective functions and denoting by \( U_1 \) the quantity \( \max_{x \in X} \| x \|_1 \) or an upper bound on this quantity, we can replace \( \max_{x \in X} \text{Var}(G_x) \) by \( U_1^2 \max_i \| \Sigma(i, i) \| \) in the expressions of \( \tau \) and \( \bar{\tau} \). Computing \( M_0 \) and \( m_0 \) also amounts to solve convex optimization problems with linear objective. Assume also that almost surely \( \| \xi \|_\infty \leq U_2 \) for some \( 0 < U_2 < +\infty \). We have \( |G_x - E[G_x]| \leq 2U_1U_2 \) and \( \|G_x - \tau\|_\infty - \|E[G_x - \tau]\|_\infty \leq U_1U_2 - \tau \) which shows that we can take \( M_1 = 2w_0U_1U_2 + \frac{w_1}{1-\alpha_1} (U_1U_2 - \tau) \). We have \( \text{E}[H'_x(x, \tau, \xi)] = w_1(1 - \frac{P(\xi^T x \geq \tau)}{1-\alpha_1}) \) so that \( |\text{E}[H'_x(x, \tau, \xi)]| \leq w_1 \text{max}(1, \frac{\alpha_1}{1-\alpha_1}) \) and \( \text{E}[H'_x(x, \tau, \xi)] \leq m(w_0 + \frac{w_1}{1-\alpha_1})^2U_2^2 \), i.e., we can take \( L = w_1 \text{max}(1, \frac{\alpha_1}{1-\alpha_1}) + m(w_0 + \frac{w_1}{1-\alpha_1})^2U_2^2 \). Next, for all \( \xi_0 \in \Xi \) we have

\[
|H'_x(x, \tau, \xi_0) - \text{E}[H'_x(x, \tau, \xi)]| = \begin{cases} \frac{w_1(1-P(\xi^T x \geq \tau))}{1-\alpha_1} & \text{if } \xi_0^T x \geq \tau, \\ \frac{w_1P(\xi^T x \geq \tau)}{1-\alpha_1} & \text{otherwise,} \end{cases}
\]

implying that \( |H'_x(x, \tau, \xi_0) - \text{E}[H'_x(x, \tau, \xi)]| \leq \frac{w_1}{1-\alpha_1} \).

Since \( \|H'_x(x, \tau, \xi_0) - \text{E}[H'_x(x, \tau, \xi)]\|_\infty \) is bounded from above by 2(\( w_0 + \frac{w_1}{1-\alpha_1} \))\( U_2 \), we can take \( M_2 = \frac{w_1^2}{(1-\alpha_1)^2} + 4 \text{m}(w_0 + \frac{w_1}{1-\alpha_1})^2U_2^2 \). In the special case when \( X = \{x_0\} \) is a singleton, denoting \( \eta = \xi^T x_0 \), we have \( \vartheta = \mathcal{R}(\eta) \), \( H(x, \tau, \xi) = H(x_0, \tau, \xi) \), \( H_\tau(x, \tau, \xi) = 0 \) almost surely and the above computations show that we can take

\[
L = w_1 \text{max}(1, \frac{\alpha_1}{1-\alpha_1}), \quad M_1 = w_0(b_0 - a_0) + \frac{w_1}{1-\alpha_1}(b_0 - \tau), \quad \text{and } M_2 = \frac{w_1}{1-\alpha_1},
\]

where \( \tau = \mathcal{E}[\eta] - \sqrt{\frac{1-\alpha_1}{\alpha_1}} \sqrt{\text{Var}(\eta)} \) with \( a_0, b_0 \) satisfying \( a_0 \leq \eta \leq b_0 \) almost surely.

Discussion: asymptotic versus nonasymptotic tests and confidence intervals for the optimal value of (1). The nonasymptotic tests of this and the previous section do not require the independence of \( \xi^{(N,1)}, \ldots, \xi^{(N,N)} \) and are valid for any sample size \( N \). On the contrary, the asymptotic tests are valid as the sample size \( N \) goes to infinity and theory does not tell us for which values of \( N \) the Gaussian distribution “approximates well” the optimal value of SAA (2) of (1). Moreover, experiments in [8] and in the next section show that this value of \( N \) depends on dimension \( m \) of \( x \).

A (known) drawback of nonasymptotic confidence bounds is their conservativeness. On the one hand, this conservativeness allows us, when the sample size \( N \) is not much larger than problem dimension \( m \), to provide confidence sets of the prescribed risk, which asymptotic confidence intervals (based on the CLT of Section 3) fail to do, see [8]. On the other hand, for testing problems (3), (4), nonasymptotic rejection regions can lead to large probabilities of type II errors. Even if the asymptotic tests of Section 5.1 are valid as the sample size tends to infinity they can work well in practice for small sample sizes (\( N = 20 \)) and problems of small to moderate size (\( m \) up to 500); see the numerical simulations of Section 6. The derivation of less conservative nonasymptotic confidence sets (especially the lower bound) is an interesting future research goal.


