European Journal of Operational Research 000 (2017) 1–11

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

Expected shortfall: Heuristics and certificates

Federico Alessandro Ramponi*, Marco C. Campi

Università degli Studi di Brescia, Dipartimento di ingegneria dell'informazione, Via Branze 38, Brescia 25123, Italy

A R T I C L E I N F O

Article history:
Received 17 May 2016
Accepted 12 November 2017
Available online xxx

Keywords:
Stochastic programming
Convex programming
Scenario approach
Expected shortfall
CVaR

A B S T R A C T

We consider the expected shortfall, a coherent risk measure that is gaining popularity outside mathematical finance and that is being applied to an increasing number of optimization problems due to its versatility and pleasant properties. A commonly used heuristic to optimize the expected shortfall consists in replacing the unknown distribution of the loss function with its empirical discrete counterpart constructed from observations. The boundary of the empirical shortfall tail is called the shortfall threshold, and, in this paper, we study the probability of incurring losses larger than the shortfall threshold. In a stationary set-up, we show that under mild conditions a striking universal result holds which says that the probability of losses exceeding the shortfall threshold is a random variable whose distribution is independent of the distribution of the loss function. This result complements previous findings on the expected shortfall and bears important practical consequences on the applications of this risk measure to stochastic optimization. The theory this result rests on is fully developed in this paper and its use is illustrated by examples.

1. Introduction

In this paper, we consider the following sample-based optimization problem:

$$x_N^* = \arg\min_{x\in\mathcal{X}} \{\text{average of the } k \text{ largest values among } L(x,\delta_1),\dots,L(x,\delta_N)\}, \tag{1}$$

where k is an integer in the range $1 \le k \le N$, $\mathcal{X}$ is a convex subset of $\mathbb{R}^d$, $L(\cdot,\delta_i)$ are convex cost functions, each one depending on a value δ_i of a random variable δ, and where the random sample $(\delta_1,\dots,\delta_N)$ is supposed to be independent and identically distributed. In a real application, the variable δ describes uncertainty, and δ_i are observations of the variable δ that come from previous experience. The quantity being minimized in (1) is the empirical estimate of a measure of risk, known in financial risk management as *Conditional Value-at-Risk* (CVaR) or *Expected Shortfall* (ES). To make the meaning of (1) concrete, we introduce at this early stage an example that will be resumed later with more explanation and numerical results.

Example 1.1 (Portfolio optimization)**.** Suppose that n_a assets $A^{[1]},\dots,A^{[n_a]}$ are available for trading. On period i, the asset $A^{[j]}$ may gain or lose value in the market, and the ratio $\delta_i^{[j]} = (P_i^{[j]} - P_{i-1}^{[j]})/P_{i-1}^{[j]}$, where $P_i^{[j]}$ is the close price of asset $A^{[j]}$ on period i, is called the *rate of return* of asset $A^{[j]}$ on period i. To cope with uncertainty, investors *diversify* among assets; thus, if an investor has 1$ to invest, s/he will invest fractions $x^{[1]},\dots,x^{[n_a]}$ of her/his dollar on $A^{[1]},\dots,A^{[n_a]}$ (we assume that $x^{[j]} \ge 0$ for all j, and $\sum_{j=1}^{n_a} x^{[j]} = 1$). The vector $x := (x^{[1]},\dots,x^{[n_a]})$ is called a *portfolio*. Letting $\delta_i := (\delta_i^{[1]},\dots,\delta_i^{[n_a]})$ be the vector of the rates of return, the scalar product $\delta_i \cdot x = \sum_{j=1}^{n_a} \delta_i^{[j]} x^{[j]}$ is the rate of return of the portfolio on period i. If $\delta_i{\cdot}x$ is positive, the investor's capital increases on period i of $\delta_i{\cdot}x$ $ for each dollar invested. Hence,

$$L_i(x) := -\delta_i \cdot x$$

quantifies the *portfolio loss* on period i.

Suppose now that the investor has observed a record of N vectors $(\delta_1,\dots,\delta_N)$ on various periods. Then s/he can choose a portfolio x_N^* by minimizing cost (1) where $\mathcal{X} = \{x \in \mathbb{R}^{n_a} : x^{[j]} \ge 0 \text{ for all } j,\ \sum_{j=1}^{n_a} x^{[j]} = 1\}$ is the simplex in $\mathbb{R}^{n_a}$. The interpretation is that the investor chooses the portfolio that incurs the lowest average loss over the empirical shortfall cases.

