Available online at www.sciencedirect.com
ScienceDirect
Operations Research Letters 35 (2007) 722–730
Operations Research Letters
www.elsevier.com/locate/orl

Large deviations bounds for estimating conditional value-at-risk

David B. Brown

The Fuqua School of Business, Duke University, 1 Towerview Drive, Durham, NC 27708, USA

Received 11 December 2006; accepted 2 January 2007
Available online 11 January 2007

Abstract

In this paper, we prove an exponential rate of convergence result for a common estimator of conditional value-at-risk for bounded random variables. The bound on optimistic deviations is tighter while the bound on pessimistic deviations is more general and applies to a broader class of convex risk measures.

Keywords: Conditional value-at-risk; Convex risk measure; Optimized certainty equivalent; Large deviations; Estimation

1. Introduction

The use of value-at-risk (VaR) as a risk measure has been the subject of significant criticism in recent years (e.g., [7,9]), primarily due to the facts that VaR does not properly account for risk diversification and that it says nothing about the magnitude of losses beyond the quantile level in question.

As a result of these drawbacks associated with VaR, the class of *coherent* risk measures was axiomatized and popularized by Artzner et al. [2] and Delbaen [8]. Of particular interest within this class of risk measures is the *conditional value-at-risk* (CVaR) risk measure (see, [14,1]), which has received considerable attention within the mathematical finance community.

Despite the interest in coherent risk measures, they have in turn received criticism because the size of such risk measures grows linearly with the size of positions, thereby ruling out many of the inherently nonlinear, certainty equivalent-type risk measures suggested by traditional utility theory [16]. Accordingly, the axioms of coherent risk measures were relaxed by Föllmer and Schied [10], who introduced the class of *convex* risk measures. These risk measures allow the use of a wide-range of nonlinear, certainty equivalent measures, including the class of *optimized certainty equivalent* (OCE) risk measures introduced by Ben-Tal and Teboulle [3,4,6], who also study these risk measures in the context of portfolio theory [5]. It is known that CVaR is in fact an OCE with a particular, piecewise-linear loss function [6].

Our focus in this paper will be primarily on CVaR and the practically relevant issue of how to estimate it from samples. While the *asymptotic* convergence properties of various estimators for CVaR have been investigated (e.g., [1]), less is known about the

E-mail address: dbbrown@duke.edu.

0167-6377/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.orl.2007.01.001

finite-sample convergence properties for estimators. Takeda and Kanamori [15] have recently shown some finite convergence results for CVaR; although these results apply to the case of optimization of CVaR, the bounds rely on statistical learning and, as a result, suffer from the conservatism of this theory.

Here we will use the classical results of McDiarmid [12] and Hoeffding [11] to derive large deviation convergence bounds for estimating CVaR from a finite number of independent samples. Our results apply to bounded random variables representing an uncertain loss (hence, we will, at times, refer to estimates that are larger than the true value to be "pessimistic" and those that are smaller than the true value to be "optimistic"). The estimator we will use is the intuitive one based on fitting the distribution via the "method of moments" and discussed in many treatments of CVaR (e.g., [14]).

For this estimator, with N independent samples, we show the following:

1. The probability that an estimate of $\mathrm{CVaR}_\alpha(X)$ is pessimistic by an amount ε decays exponentially in $\alpha^2\varepsilon^2 N$. This bound applies to more general OCE risk measures as well, and we show the CVaR bound as a special case of it.
2. When X, in addition, has a continuous probability distribution, the probability that an estimate is optimistic by an amount ε decays exponentially in $\alpha\varepsilon^2 N$.

Notice that the bound on optimistic (lower) deviations is tighter than the bound on pessimistic deviations, but less general in that it only applies to CVaR. Although we have yet to show it, we believe a similar bound decaying exponentially only in $\alpha\varepsilon^2 N$ should apply to the pessimistic estimates as well. Furthermore, though we also do not show it, we believe the bound on optimistic estimates is tight up to constants.

For a real-valued function $f : \mathbb{R}^n \to \mathbb{R}$, recall the definition of the conjugate is $f^*(\boldsymbol{y}) \triangleq \sup_{\boldsymbol{x}\in \mathrm{dom}\, f}\{\boldsymbol{y}'\boldsymbol{x} - f(\boldsymbol{x})\}$.

2. Background

Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and let $\mathscr{X}$ be a set of real-valued random variables on Ω, i.e., a set of functions $X : \Omega \to \mathbb{R}$. Here, X represents an uncertain *loss*. We have the following definition.

Definition 2.1. A *risk measure* is a function $\mu : \mathscr{X} \to \mathbb{R}$.

A risk measure μ induces a preference order $\succcurlyeq_\mu$ over random variables in $\mathscr{X}$. Specifically, we have, for all $X_1, X_2 \in \mathscr{X}$, $X_1 \succcurlyeq_\mu X_2 \Leftrightarrow \mu(X_1) \leqslant \mu(X_2)$.

We begin by recalling a class of *certainty equivalent* risk measures introduced by Ben-Tal and Teboulle [3] and further developed in [4,6].