6.1. Comparing the risk of two distributions: tests (3) with a singleton for \( \mathcal{X} \). We consider test (3) with \( V = 2 \) and \( \mathcal{X} \) a singleton. We use the rejection
regions given in Section 5.2 (resp. given by (43)) in the nonasymptotic (resp. asymptotic) case. In this situation, the test aims at comparing the risk of two distributions. We use the notation \( \mathcal{N}(m_0, \sigma^2; a_0, b_0) \) for the normal distribution with mean \( m_0 \) and variance \( \sigma^2 \) conditional on this random variable being in \([a_0, b_0]\) (truncated normal distribution with support \([a_0, b_0]\)). More precisely, we compare the risks \( R(\xi_1) \) and \( R(\xi_2) \) of two truncated normal (loss) distributions \( \xi_1 \) and \( \xi_2 \) with support \([a_0, b_0]\) in three cases: (I) \( \xi_1 \sim \mathcal{N}(10, 1; 0, 30), \xi_2 \sim \mathcal{N}(20, 1; 0, 30), \) (II) \( \xi_1 \sim \mathcal{N}(5, 1; 0, 30), \xi_2 \sim \mathcal{N}(10, 25; 0, 30), \) and (III) \( \xi_1 \sim \mathcal{N}(10, 49; 0, 30), \xi_2 \sim \mathcal{N}(14, 0.25; 0, 30). \) For these three cases, the densities of \( \xi_1 \) and \( \xi_2 \) are represented in Figure 1 (left for (I), middle for (II), right for (III)).

![Densities of truncated normal loss distributions](image)

*Fig. 1. Densities of truncated normal loss distributions \( \xi_1 \) and \( \xi_2 \). Left plot: \( \xi_1 \sim \mathcal{N}(10, 1; 0, 30) \) and \( \xi_2 \sim \mathcal{N}(20, 1; 0, 30) \). Middle plot: \( \xi_1 \sim \mathcal{N}(5, 1; 0, 30) \) and \( \xi_2 \sim \mathcal{N}(10, 25; 0, 30) \). Right plot: \( \xi_1 \sim \mathcal{N}(10, 49; 0, 30) \) and \( \xi_2 \sim \mathcal{N}(14, 0.25; 0, 30) \).*

We take for \( R \) the risk measure \( R(\xi) = w_0 E[\xi] + w_1 \text{AVaR}_\alpha(\xi) \) for \( 0 < \alpha < 1 \) where \( w_0, w_1 \geq 0 \) with \( w_0 + w_1 = 1 \). We assume that only the support \([a_0, b_0]\) of \( \xi_1 \) and \( \xi_2 \) and two samples \( \xi_1^N \) and \( \xi_2^N \) of size \( N \) of respectively \( \xi_1 \) and \( \xi_2 \) are known. Since the distribution of \( \xi \) has support \([a_0, b_0]\), we can write

\[
(57) \quad R(\xi) = \min_{\tau \in [a_0, b_0]} w_0 E[\xi] + w_1 \left( \tau + \frac{1}{1-\alpha} E[\xi - \tau]_+ \right)
\]

which is of form (1) with a risk-neutral objective function, \( G(\tau, \xi) = w_0 \xi + w_1 \tau + \frac{w_1}{1-\alpha} [\xi - \tau]_+ \) and \( \mathcal{X} \) the compact set \( \mathcal{X} = [a_0, b_0] = [0, 30] \).

It follows that the RSA algorithm can be used to estimate \( R(\xi_1) \) and \( R(\xi_2) \) and to compute the confidence bounds (49) and (50) with \( L, M_1, \) and \( M_2 \) given by (56). In these formulas, we replace \( \tau \) by its lower bound 0 since we do not assume the mean and standard deviation of \( \xi_1 \) and \( \xi_2 \) known. We obtain \( L = w_1 \max(1, \frac{\alpha}{1-\alpha}), M_2 = \frac{w_1}{1-\alpha}, \) and \( M_1 = 30(w_0 + \frac{w_1}{1-\alpha}) \).

**Case (I).** We first illustrate Corollary 3 computing the empirical estimation \( R(\hat{F}_{N,1}) \) of \( R(\xi_1) \) on 200 samples of size \( N \) of \( \xi_1 \sim \mathcal{N}(10, 1; 0, 30) \) for \( w_0 = 0.1, w_1 = 0.9, \) and various values of \( \alpha \) and of the sample size \( N \). For this experiment, the QQ-plots of the empirical distribution of \( R(\hat{F}_{N,1}) \) versus the normal distribution with parameters the empirical mean and standard deviation of this empirical distribution are reported in the supplementary materials of this article. We see that even for small values of \( 1-\alpha \) and \( N \) as small as 20, the distribution of \( R(\hat{F}_{N,1}) \) is well approximated by a Gaussian distribution: for \( N = 20 \) the Jarque-Bera test accepts the hypothesis of normality at the significance level 0.05 for \( 1-\alpha = 0.01 \) and \( 1-\alpha = 0.5 \).