CVaR is a coherent risk measure in the sense of Artzner, Delbaen, Eber, and Heath (1999), which has been introduced and pop-

* Corresponding author.
E-mail addresses: federico.ramponi@unibs.it (F.A. Ramponi), marco.campi@unibs.it (M.C. Campi).

https://doi.org/10.1016/j.ejor.2017.11.022
0377-2217/© 2017 Elsevier B.V. All rights reserved.

ularized by Rockafellar and Uryasev (2000) and Rockafellar and Uryasev (2002) in their papers.[1] ES (see e.g. Christoffersen, 2012 or Fabozzi, Kolm, Pachamanova, & Focardi, 2007) is defined similarly to CVaR, and the difference between CVaR and ES arises only when the distribution of $L(x,\cdot)$ has point masses (that is, there are single values that have non-zero probability to occur). Moreover, in finance, often such a difference is not even considered and a definition of ES completely equivalent to CVaR is used, see e.g. Acerbi and Tasche (2002) and McNeil, Frey, and Embrechts (2015, Chapter 2). In this paper we deal with distributions *without* point masses, and use the definition of ES (or CVaR) that is given in formula (4) below. Terminology and technicalities aside, the concept of expected shortfall is gaining popularity in fields well outside the realm of financial analysis. For example, ES as a measure of risk has recently seen applications to breast cancer therapy (Chan, Mahmoudzadeh, & Purdie, 2014), scheduling (Quan, He, & He, 2014; Sarin, Sherali, & Liao, 2014), and machine learning (Takeda, 2009; Takeda & Kanamori, 2009, 2014; Wang, Dang, & Wang, 2015).[2]

As said before, Problem (1) is a heuristic towards the minimization of an ES risk measure and in this paper we provide rigorous results that certify the properties of this heuristic. More precisely, we introduce a notion of *shortfall threshold* $\bar{L}_N$ (see Eq. (6)) which is interpreted as the empirical boundary of shortfall cases and consider the event where "a further function $L(\cdot,\delta)$, with δ sampled from P independently of the already seen values $(\delta_1,\dots,\delta_N)$, incurs a cost $L(x^*_N,\delta)$ bigger than $\bar{L}_N$". Such a probability is written as $\mathsf{P}\left\{\delta : L(x^*_N,\delta) > \bar{L}_N\right\}$, and is a random variable because it depends on x^*_N and $\bar{L}_N$, which in turn depend on the random sample $\delta_1,\dots,\delta_N$.

We show that a probabilistic certificate of the form

$$\mathsf{P}\left\{\delta : L(x^*_N,\delta) > \bar{L}_N\right\} \le \varepsilon \quad \text{with confidence } 1-\beta \tag{2}$$

can be attached to the solution of (1). This result has a *universal* validity, that is, it holds true regardless of the distribution P by which the δ_i's are sampled. Hence, an experimenter unaware of P can still append to the solution of Problem (1) a probabilistic certificate in the form of (2). This paper also shows the usefulness of this result by providing a set of corollaries that have a practical use, as well as application examples with real data.

1.1. Structure of the paper

Relevant definitions are given in Section 2. In Section 3 the main result that the random variable $\mathsf{P}\left\{\delta : L(x^*_N,\delta) > \bar{L}_N\right\}$ has a universal distribution is stated and proven, followed by two corollaries regarding the statistics and the long-run behavior of such random variable. Section 4 presents two applications exploring, respectively, the choice of k, and the long-run behavior of a sequence of optimization problems solved in a "sliding window" fashion. In Section 5, the results from Sections 3 and 4 are applied to the optimization of a portfolio that includes shares of 10 companies with high market capitalization traded on the New York Stock Exchange and the NASDAQ. The paper ends with some conclusions and acknowledgments.

2. Formal definitions and problem position

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a convex set, $(\Delta,\mathcal{F},\mathsf{P})$ be a probability space, and $L : \mathcal{X}\times\Delta \to \mathbb{R}$ be a function such that

1. For any $x\in\mathcal{X}$, $L(x,\cdot)$ is a random variable on $(\Delta,\mathcal{F},\mathsf{P})$;
2. For any $\delta\in\Delta$, $L(\cdot,\delta)$ is a convex function on $\mathcal{X}$.

L is interpreted as a cost function whose value depends on an optimization variable x and a variable δ (uncertainty variable) that accounts for all other sources of variation of L besides x. If $(\delta_1,\dots,\delta_N)$ is a sample of independent realizations from $(\Delta,\mathcal{F},\mathsf{P})$, we shall often use the shorthand notation $L_i := L(\cdot,\delta_i)$, and $L_i(x) := L(x,\delta_i)$, $i=1,\dots,N$.