Definition 2.2. Let Φ be the class of all functions $\phi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ which are closed, convex, have a minimum value of 0 attained at 1, and satisfy $\mathrm{dom}\,\phi \subseteq \mathbb{R}_+$. The OCE of a random variable $X \in \mathscr{X}$ under $\phi \in \Phi$ is

$$\mu_\phi(X) \triangleq \inf_{v\in\mathbb{R}}\{v + \mathbb{E}[\phi^*(X - v)]\}. \quad (1)$$

The OCE can be interpreted as the value obtained by optimally paying an uncertain debt X between two periods. The sure amount v is paid today, and the remainder, $X - v$, is paid later. The loss function ϕ^* captures the "present value" of this uncertain, future debt. μ_ϕ is then the minimum possible present value over all possible two-period payment plans; in other words, computation of the risk measure μ_ϕ is *itself* an optimization problem.

The framework for OCEs in [6] is actually developed in terms of gains and a concave utility function u. Specifically, the authors consider the measure $S_u(X) = \sup_{v\in\mathbb{R}}\{v + \mathbb{E}[u(X - v)]\}$, which is perhaps more easily interpreted than μ_ϕ as the optimal two-period *consumption* of an uncertain *income* X with u representing the utility of the future income $X - v$. It is easy to show that $\mu_\phi(X) = -S_u(-X)$ when $\phi^*(t) = -u(-t)$. Therefore, μ_ϕ is a simple dual of S_u with appropriate sign changes to measure losses rather than gains.

Some examples of OCEs are the following (here $\mathrm{dom}\,\phi = \mathbb{R}_+$ is implicitly assumed):

Exponential loss:

$$[\phi(t) = t\log t - t + 1] \;\Rightarrow\; [\phi^*(u) = \mathrm{e}^u - 1]$$

$$\Rightarrow [\mu_\phi(X) = \log \mathbb{E}[\mathrm{e}^X]].$$

Quadratic loss: For a random variable $X \geqslant -1$,

$$[\phi(t) = \tfrac{1}{2}(t-1)^2] \Rightarrow \left[\phi^*(u) = \begin{cases} \frac{1}{2}u^2 + u & \text{if } u \geqslant -1 \\ -\frac{1}{2} & \text{otherwise} \end{cases}\right] \Rightarrow [\mu_\phi(X) = \mathbb{E}[X] + \tfrac{1}{2}\sigma^2(X)].$$

CVaR also fits in as a special case of the OCE risk measure.

Definition 2.3. For $\alpha \in (0,1]$, the *conditional value-at-risk at level* α of a random variable $X \in \mathscr{X}$ is

$$\mathrm{CVaR}_\alpha(X) \triangleq \inf_v \left\{ v + \frac{1}{\alpha}\mathbb{E}[(X-v)^+] \right\}. \tag{2}$$

Thus, as detailed in [6], CVaR is a special case of the OCE measure with a particular, piecewise linear loss function. It is well known (e.g., [1]) that, when X has a continuous distribution, that $\mathrm{CVaR}_\alpha(X) = \mathbb{E}[X \mid X \geqslant \mathrm{VaR}_\alpha(X)]$, where

$$\mathrm{VaR}_\alpha(X) \triangleq \sup_x \{x \mid \mathbb{P}(X \geqslant x) \geqslant \alpha\} \tag{3}$$

is the α-quantile (or "value-at-risk") of X. While this relationship does not necessarily hold if X has a discontinuous distribution, we can nonetheless roughly interpret $\mathrm{CVaR}_\alpha(X)$ as the expected loss over the α% worst cases.

As the thrust of this paper is the rate of convergence for estimating risk measures (specifically, CVaR), we now define the estimator for which we will prove convergence results. We will implicitly assume the random variable in question obeys the following assumption in everything that follows.

Assumption 2.1. The random variable $X \in \mathscr{X}$ satisfies $\mathrm{supp}(X) \subseteq [0, U]$ and the samples $X_1, \ldots, X_N$ used for estimation are independent.

A boundedness criterion is common in many concentration inequalities and is in fact needed to apply the results of McDiarmid [12] and Hoeffding [11]; we will, without loss of generality, use $\mathrm{supp}(X) \subseteq [0, U]$, though the results extend in straightforward fashion to more general intervals $[a, b] \subseteq \mathbb{R}$. The reason that this is so is that OCEs (and hence CVaR) satisfy translation invariance, i.e., for any $c \in \mathbb{R}$, we have $\mu_\phi(X + c) = \mu_\phi(X) + c$ (e.g., [6]). Therefore, we may always shift all samples by a constant amount so that they are in $[0, U]$, perform our estimation procedure, and then subtract the constant from the estimate. The corresponding convergence rates in such cases, then, apply directly with U simply replaced by $b - a$.

The estimator we will use is as follows.