We fix again the distribution \( \xi_1 \sim \mathcal{N}(10, 1; 0, 30) \) and approximately compute \( R(\xi_1) \) for various values of \( (w_0, w_1, \alpha, N) \) using the RSA and SAA methods on samples \( \xi_1^N \) of size \( N \) of \( \xi_1 \). For a sample of size \( N \) of \( \xi_1 \), let \( R_{N,\text{RSA}}(\xi_1) \) and \( R_{N,\text{SAA}}(\xi_1) = R(\hat{F}_{N,1}) \) be these estimations using respectively RSA and SAA. For fixed \( (w_0, w_1, \alpha, N) \),
Using RSA, unless the sample size is very large. More precisely, we compute the nonasymptotic tests of Section 5.2 based on the confidence intervals computed especially when $1 - \alpha$ decreases with the sample size $N$, as expected. We also naturally observe that the more weight is given to the A\VaR and the smaller $1 - \alpha$ the more difficult it is to estimate the risk measure, i.e., the more distant the expectation of the approximation is to the optimal value and the larger the sample size needs to be to obtain an expected approximation with given accuracy.

We now study for case (I) the test
\begin{equation}
H_0 : \mathcal{R}(\xi_1) = \mathcal{R}(\xi_2) \quad \text{against} \quad H_1 : \mathcal{R}(\xi_1) \neq \mathcal{R}(\xi_2).
\end{equation}

We fix $\alpha = 0.1$ for the maximal type I error and $1 - \alpha = 0.1$. Since in case (I) we have $\mathcal{R}(\xi_1) \neq \mathcal{R}(\xi_2)$ (see Figure 1), from this experiment we expect to obtain a large probability of type II error using the nonasymptotic tests of Section 5.2 based on the confidence intervals computed using RSA, unless the sample size is very large. More precisely, we compute the nonasymptotic confidence interval is given by (49)-(50). Recalling that $\mathcal{R}(\xi)$ is the optimal value of optimization problem (57) which is of the form (1), we compute for $\mathcal{R}(\xi)$ the asymptotic confidence interval $\left[\hat{\vartheta}_N - \Phi^{-1}(1 - \beta/2) \sqrt{\frac{\hat{\vartheta}_N}{N}}, \hat{\vartheta}_N + \Phi^{-1}(1 - \beta/2) \sqrt{\frac{\hat{\vartheta}_N}{N}}\right]$, where $\hat{\vartheta}_N$ is the optimal value of the SAA of (57). Note that in this case the optimal value $\hat{\vartheta}_N$ of the SAA problem is the $\alpha$-quantile of the distribution of $\xi$ (no optimization step is necessary to solve the SAA problem).

| Sample size $N$ | Sample size $N$ | \begin{smallmatrix} (w_0, w_1) \\ \alpha, N \end{smallmatrix} | \begin{smallmatrix} \text{Method} \\ 1 - \alpha \end{smallmatrix} | \begin{array}{cccccc}
20 & 50 & 10^3 & 10^4 & 10^5 & 10^6 \\
\hline
\alpha, N & 1 - \alpha & \text{Method} & 20 & 50 & 10^3 & 10^4 & 10^5 & 10^6 \\
\hline
(0.1, 0.9) & 10^{-2} & \text{SAA} & 11.71 & 12.00 & 12.21 & 12.37 & 12.40 & 12.40 & 12.40 \\
(0.1, 0.9) & 10^{-2} & \text{RSA} & 14.35 & 14.26 & 14.16 & 13.46 & 12.75 & 12.51 & 12.43 \\
(0.1, 0.9) & 0.1 & \text{SAA} & 11.51 & 11.50 & 11.54 & 11.58 & 11.58 & 11.58 & 11.58 \\
(0.1, 0.9) & 0.1 & \text{RSA} & 20.50 & 16.78 & 15.10 & 12.61 & 11.90 & 11.68 & 11.61 \\
(0.1, 0.9) & 0.5 & \text{SAA} & 10.71 & 10.69 & 10.72 & 10.72 & 10.72 & 10.72 & 10.72 \\
(0.1, 0.9) & 0.5 & \text{RSA} & 11.42 & 11.12 & 11.02 & 10.81 & 10.75 & 10.73 & 10.72 \\
(0.9, 0.1) & 10^{-2} & \text{RSA} & 10.19 & 10.23 & 10.25 & 10.26 & 10.27 & 10.27 & 10.27 \\
(0.9, 0.1) & 10^{-2} & \text{SAA} & 10.49 & 10.48 & 10.47 & 10.38 & 10.31 & 10.28 & 10.27 \\
(0.9, 0.1) & 0.1 & \text{SAA} & 10.17 & 10.16 & 10.19 & 10.18 & 10.18 & 10.18 & 10.18 \\
(0.9, 0.1) & 0.1 & \text{RSA} & 10.34 & 10.28 & 10.27 & 10.20 & 10.18 & 10.18 & 10.18 \\
(0.9, 0.1) & 0.5 & \text{SAA} & 10.09 & 10.07 & 10.08 & 10.08 & 10.08 & 10.08 & 10.08 \\
(0.9, 0.1) & 0.5 & \text{RSA} & 10.17 & 10.11 & 10.12 & 10.09 & 10.08 & 10.08 & 10.08 \\
\end{array}
\end{table}

\text{Table 1: Estimation of the risk measure value $\mathcal{R}(\xi_1)$ for $\xi_1 \sim \mathcal{N}(10, 1; 0, 30)$ using SAA and RSA for various values of $(w_0, w_1, \alpha)$ and various sample sizes $N$.}
probability of type II error for (58) considering asymptotic and nonasymptotic rejection regions using various sample sizes \( N \in \{20, 50, 100, 1000, 5000, 10000, 20000, 50000, 100000, 150000\} \), taking \( 1 - \alpha = 0.1 \) and \( (w_0, w_1) \in \{(0, 1), (0.1, 0.9), (0.2, 0.8), (0.3, 0.7), (0.4, 0.6), (0.5, 0.5), (0.6, 0.4), (0.7, 0.3), (0.8, 0.2), (0.9, 0.1)\} \). For fixed \( N \), the probability of type II error is estimated using 100 samples of size \( N \) of \( \xi_1 \) and \( \xi_2 \). Using the asymptotic rejection region, we reject \( H_0 \) for all realizations and all parameter combinations, meaning that the probability of type II error is null (since \( H_1 \) holds for all parameter combinations). For the nonasymptotic test, the probabilities of type II errors are reported in Table 4. For sample sizes less than 5,000, the probability of type II error is always 1 (the nonasymptotic test always takes the wrong decision) and the larger \( w_1 \) the larger the sample size \( N \) needs to be to obtain a probability of type II error of zero. In particular, if \( w_1 = 1 \) (we estimate the AVaR of the distribution) as much as 150,000 observations are needed to obtain a null probability of type II error. However, if the sample size is sufficiently large, both tests always take the correct decision \( \mathcal{R}(\xi_1) \neq \mathcal{R}(\xi_2) \).

Given (possibly small) samples of size \( N \) of \( \xi_1 \) and \( \xi_2 \), to know which of the two risks \( \mathcal{R}(\xi_1) \) and \( \mathcal{R}(\xi_2) \) is the smallest, we now consider the test

\[
H_0 : \mathcal{R}(\xi_1) \geq \mathcal{R}(\xi_2) \quad \text{against} \quad H_1 : \mathcal{R}(\xi_1) < \mathcal{R}(\xi_2).
\]

Computing \( \mathcal{R}(\xi_1) \) and \( \mathcal{R}(\xi_2) \) with a very large sample (of size \( 10^6 \)) of \( \xi_1 \) and \( \xi_2 \) either with SAA or RSA or looking at Figure 1, we know that \( \mathcal{R}(\xi_1) < \mathcal{R}(\xi_2) \). We again analyze the probability of type II error using the asymptotic and nonasymptotic rejection regions when the decision is taken on the basis of a much smaller sample.

\(^3\)All computed probabilities of type II error are empirical probabilities. However, for short, we will use in the sequel the term probabilities of type II error.
For the nonasymptotic test, the empirical probabilities of type II error for various sample sizes (estimated, for fixed $N$, using 100 samples of size $N$ of $\xi_1$ and $\xi_2$) are exactly those obtained for test (58) and are given in Table 4. The asymptotic test again always takes the correct decision $R(\xi_1) < R(\xi_2)$ while a large sample size is needed to always take the correct decision using the nonasymptotic test, as large as 150,000 for $w_1 = 1$.

We now consider tests (58) and (59) for case (II). In this case, there is a larger overlap between the distributions of $\xi_1$ and $\xi_2$. However, from Figure 1 and computing $R(\xi_1)$ and $R(\xi_2)$ with a very large sample (say of size $10^6$) of $\xi_1$ and $\xi_2$ either using SAA or RSA, we check that we have again $R(\xi_2) > R(\xi_1)$ for all values of $(w_0, w_1)$.

The empirical probabilities of type II error are null for the asymptotic test for all sample sizes $N$ tested while for the nonasymptotic test, the probabilities of type II error are given in Table 5 for both tests (58) and (59). As a result, here again, the asymptotic test always takes the correct decision $R(\xi_1) < R(\xi_2)$ while a large sample size is needed to always take the correct decision using the nonasymptotic test (as large as 110,000 for $w_1 = 1$). For sample sizes less than 10,000, the empirical probability of type II error with the nonasymptotic test is 1. We see that for fixed $(w_0, w_1)$, in most cases, we need a larger sample size than in case (I) to have a null probability of type II error, due the overlap of the two distributions.