Definition 2.4. Let μ_ϕ be an OCE with $\phi \in \Phi$. The *simple estimate* for $\mu_\phi(X)$ is

$$\hat{\mu}_\phi(X_1, \ldots, X_N) \triangleq \inf_v \left\{ v + \frac{1}{N}\sum_{i=1}^N \phi^*(X_i - v) \right\}. \tag{4}$$

For the case of $\mu_\phi = \mathrm{CVaR}_\alpha$, we denote the simple estimator by $\widehat{\mathrm{CVaR}}_\alpha$, i.e.,

$$\widehat{\mathrm{CVaR}}_\alpha(X_1, \ldots, X_n) \triangleq \inf_v \left\{ v + \frac{1}{N\alpha}\sum_{i=1}^N (X_i - v)^+ \right\}. \tag{5}$$

The estimator (4) is an intuitive one based on the method of moments; such estimators are efficiently computed for large N and most convex loss functions ϕ^*. The estimator (5) is the one typically used in most treatments discussing estimation of CVaR from samples (e.g., [14,15]). Our focus now will be quantifying the estimation error associated with (4) and, in particular, (5).

3. A bound on upper deviations via McDiarmid

The main machinery for proving a rate of convergence on upper deviations will be the following, powerful result from McDiarmid [12]. We will prove this convergence for general OCE risk measures, and CVaR follows as a special case.

Theorem 3.1 (*McDiarmid [12]*). *Consider a function $f : S^n \to \mathbb{R}$ which satisfies*

$$\sup_{x_1,\ldots,x_n,x_i' \in S} |f(x_1,\ldots,x_n) - f(x_1,\ldots,x_i',\ldots,x_n)| \leqslant c_i, \tag{6}$$

for all $i = 1, \ldots, n$. *Let* $X_1, \ldots, X_n$ *be independent random variables taking values in S. Then*

$$\mathbb{P}\left(|f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)]| \geqslant \varepsilon\right) \leqslant 2e^{-2\varepsilon^2/\sum_{i=1}^n c_i^2}. \tag{7}$$

Though not explicit in the statement of Theorem 3.1, McDiarmid's inequality is in fact symmetric; we apply the bound on upper deviations to derive a rate of convergence on pessimistic estimation errors for $\mu_\phi(X)$. We are able to do this because the estimator (4) lower bounds μ_ϕ in expectation, as we now show.

Proposition 3.1. *The estimator* (4) *satisfies the following*:

$$\mathbb{E}[\hat{\mu}_\phi(X_1, \ldots, X_N)] \leqslant \mu_\phi(X). \tag{8}$$

Proof. We prove the result for continuous probability spaces, but it easily extends to the more general case. We have

$$\mathbb{E}[\hat{\mu}_\phi(X_1, \ldots, X_N)] = \int_{\omega \in \Omega^N} \inf_v \left\{ v + \frac{1}{N} \sum_{i=1}^N \phi^*(X_i(\omega) - v) \right\} \mathrm{d}\mathbb{P}^N(\omega).$$

Since $\operatorname{supp}(X) \subseteq [0, U]$, there exists a $v^*(\omega) \in [0, U]$ attaining the infimum for every $\omega \in \Omega^N$ (see [6, Proposition 2.1]). By construction, $v^*(\omega)$ satisfies

$$v^*(\omega) + \frac{1}{N} \sum_{i=1}^N \phi^*(X_i(\omega) - v^*(\omega)) \leqslant v + \frac{1}{N} \sum_{i=1}^N \phi^*(X_i(\omega) - v) \quad \forall v \in \mathbb{R}. \tag{9}$$

In addition, let $\hat{v}^*$ be any v attaining the infimum for $\mu_\phi(X)$, i.e., $\hat{v}^* \in \arg\inf_v \{v + \mathbb{E}[\phi^*(X - v)]\}$. By similar reasoning, such a $\hat{v}^*$ is guaranteed to exist in $[0, U]$. We then have

$$\begin{aligned}
&\mathbb{E}[\hat{\mu}_\phi(X_1, \ldots, X_N)] \\
&= \int_{\omega \in \Omega^N} \inf_v \left\{ v + \frac{1}{N} \sum_{i=1}^N \phi^*(X_i(\omega) - v) \right\} \times \mathrm{d}\mathbb{P}^N(\omega) \\
&= \int_{\omega \in \Omega^N} \left(v^*(\omega) + \frac{1}{N} \sum_{i=1}^N \phi^*(X_i(\omega) - v^*(\omega)) \right) \times \mathrm{d}\mathbb{P}^N(\omega) \\
&\leqslant \int_{\omega \in \Omega^N} \left(\hat{v}^* + \frac{1}{N} \sum_{i=1}^N \phi^*(X_i(\omega) - \hat{v}^*) \right) \mathrm{d}\mathbb{P}^N(\omega) \\
&= \hat{v}^* + \frac{1}{N} \int_{\omega \in \Omega^N} \sum_{i=1}^N \phi^*(X_i(\omega) - \hat{v}^*)\, \mathrm{d}\mathbb{P}^N(\omega) \\
&= \hat{v}^* + \mathbb{E}[\phi^*(X - \hat{v}^*)] \\
&= \mu_\phi(X),
\end{aligned}$$

where the inequality follows by (9). □

The inequality in Proposition 3.1 will, in general, be strict. Indeed, consider the case of exponential utility, i.e., $\phi^*(u) = e^u - 1$. In this case, $\mu_\phi(X) = \log \mathbb{E}[e^X]$, but one also notes that $\inf_v \{v + e^{x-v} - 1\} = x$, which means, when $N = 1$, that $\mathbb{E}[\hat{\mu}_\phi(X_1)] = \mathbb{E}[X_1] = \mathbb{E}[X]$. From here, Jensen's inequality tells us that $\log \mathbb{E}[e^X] > \mathbb{E}[X]$, provided that X is not a constant random variable, and therefore the inequality is strict in this case.