We finally consider Case (III) where the choice between $\xi_1$ and $\xi_2$ is more delicate and depends on the pair $(w_0, w_1)$. In this case, we have (see Figure 1) $E[\xi_2] > E[\xi_1]$ and $AVaR_\alpha(\xi_2) < AVaR_\alpha(\xi_1)$ for $1 - \alpha = 0.1$. It follows that for pairs $(w_0, w_1)$ summing to one, when

$$0 \leq w_0 < w_{\text{crit}} = \frac{AVaR_\alpha(\xi_1) - AVaR_\alpha(\xi_2)}{E[\xi_2] - E[\xi_1] + AVaR_\alpha(\xi_1) - AVaR_\alpha(\xi_2)}$$

then $R(\xi_2) < R(\xi_1)$ and for $w_0 > w_{\text{crit}}$ then $R(\xi_2) > R(\xi_1)$. The empirical estimation of $w_{\text{crit}}$ (estimated using a sample of size $10^6$) is 0.71. For $w_0$ close to $w_{\text{crit}}$, $R(\xi_1)$ and $R(\xi_2)$ are close and the probability of type II error for test (58) can be large even for the asymptotic test if the sample size is not sufficiently large. More precisely, for the asymptotic test, when $(w_0, w_1) = (0.7, 0.3)$, the empirical probabilities of type II error are given in Table 6 for $N \in \{20, 50, 100, 200, 500, 1,000, 2,000, 5,000\}$, and are 0.28, 0.11, 0.01, and 0 for respectively $N = 10,000, 20,000, 40,000$, and 45,000. For the

<table>
<thead>
<tr>
<th>$(w_0, w_1)$</th>
<th>5,000</th>
<th>10,000</th>
<th>20,000</th>
<th>50,000</th>
<th>100,000</th>
<th>130,000</th>
<th>150,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.0, 1.0)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(0.1, 0.9)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(0.2, 0.8)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(0.3, 0.7)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(0.4, 0.6)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(0.5, 0.5)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(0.6, 0.4)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(0.7, 0.3)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(0.8, 0.2)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(0.9, 0.1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4

Empirical probabilities of type II error for tests (58) and (59) using a nonasymptotic rejection region when $\xi_1 \sim N(10, 1; 0, 30)$, $\xi_2 \sim N(20, 1; 0, 30)$, and $1 - \alpha = 0.1$. 

23
remaining values of $w_0$ the empirical probabilities of type II error are given in Table 6 for the asymptotic test. For the nonasymptotic test, the empirical probabilities of type II error for test (58) are given in Table 7. It is seen that much larger sample sizes are needed in this case to obtain a small probability of type II error. However, for the sample size $N = 5 \times 10^6$, the nonasymptotic test still always takes the wrong decision for the difficult case $w_0 = 0.7$.

For $w_0 < w_{\text{crit}}$ with $w_0 \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$, we are interested in the probability of type II error of the test

\begin{equation}
H_0 : R(\xi_2) \geq R(\xi_2) \quad \text{against} \quad H_1 : R(\xi_2) < R(\xi_1)
\end{equation}

since $H_1$ holds in this case. Using the asymptotic rejection region, except for the difficult case $w_0 = 0.7$ where the probability of type II error is still positive for $N = 30000$, the empirical probability of type II error is null for small to moderate (at most 1000) sample sizes; see Table 8. Using the nonasymptotic rejection region, much larger sample sizes are necessary to obtain a small probability of type II error, see Table 9.

For $w_0 > w_{\text{crit}}$ with $w_0 \in \{0.8, 0.9\}$, we are interested in the probability of type II error of test (59) since $H_1$ holds in this case. The probability of type II error for
This test using the nonasymptotic rejection region is 1 (resp. 0) for $(N, w_0, w_1) = (10^6, 0.8, 0.2)$ (resp. $(N, w_0, w_1) = (10^6, 0.9, 0.1)$), and null for $(N, w_0, w_1) = (5 \times 10^6, 0.8, 0.2), (5 \times 10^6, 0.9, 0.1)$, meaning that we always take the correct decision $R(\xi_1) < R(\xi_2)$ for $N = 5 \times 10^6$ and $(w_0, w_1) = (0.8, 0.2), (0.9, 0.1)$. Using the asymptotic rejection region, the probabilities of type II errors are null already for $N = 1000$. For $N = 100$, we get probabilities of type II error of 0.09 and 0.42 for respectively $(w_0, w_1) = (0.8, 0.2)$ and $(w_0, w_1) = (0.9, 0.1)$.