In general, then, the estimator (4) is optimistically biased. As we will prove in the next section, however, for the case of CVaR, the probability that the simple estimate will deviate optimistically by an amount ε decays exponentially in $\varepsilon^2 N$.

For now, we are ready for the main result of this section, which is instead a large deviations bound on the estimator being overly *pessimistic*.

Theorem 3.2. *Let* μ_ϕ *be an OCE with* $\phi \in \Phi$. *Then*

$$\mathbb{P}(\hat{\mu}_\phi(X_1, \ldots, X_N) \geqslant \mu_\phi(X) + \varepsilon) \leqslant e^{-2(\varepsilon/\phi^*(U))^2 \cdot N}. \tag{10}$$

Proof. The main task is to find the bounded differences, then apply McDiarmid's inequality and Proposition 3.1. Without loss of generality, we may assume the term on the left below (the term with X_1 not X_1') is the larger of the two, and let v_R^* be a value of v which achieves the infimum for the term on the right (again, such a v_R^* is guaranteed to exist). We then have

$$\begin{aligned}
&\sup_{(X_1,\ldots,X_N,X_1')\in[0,U]^{N+1}} [\hat{\mu}_\phi(X_1,\ldots,X_N) - \hat{\mu}_\phi(X_1',\ldots,X_N)] \\
&= \sup_{(X_1,\ldots,X_N,X_1')\in[0,U]^{N+1}} \times \left[\inf_v \left\{ v + \frac{1}{N}\sum_{i=1}^N \phi^*(X_i - v) \right\} - \inf_v \left\{ v + \frac{1}{N}\left(\sum_{i=2}^N \phi^*(X_i - v)\right) + \frac{1}{N}\phi^*(X_1' - v) \right\} \right] \\
&\leqslant \sup_{(X_1,\ldots,X_N,X_1')\in[0,U]^{N+1}} \times \left[v_R^* + \frac{1}{N}\sum_{i=1}^N \phi^*(X_i - v_R^*) - \left(v_R^* + \frac{1}{N}\left(\sum_{i=2}^N \phi^*(X_i - v_R^*)\right) + \frac{1}{N}\phi^*(X_1' - v_R^*)\right) \right] \\
&= \frac{1}{N} \sup_{(X_1,X_1')\in[0,U]^2} [\phi^*(X_1 - v_R^*) - \phi^*(X_1' - v_R^*)] \\
&\leqslant \frac{\phi^*(U)}{N},
\end{aligned}$$

where the last inequality follows from the fact that ϕ^* is convex and nondecreasing. Indeed, it is well known that any conjugate function is convex (e.g., [13]); to see that $\phi \in \Phi$ implies that ϕ^* is also nondecreasing, consider $y_1 \geqslant y_2$, and let $x_2^* \in \arg\sup\{y_2 x - \phi(x)\}$. Then $\phi^*(y_1) = \sup_{x\in\operatorname{dom}\phi}\{y_1 x - \phi(x)\} \geqslant y_1 x_2^* - \phi(x_2^*) \geqslant y_2 x_2^* - \phi(x_2^*) = \phi^*(y_2)$, where the last inequality follows from the fact that $x_2^* \in \operatorname{dom}\phi \subseteq \mathbb{R}_+$.

The bound on the negative difference follows by an identical but reversed argument. We then have

$$\begin{aligned}
&\mathbb{P}(\hat{\mu}_\phi(X_1,\ldots,X_N) \geqslant \mu_\phi(X) + \varepsilon) \\
&\leqslant \mathbb{P}(\hat{\mu}_\phi(X_1,\ldots,X_N) \geqslant \mathbb{E}[\hat{\mu}_\phi(X_1,\ldots,X_N)] + \varepsilon) \\
&\leqslant e^{-2\varepsilon^2/(N\cdot(\phi^*(U)/N)^2)} \\
&= e^{-2(\varepsilon/\phi^*(U))^2\cdot N},
\end{aligned}$$

where the first inequality follows from Proposition 3.1 and the second follows from Theorem 3.1, which can be applied because of the bounded difference $\phi^*(U)/N$ computed above. □

Theorem 3.2 leads us immediately to the upper deviation bound on CVaR.