### 6.2. Tests on the optimal value of two risk averse stochastic programs.

We illustrate the results of Section 3 on the risk averse problem

$$\min_{w_0} E[\sum_{i=1}^m \xi_i x_i] + w_1 \left( x_0 + \mathbb{E} \left[ \frac{1}{1-\alpha} \sum_{i=1}^m \xi_i x_i - x_0 \right] \right) + \lambda_0 \| \langle x_0; x_1; \ldots; x_m \rangle \|_2^2 + c_0$$

$$-1 \leq x_0 \leq 1, \sum_{i=1}^m x_i = 1, x_i \geq 0, i = 1, \ldots, m,$$

where $\xi$ is a random vector with i.i.d. Bernoulli entries: $\mathbb{P}(\xi_i = 1) = \Psi_i$, $\mathbb{P}(\xi_i = -1) = 1 - \Psi_i$, with $\Psi_i$ randomly drawn over $[0, 1]$. This problem amounts to minimizing a linear combination of the expectation and the AVaR of $\sum_{i=1}^m \xi_i x_i$ plus a penalty obtained taking $\lambda_0 > 0$. Therefore, it has a unique optimal solution. SAA formulation

---

### Table 7

<table>
<thead>
<tr>
<th>$(w_0, w_1)$</th>
<th>20</th>
<th>100</th>
<th>200</th>
<th>1000</th>
<th>5000</th>
<th>10000</th>
<th>30000</th>
<th>50000</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.0, 1.0)</td>
<td>0.11</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.1, 0.9)</td>
<td>0.26</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.2, 0.8)</td>
<td>0.28</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.3, 0.7)</td>
<td>0.35</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.4, 0.6)</td>
<td>0.51</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.5, 0.5)</td>
<td>0.60</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.6, 0.4)</td>
<td>0.83</td>
<td>0.53</td>
<td>0.22</td>
<td>0.81</td>
<td>0.81</td>
<td>0.61</td>
<td>0.39</td>
<td>0.05</td>
</tr>
<tr>
<td>(0.7, 0.3)</td>
<td>0.87</td>
<td>0.88</td>
<td>0.90</td>
<td>0.39</td>
<td>0.39</td>
<td>0.05</td>
<td>0.05</td>
<td>0.00</td>
</tr>
</tbody>
</table>

### Table 8

<table>
<thead>
<tr>
<th>$(w_0, w_1)$</th>
<th>20</th>
<th>100</th>
<th>200</th>
<th>1000</th>
<th>5000</th>
<th>10000</th>
<th>30000</th>
<th>50000</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.0, 1.0)</td>
<td>0.11</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.1, 0.9)</td>
<td>0.26</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.2, 0.8)</td>
<td>0.28</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.3, 0.7)</td>
<td>0.35</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.4, 0.6)</td>
<td>0.51</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.5, 0.5)</td>
<td>0.60</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>(0.6, 0.4)</td>
<td>0.83</td>
<td>0.53</td>
<td>0.22</td>
<td>0.81</td>
<td>0.81</td>
<td>0.61</td>
<td>0.39</td>
<td>0.05</td>
</tr>
<tr>
<td>(0.7, 0.3)</td>
<td>0.87</td>
<td>0.88</td>
<td>0.90</td>
<td>0.39</td>
<td>0.39</td>
<td>0.05</td>
<td>0.05</td>
<td>0.00</td>
</tr>
</tbody>
</table>
large samples are needed to obtain a good approximation with RSA. We also report in SAA, the optimal value is already well approximated with small sample sizes while values increase (resp. decrease) with the sample size for SAA (resp. RSA). With \( \xi \), the Jarque-Bera test accepts the null hypothesis (the data comes from a normal distribution with unknown mean and variance) at the 5% significance level.

We first compare the estimation of the optimal value of \( I_2 \) using RSA and SAA. We observe again that this distribution is well approximated by a Gaussian distribution even when the sample size is small (\( N = 20 \)): for all problem sizes (\( m = 100, m = 500, m = 10^3 \), and \( m = 10^4 \)) and the smallest sample size tested (\( N = 20 \)), the Jarque-Bera test accepts the null hypothesis (the data comes from a normal distribution with unknown mean and variance) at the 5% significance level.

We now define in Table 10 six instances \( I_1, I_2, I_3, I_4, I_5, \) and \( I_6 \) of problem (61). \( \Psi_1 \) and \( \Psi_2 \) are vectors with entries drawn independently and randomly over \([0, 1] \).

\[
\begin{array}{cccccc}
\text{Instance} & (w_0, w_1, 1 - \alpha, \lambda_0) & c_0 & m & (P(\xi_i = 1))_i \\
I_1 & (0.9, 0.1, 0.1, 2) & 0 & 100 & \Psi_1 \\
I_2 & (0.9, 0.1, 0.1, 2) & 0 & 100 & 0.8\Psi_1 \\
I_3 & (0.9, 0.1, 0.1, 2) & -3 & 100 & 0.8\Psi_1 \\
I_4 & (0.9, 0.1, 0.1, 2) & 0 & 500 & \Psi_2 \\
I_5 & (0.9, 0.1, 0.1, 2) & 0 & 500 & 0.8\Psi_2 \\
I_6 & (0.9, 0.1, 0.1, 2) & -3 & 500 & 0.8\Psi_2 \\
\end{array}
\]

Definition of instances \( I_1, I_2, I_3, I_4, I_5, \) and \( I_6 \) of problem (61) (\( \Psi_1 \) and \( \Psi_2 \) are vectors with entries drawn independently and randomly over \([0, 1] \)).