Corollary 3.1. *When $\mu_\phi(X) = \mathrm{CVaR}_\alpha(X)$ for some $\alpha \in (0, 1]$, we have*

$$\mathbb{P}(\widehat{\mathrm{CVaR}}_\alpha(X_1,\ldots,X_N) \geqslant \mathrm{CVaR}_\alpha(X) + \varepsilon) \leqslant e^{-2(\alpha\varepsilon/U)^2\cdot N}. \tag{11}$$

Proof. Immediate from the fact that $\phi^*(u) = \alpha^{-1}u^+$ for CVaR_α. □

4. A bound on lower deviations via Hoeffding

In this section, we are able to exploit the structure of the estimator for CVaR to bound lower deviation errors. The main tool we will use is a result from Hoeffding [11], which is a special case of Theorem 3.1.

Theorem 4.1 (*Hoeffding [11]*). *Let $X_1,\ldots,X_N$ be i.i.d. random variables with* $\operatorname{supp}(X) \subseteq [0, U]$, *where $U \geqslant 0$. Then, for any $\varepsilon \geqslant 0$, we have*

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^N X_i - \mathbb{E}[X]\right| \geqslant \varepsilon\right) \leqslant 2e^{-2(\varepsilon/U)^2 N}.$$

Hoeffding's inequality says we need on the order of $N_H \triangleq (U/\varepsilon)^2 \log(1/\delta)$ samples to estimate the sample mean within a precision of ε with probability at least $1-\delta$. We would expect for CVaR_α to need on the order of N_H/α samples, as CVaR is essentially a conditional expectation of the α-tail of the distribution, and roughly α% samples fall in the α-tail. In fact, for the lower deviation bound, we can use the structure of the estimator for CVaR in conjunction with Hoeffding's result to match this intuition.

To show this, we first need the following, straightforward fact.

Proposition 4.1. *The inequality*

$$\widehat{\mathrm{CVaR}}_\alpha(X_1,\ldots,X_N) \geqslant \frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha\rfloor} X_{(i)}, \tag{12}$$

holds, where $X_{(i)}$ are the decreasing order statistics of X_i, i.e., $X_{(1)} \geqslant X_{(2)} \geqslant \cdots \geqslant X_{(N)}$.

Proof. The proof follows by simply carrying out the minimization of the piecewise linear, convex function $v + (1/N\alpha)\sum_{i=1}^{N}(X_i - v)^+$, which has $N+1$ pieces. Quick inspection shows that the slope of this function changes sign from positive to negative at $v = X_{(\lceil N\alpha \rceil)}$, which means $v^* = X_{(\lceil N\alpha \rceil)}$. We then have

$$\begin{aligned}
&\widehat{\mathrm{CVaR}}_\alpha(X_1, \ldots, X_N) \\
&\quad = v^* + \frac{1}{N\alpha}\sum_{i=1}^{N}(X_i - v^*)^+ \\
&\quad = X_{(\lceil N\alpha \rceil)} + \frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha \rfloor}(X_{(i)} - X_{(\lceil N\alpha \rceil)}) \\
&\quad = \left(1 - \frac{\lfloor N\alpha \rfloor}{N\alpha}\right) X_{(\lceil N\alpha \rceil)} + \frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{(i)} \\
&\quad \geqslant \frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{(i)},
\end{aligned}$$

where the inequality follows from $X_i \geqslant 0$. □

We can now generalize the Hoeffding's inequality to apply to the simple estimates of $\mathrm{CVaR}_\alpha(X)$ for random variables with bounded support. Our result applies to underlying random variables with a continuous distribution function.

Theorem 4.2. *When X has a continuous distribution function, we have, for any $\varepsilon \geqslant 0$,*

$$\begin{aligned}
&\mathbb{P}(\widehat{\mathrm{CVaR}}_\alpha(X_1, \ldots, X_N) \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon) \\
&\quad \leqslant 3\mathrm{e}^{-(1/5)\alpha(\varepsilon/U)^2 \cdot N}.
\end{aligned} \tag{13}$$

Proof. We have, by Proposition 4.1,

$$\begin{aligned}
&\mathbb{P}(\widehat{\mathrm{CVaR}}_\alpha(X_1, \ldots, X_N) \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon) \\
&\quad \leqslant \mathbb{P}\left(\frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon\right).
\end{aligned}$$

Since X has a continuous distribution, we have that $\mathrm{CVaR}_\alpha(X) = \mathbb{E}[X | X \geqslant \mathrm{VaR}_\alpha(X)]$. Thus, from here, we are basically trying to bound the error in estimating conditional expectation. The key to doing this is to condition on the random variable $K_{N,\alpha}$, defined as $K_{N,\alpha} = \max\{i | X_{(i)} \in [\mathrm{VaR}_\alpha(X), U]\}$. Note that $K_{N,\alpha}$ is a function of $(X_1, \ldots, X_N)$, but we omit this dependence for brevity. Clearly $K_{N,\alpha}$ is binomially distributed with parameters N and α.

From here, if we condition on $K_{N,\alpha} = k$, one can see by symmetry that $1/k\sum_{i=1}^{k} X_{(i)}$ is equal in distribution to $1/k\sum_{i=1}^{k}\tilde{X}_i$, where $\tilde{X}_i$ are i.i.d. and equal in distribution to $\{X | X \in [\mathrm{VaR}_\alpha(X), U]\}$. We then have

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon\right) \\
&\quad = \sum_{k=0}^{N}\mathbb{P}(K_{N,\alpha} = k)\mathbb{P}\left(\frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon \,\middle|\, K_{N,\alpha} = k\right) \\
&\quad = \underbrace{\sum_{k=0}^{\lfloor N\alpha \rfloor}\mathbb{P}(K_{N,\alpha} = k)\mathbb{P}\left(\frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon \,\middle|\, K_{N,\alpha} = k\right)}_{I_2} \\
&\qquad + \underbrace{\sum_{k=\lfloor N\alpha \rfloor + 1}^{N}\mathbb{P}(K_{N,\alpha} = k)\mathbb{P}\left(\frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon \,\middle|\, K_{N,\alpha} = k\right)}_{I_1}.
\end{aligned}$$

We now distinguish two cases.

Case 1: $k \geqslant \lfloor N\alpha \rfloor + 1$. Then

$$\mathbb{P}\left(\frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon \,\middle|\, K_{N,\alpha} = k\right)$$

$$\leqslant \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon \,\middle|\, K_{N,\alpha} = k\right)$$

$$(k \geqslant \lfloor N\alpha \rfloor + 1 \geqslant N\alpha)$$

$$= \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} \tilde{X}_i \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon \,\middle|\, K_{N,\alpha} = k\right)$$

$$\leqslant \mathrm{e}^{-2(\varepsilon/U)^2 k} \quad \text{(Theorem. 4.1)}.$$

Therefore, we have

$$I_1 \leqslant \sum_{k=\lfloor N\alpha \rfloor+1}^{N} \binom{N}{k} \alpha^k (1-\alpha)^{N-k} \mathrm{e}^{-2(\varepsilon/U)^2 k}$$

$$\leqslant \mathrm{e}^{-2(\varepsilon/U)^2(\lfloor N\alpha \rfloor+1)} \sum_{k=\lfloor N\alpha \rfloor+1}^{N} \binom{N}{k} \alpha^k (1-\alpha)^{N-k}$$

$$\leqslant \mathrm{e}^{-2(\varepsilon/U)^2(\lfloor N\alpha \rfloor+1)}$$

$$\leqslant \mathrm{e}^{-2\alpha(\varepsilon/U)^2 N}.$$

Case 2: $k \leqslant \lfloor N\alpha \rfloor$. Then

$$\mathbb{P}\left(\frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon \,\middle|\, K_{N,\alpha} = k\right)$$

$$\leqslant \mathbb{P}\left(\frac{1}{N\alpha}\sum_{i=1}^{k} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon \,\middle|\, K_{N,\alpha} = k\right)$$

$$(X_i \geqslant 0)$$

$$= \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} X_{(i)} \leqslant \frac{N\alpha}{k}(\mathrm{CVaR}_\alpha(X) - \varepsilon) \,\middle|\, K_{N,\alpha} = k\right)$$

$$\leqslant \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} X_{(i)} \leqslant \frac{\lceil N\alpha \rceil}{k}(\mathrm{CVaR}_\alpha(X) - \varepsilon) \,\middle|\, K_{N,\alpha} = k\right)$$

$$\leqslant \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \frac{\lceil N\alpha \rceil}{k}\varepsilon + \left(\frac{\lceil N\alpha \rceil - k}{k}\right) U \,\middle|\, K_{N,\alpha} = k\right)$$

$$(\mathrm{CVaR}_\alpha(X_i) \leqslant U)$$

$$\leqslant \mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon'(k) \,\middle|\, K_{N,\alpha} = k\right),$$

where $\varepsilon'(k) = (\lceil N\alpha \rceil/k)(\varepsilon - U) + U$. Note that $\varepsilon'(k) \geqslant 0$ if and only if $k \geqslant (1 - \varepsilon/U)\lceil N\alpha \rceil$. Let $\gamma = 1 - \varepsilon/U$ and let $k^* = \lceil \gamma \lceil N\alpha \rceil \rceil$. Finally, for some $\beta \in [0, 1]$, let $k^*_\beta = \lceil (\beta\gamma + (1-\beta))\lceil N\alpha \rceil \rceil = \lceil (1 - \beta(\varepsilon/U))\lceil N\alpha \rceil \rceil \in [k^*, \lceil N\alpha \rceil]$.We furthermore note that, for any $k \in [k^*, \lceil N\alpha \rceil]$, we have

$$(\varepsilon'(k)/U)^2 = \left(1 - \frac{\gamma\lceil N\alpha \rceil}{k}\right)^2$$

$$= \left(\frac{\lceil N\alpha \rceil}{k}\right)^2 \left(\frac{k}{\lceil N\alpha \rceil} - \gamma\right)^2$$

$$\geqslant \left(\frac{k}{\lceil N\alpha \rceil} - \gamma\right)^2. \tag{14}$$

Therefore, we have

$$\mathbb{P}\left(\frac{1}{k}\sum_{i=1}^{k} X_{(i)} \leqslant \mathrm{CVaR}_\alpha(X) - \varepsilon'(k) \,\middle|\, K_{N,\alpha} = k\right)$$

$$\leqslant \begin{cases} \mathrm{e}^{-2(\varepsilon'(k)/U)^2 k}, & k^* \leqslant k \leqslant \lfloor N\alpha \rfloor \\ 1, & k < k^* \end{cases} \quad \text{(Theorem. 4.1)},$$

$$\leqslant \begin{cases} \mathrm{e}^{-2k(k/\lceil N\alpha \rceil - \gamma)^2}, & k^* \leqslant k \leqslant \lfloor N\alpha \rfloor \\ 1, & k < k^* \end{cases} \quad \text{(Eq. (14))}.$$