We now define in Table 10 six instances \( I_1, I_2, I_3, I_4, I_5, \) and \( I_6 \) of problem (61). We first compare the estimation of the optimal value of \( I_2 \) using RSA and SAA. For the RSA algorithm, we take \( ||·|| = ||·||_2 = ||·||_\alpha \), and (see [7]) \( M_1 = 2(w_0 + \frac{\sqrt{1 - \alpha}}{1 - \alpha}) \), \( L = \sqrt{\left(\frac{w_1}{1 - \alpha}\right)^2 + m(w_0 + \frac{w_1}{1 - \alpha})^2 + 2\lambda_0} \), \( M_2 = \sqrt{\left(\frac{w_1}{1 - \alpha}\right)^2 + m(w_0 + \frac{w_1}{1 - \alpha})^2} \). The average approximate optimal value of instance \( I_2 \) (averaging taking 100 samples of \( \xi^N \)) using RSA and SAA is given in Table 11 for various sample sizes \( N \). These values increase (resp. decrease) with the sample size for SAA (resp. RSA). With SAA, the optimal value is already well approximated with small sample sizes while large samples are needed to obtain a good approximation with RSA. We also report in

<table>
<thead>
<tr>
<th>Sample size ( N )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 300,000 )</td>
</tr>
<tr>
<td>( 400,000 )</td>
</tr>
<tr>
<td>( 500,000 )</td>
</tr>
<tr>
<td>( 700,000 )</td>
</tr>
<tr>
<td>( 900,000 )</td>
</tr>
<tr>
<td>( 2 \times 10^6 )</td>
</tr>
<tr>
<td>( 5 \times 10^6 )</td>
</tr>
</tbody>
</table>

\( \Psi \)

<table>
<thead>
<tr>
<th>( (w_0, w_1) )</th>
<th>( 300,000 )</th>
<th>( 400,000 )</th>
<th>( 500,000 )</th>
<th>( 700,000 )</th>
<th>( 900,000 )</th>
<th>( 2 \times 10^6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (0.0, 1.0) )</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>( (0.1, 0.9) )</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>( (0.2, 0.8) )</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( (0.3, 0.7) )</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( (0.4, 0.6) )</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( (0.5, 0.5) )</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( (0.6, 0.4) )</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>( (0.7, 0.3) )</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 12 the average values of the asymptotic and nonasymptotic confidence bounds (computed using 100 samples of $\xi^N$) on the optimal values of instances $I_1$ and $I_2$ and various sample sizes. Knowing that the optimal values of $I_1$ and $I_2$, estimated using SAA with a sample of size $10^6$, are respectively $\hat{\varphi}_1 = -0.6515$ and $\hat{\varphi}_2 = -0.6791$, we observe that the asymptotic confidence interval is in mean much closer to the optimal value and of small width while large samples are needed to obtain a nonasymptotic confidence interval of small width. However, the confidence bounds on the optimal value obtained using RSA are almost independent on the problem size and as for the one dimensional problem of the previous section the sample size $N = 10^5$ provides confidence intervals of small width and allows us to have small probabilities of type I and type II errors for nonasymptotic tests on the optimal value of two instances of (61) if their optimal values are sufficiently distant (see Lemma 6). To check that and the superiority of the asymptotic tests for problems of moderate sizes ($m = 100$ and $m = 500$), we compare the empirical probabilities of type II error of several tests of form (3) with $K = 2$ for which $H_1$ holds and where $\varphi_i$ is the optimal value of instance $I_i$.

More precisely, the empirical probabilities of type II error of asymptotic and nonasymptotic tests of form

$$H_0 : \varphi_p = \varphi_q \text{ against } H_1 : \varphi_p \neq \varphi_q,$$

are reported in Table 13 (for all these tests, we check that $H_1$ holds computing $\varphi_i$, solving the SAA problem of instance $I_i$ with a sample of $\xi$ of size $10^6$: $\varphi_1 = -0.6515$, $\varphi_2 = -0.6791$, $\varphi_3 = -3.6791$, $\varphi_4 = -0.7725$, $\varphi_5 = -0.7868$, and $\varphi_6 = -3.7868$).