Continuing further, we have

$$\begin{cases} e^{-2k}(k/\lceil N\alpha\rceil-\gamma)^2, & k^*\leqslant k\leqslant \lfloor N\alpha\rfloor,\\ 1, & k<k^*,\end{cases}$$

$$\leqslant \begin{cases} e^{-2k}(k/\lceil N\alpha\rceil-\gamma)^2, & k^*_\beta\leqslant k\leqslant \lfloor N\alpha\rfloor,\\ 1, & k<k^*_\beta,\end{cases}$$

$$\leqslant \begin{cases} e^{-2k(k^*_\beta/\lceil N\alpha\rceil-\gamma)^2}, & k^*_\beta\leqslant k\leqslant \lfloor N\alpha\rfloor,\\ 1, & k<k^*_\beta,\end{cases}$$

$$\leqslant \begin{cases} e^{-2k(\lceil(1-\beta(\varepsilon/U))\lceil N\alpha\rceil\rceil/\lceil N\alpha\rceil-\gamma)^2}, & k^*_\beta\leqslant k\leqslant \lfloor N\alpha\rfloor,\\ 1, & k<k^*_\beta,\end{cases}$$

$$\leqslant \begin{cases} e^{-2k(1-\beta(\varepsilon/U)-\gamma)^2}, & k^*_\beta\leqslant k\leqslant \lfloor N\alpha\rfloor,\\ 1, & k<k^*_\beta,\end{cases}$$

$$= \begin{cases} e^{-2k((1-\beta)\varepsilon/U)^2}, & k^*_\beta\leqslant k\leqslant \lfloor N\alpha\rfloor,\\ 1, & k<k^*_\beta.\end{cases}$$

Going back to our original expansion, we have

$$I_2=\sum_{k=0}^{\lfloor N\alpha\rfloor}\mathbb{P}\left(K_{N,\alpha}=k\right)\times\mathbb{P}\left(\frac{1}{N\alpha}\sum_{i=1}^{\lfloor N\alpha\rfloor}X_{(i)}\leqslant \mathrm{CVaR}_\alpha(X)-\varepsilon \,\middle|\, K_{N,\alpha}=k\right)$$

$$\leqslant \underbrace{\left[\sum_{k=0}^{k^*_\beta-1}\binom{N}{k}\alpha^k(1-\alpha)^{N-k}\right]}_{I_{2a}} + \underbrace{\left[\sum_{k=k^*_\beta}^{\lfloor N\alpha\rfloor}\binom{N}{k}\alpha^k(1-\alpha)^{N-k}e^{-2k((1-\beta)\varepsilon/U)^2}\right]}_{I_{2b}}.$$

For the first term, we have $I_{2a}\leqslant\mathbb{P}(K_{N,\alpha}\leqslant(1-\beta(\varepsilon/U))N\alpha)\leqslant e^{-(\alpha/2)(\beta(\varepsilon/U))^2N}$, where the second inequality follows by the Chernoff bound. For the second term, we have

$$\begin{aligned}I_{2b}&=\sum_{k=k^*_\beta}^{\lfloor N\alpha\rfloor}\binom{N}{k}(\alpha\cdot e^{-2((1-\beta)\varepsilon/U)^2})^k(1-\alpha)^{N-k}\\ &\leqslant(1-(1-e^{-2((1-\beta)\varepsilon/U)^2})\alpha)^N\\ &\quad\text{(Binomial expansion)}\\ &\leqslant e^{-(1-\exp(-2((1-\beta)\varepsilon/U)^2)\alpha N)}\\ &\quad(1-\Delta\leqslant e^{-\Delta}\text{ for }\Delta\in[0,1])\\ &\leqslant e^{-\alpha(2\beta(2-\beta)(1-\beta)^2)(\varepsilon/U)^2N},\end{aligned}$$

where the last line follows from the fact that, for any $\rho\in[0,1]$ and $x\in[0,1]$, we have

$$\begin{aligned}1-e^{-2\rho^2x^2}&=1-\sum_{k=0}^{\infty}(-1)^k\frac{(2\rho^2x^2)^k}{k!}\\ &=2\rho^2x^2-2\rho^4x^4\\ &\quad+\sum_{k=1}^{\infty}\left[1-\left(\frac{2\rho^2x^2}{2k+2}\right)\right]\frac{(2\rho^2x^2)^{2k+1}}{(2k+1)!}\\ &\geqslant 2\rho^2x^2(1-\rho^2x^2)\\ &\geqslant 2\rho^2(1-\rho^2)x^2,\end{aligned}$$

where the first inequality is a consequence of the fact that $2\rho^2x^2/(2k+2)\leqslant 1$ on $k\geqslant 1$ (since $\rho x\leqslant 1$). Using $\rho=1-\beta$ and $x=\varepsilon/U$ results in the above bound. To combine the three terms into a single, exponential bound with the highest decay coefficient, we want to choose $\beta\in[0,1]$ such that the $f(\beta)=\min(\beta^2/2, 2\beta(2-\beta)(1-\beta)^2)$ is maximum, which occurs when these two terms in the minimization are equal, or, equivalently, when $\beta/2=2(2-\beta)(1-\beta)^2$, which has a unique root $\hat\beta$ on $[0,1]$ satisfying $\hat\beta^2/2\geqslant .2172\geqslant\frac{1}{5}$. Putting everything together, we have

$$\begin{aligned}&\mathbb{P}(\widehat{\mathrm{CVaR}}_\alpha(X_1,\ldots,X_N)\leqslant \mathrm{CVaR}_\alpha(X)-\varepsilon)\\ &\quad\leqslant I_1+I_{2a}+I_{2b}\\ &\quad\leqslant e^{-2\alpha(\varepsilon/U)^2N}+e^{-\alpha(\beta^2/2)(\varepsilon/U)^2N}\\ &\qquad+e^{-\alpha(2\beta(2-\beta)(1-\beta)^2)(\varepsilon/U)^2N}\\ &\quad\leqslant 3e^{-\alpha\min(2,\beta^2/2,2\beta(2-\beta)(1-\beta)^2)(\varepsilon/U)^2N}\\ &\quad\leqslant 3e^{-\hat\beta^2/2\cdot\alpha(\varepsilon/U)^2\cdot N}\\ &\quad\leqslant 3e^{-(1/5)\alpha(\varepsilon/U)^2\cdot N}.\qquad\square\end{aligned}$$

5. Conclusions

In this paper, we have proven an exponential convergence result for an estimator of CVaR for bounded random variables representing an uncertain loss. Open directions include the following:

1. Improving the bound on pessimistic errors. We believe the α^2 in this bound in Theorem 3.2 can be improved to α as in the optimistic case.
2. Establishing tightness of the bounds. Intuition suggests that the bound on optimistic errors is tight up to constants. It is also interesting to see how these bounds perform over more restricted classes of distributions (e.g., truncated normal, uniform, distributions with moment constraints, etc.).
3. Extending these bounds to optimization problems. CVaR and, more generally, convex risk measures, have gained popularity in optimization problems with uncertainty, particularly portfolio optimization. A challenging, open question is whether similar bounds can be derived for optimal solutions to finite-sample optimization problems using CVaR and other risk measures.

Acknowledgement

We would like to thank an anonymous referee for a number of helpful comments which improved this manuscript.

References

[1] C. Acerbi, D. Tasche, On the coherence of expected shortfall, J. Banking Finance 26 (7) (2002).
[2] P. Artzner, F. Delbaen, J. Eber, D. Heath, Coherent measures of risk, Math. Finance 9 (1999) 203–228.
[3] A. Ben-Tal, M. Teboulle, Expected utility, penalty functions and duality in stochastic nonlinear programming, Manage. Sci. 32 (1986).
[4] A. Ben-Tal, M. Teboulle, Penalty functions and duality in stochastic programming via ϕ-divergence functionals, Math. Oper. Res. 12 (1987).
[5] A. Ben-Tal, M. Teboulle, Portfolio theory for the recourse certainty equivalent maximizing investor, Ann. Oper. Res. 31 (1991).
[6] A. Ben-Tal, M. Teboulle, An old-new concept of convex risk measures: the optimized certainty equivalent, J. Math. Finance (2006), to appear.
[7] F. Delbaen, Pisa lecture notes: coherent risk measures, Scuola Normale Superiore di Pisa, ⟨http://www.math.ethz.ch/~delbaen⟩, 2000.
[8] F. Delbaen, Coherent measures of risk on general probability spaces, in: K. Sandmann, P.J. Schonbucher (Eds.), Advances in Finance and Stochastics, Essays in Honor of Dieter Sondermann, Springer, Berlin, 2002, pp. 1–37.
[9] P. Embrechts, Extreme value theory: potential and limitations as an integrated risk management tool, Derivatives Use, Trading & Regulation 6 (2000).
[10] H. Föllmer, A. Schied, Convex measures of risk and trading constraints, Finance Stochastics 6 (2002).
[11] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Am. Stat. Assoc. 58 (1963).
[12] C. McDiarmid, On the method of bounded differences, in: Surveys in Combinatorics 1989, Cambridge University Press, Cambridge, 1989.
[13] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[14] R.T. Rockafellar, S.P. Uryasev, Optimization of conditional value-at-risk, J. Risk 2 (2000) 21–41.
[15] A. Takeda, T. Kanamori, A robust optimization approach based on conditional value-at-risk measure and its applications to statistical learning problems, Working paper, 2005.
[16] J. von Neumann, O. Morgenstern, Theory of Games and Economic Behavior, Princeton University Press, Princeton, 1947.