Table 11
<table>
<thead>
<tr>
<th>Method</th>
<th>$N = 20$</th>
<th>$N = 50$</th>
<th>$N = 10^3$</th>
<th>$N = 10^4$</th>
<th>$N = 10^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAA</td>
<td>-0.7205</td>
<td>-0.6965</td>
<td>-0.6883</td>
<td>-0.6799</td>
<td>-0.6791</td>
</tr>
<tr>
<td>RSA</td>
<td>-0.4615</td>
<td>-0.5274</td>
<td>-0.5646</td>
<td>-0.6389</td>
<td>-0.6654</td>
</tr>
</tbody>
</table>

Average values of the asymptotic and nonasymptotic confidence bounds (computed using 100 samples of $\xi^N$) for instances $I_1$ and $I_2$ and various sample sizes. For instance $I_i$, the average asymptotic confidence interval is $[L-\text{NA}1, U-\text{NA}1]$ and the average nonasymptotic confidence interval is $[L-\text{NA}2, U-\text{NA}2]$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$L-A1$</th>
<th>$U-A1$</th>
<th>$L-\text{NA}1$</th>
<th>$U-\text{NA}1$</th>
<th>$L-A2$</th>
<th>$U-A2$</th>
<th>$L-\text{NA}2$</th>
<th>$U-\text{NA}2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>-0.7207</td>
<td>-0.6666</td>
<td>-95.7926</td>
<td>2.5227</td>
<td>-0.7443</td>
<td>-0.6967</td>
<td>-95.8354</td>
<td>2.4799</td>
</tr>
<tr>
<td>50</td>
<td>-0.6888</td>
<td>-0.6475</td>
<td>-60.8057</td>
<td>1.3743</td>
<td>-0.7148</td>
<td>-0.6781</td>
<td>-60.8472</td>
<td>1.3329</td>
</tr>
<tr>
<td>$10^3$</td>
<td>-0.6752</td>
<td>-0.6444</td>
<td>-43.1779</td>
<td>0.7900</td>
<td>-0.7019</td>
<td>-0.6746</td>
<td>-43.2171</td>
<td>0.7508</td>
</tr>
<tr>
<td>$10^4$</td>
<td>-0.6573</td>
<td>-0.6474</td>
<td>-14.0952</td>
<td>-0.1913</td>
<td>-0.6843</td>
<td>-0.6755</td>
<td>-14.1269</td>
<td>-0.2230</td>
</tr>
<tr>
<td>$10^5$</td>
<td>-0.6532</td>
<td>-0.6501</td>
<td>-4.9019</td>
<td>-0.5051</td>
<td>-0.6805</td>
<td>-0.6777</td>
<td>-4.9307</td>
<td>-0.5339</td>
</tr>
</tbody>
</table>

Table 12

Computing $e^x$ is done using the log of the input (and the output).
is much lower than the coverage probability of the nonasymptotic confidence interval and than the target coverage probability, the asymptotic confidence bounds are much closer to each other and much closer to the optimal value than the nonasymptotic confidence bounds. This explains why the probability of type II error of the asymptotic test is much less than the probability of type II error of the nonasymptotic test, even for small sample sizes and a smaller sample is needed to always take the correct decision $H_1$ with the asymptotic test, i.e., to obtain a null probability of type II error. Of course, in both cases, for fixed $N$, the empirical probability of type II error depends on the distance between $\vartheta_p$ and $\vartheta_q$.

Similar conclusions can be drawn from Table 14 which reports the empirical probability of type II error for various tests of form

$$H_0 : \vartheta_p \leq \vartheta_q \quad \text{against} \quad H_1 : \vartheta_q < \vartheta_p.$$ 

In particular, from these results, we see that we always take the correct decision $H_1$ with the asymptotic test for sample sizes above $N = 100$.

REFERENCES


---

Table 13

<table>
<thead>
<tr>
<th>Sample size $N$</th>
<th>$H_0$</th>
<th>$H_1$</th>
<th>Test type</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>1000</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\vartheta_1 = \vartheta_2$</td>
<td>Asymptotic</td>
<td>0.72</td>
<td>0.45</td>
<td>0.29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\vartheta_1 = \vartheta_3$</td>
<td>Nonasymptotic</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\vartheta_1 = \vartheta_4$</td>
<td>Asymptotic</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\vartheta_1 = \vartheta_5$</td>
<td>Nonasymptotic</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\vartheta_1 = \vartheta_6$</td>
<td>Nonasymptotic</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 14

<table>
<thead>
<tr>
<th>Sample size $N$</th>
<th>$H_0$</th>
<th>$H_1$</th>
<th>Test type</th>
<th>20</th>
<th>50</th>
<th>100</th>
<th>1000</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\vartheta_1 \leq \vartheta_2$</td>
<td>Asymptotic</td>
<td>0.54</td>
<td>0.38</td>
<td>0.16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\vartheta_1 \leq \vartheta_3$</td>
<td>Nonasymptotic</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\vartheta_1 \leq \vartheta_4$</td>
<td>Asymptotic</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\vartheta_1 \leq \vartheta_5$</td>
<td>Nonasymptotic</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\vartheta_1 \leq \vartheta_6$</td>
<td>Nonasymptotic</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Empirical probabilities of type II error for tests of form (62).