For any $x\in\mathcal{X}$, denote by $L_{(i)}(x)$, $i=1,\dots,N$, the values attained by $L_1(x),\dots,L_N(x)$ taken in descending order:

$$L_{(1)}(x) \ge L_{(2)}(x) \ge \cdots \ge L_{(N)}(x).$$

In statistical terminology, David and Nagaraja (2003), $L_{(N-i+1)}(x)$ is called the ith order statistic of the random sample $L_1(x),\dots,L_N(x)$. Problem (1) can now be restated as follows:

$$\min_{x\in\mathcal{X}} \frac{1}{k}\sum_{i=1}^{k} L_{(i)}(x), \tag{3}$$

where $1\le k\le N$.

We next introduce a definition of expected shortfall. If L is a random variable modeling a loss, $\alpha\in[0,1]$, and F_L is the cumulative distribution function of L, the *Value at Risk* (VaR) and *Expected Shortfall* (ES) of L are given by:

$$\begin{aligned} \mathrm{VaR}_\alpha(L) &:= \min\{l\in\mathbb{R} : F_L(l)\ge\alpha\}, \\ \mathrm{ES}_\alpha(L) &:= \mathsf{E}[L \mid L > \mathrm{VaR}_\alpha(L)]. \end{aligned} \tag{4}$$

$\mathrm{VaR}_\alpha(L)$ is the threshold value at the boundary of the fraction α of highest losses. VaR is currently the most widely adopted risk measure in banking and finance despite some of its shortcomings seem to suggest that it would be better replaced by other measures like ES (refer e.g. to Christoffersen (2012) and Fabozzi et al. (2007) for examples and practical uses, and to Rockafellar and Uryasev (2002) and Hong, Hu, and Liu (2014) for a comparison of the properties of VaR and ES). $\mathrm{ES}_\alpha(L)$ is instead the expected loss suffered when the threshold $\mathrm{VaR}_\alpha(L)$ is exceeded. When the loss L depends on a choice $x\in\mathcal{X}$, i.e., $L=L(x)$, it makes sense to minimize the expected shortfall for a selected value of α:

$$\min_{x\in\mathcal{X}} \mathrm{ES}_\alpha(L(x)). \tag{5}$$

Problem (3) is indeed an empirical version of Problem (5) for $\alpha = 1-\frac{k}{N}$, based on the N observations $\delta_1,\dots,\delta_N$. Hence, we call Problem (3) the *empirical expected shortfall* problem.

Let x^*_N be the minimizer of (3), assume that it exists and is unique and, assuming also that $N\ge k+d$, define

$$\bar{L}_N := L_{(k+d)}(x^*_N). \tag{6}$$

We call $\bar{L}_N$ the *shortfall threshold*. In typical cases the interpretation of $\bar{L}_N$ is that it separates shortfall empirical functions from functions attaining a lower value at the minimizer. This is easily understood by making reference to a simple case where $d=1$ and $k=2$, as shown in Fig. 1(a). The dashed function is $\frac{1}{2}(L_{(1)}(x)+L_{(2)}(x))$. x^*_N minimizes this dashed function, which happens at the intersection of two functions L_i. $\bar{L}_N = L_{(2)}(x^*_N) = L_{(3)}(x^*_N)$ is at the boundary of the values attained by the functions L_i that are averaged to determine the solution. Notice, however, that there are cases where $\bar{L}_N$ takes a value lower than the boundary value. For example, in Fig. 1(b) the solution is determined by two functions only, and $\bar{L}_N$ is obtained by "digging" at x^*_N until the third value $L_{(3)}(x^*_N)$ is reached. This situation may occur when the cost functions are not linear, as in Fig. 1(b), or even when they are linear and the solution x^*_N is obtained at a boundary point of the optimization domain $\mathcal{X}$. The reason why $\bar{L}_N$ is defined to always be the $(k+d)$th largest cost is that the theoretical certificate introduced in this paper holds true rigorously for this choice only.

[1] A recent work of Mafusalov and Uryasev (2016) has generalized the concept of CVaR from that of a risk measure to that of a norm over a space of random variables. In fact the average of the k greatest values among $|l_1|,\dots,|l_N|$ of a vector $(l_1,\dots,l_N)\in\mathbb{R}^N$ is a norm on $\mathbb{R}^N$; for $k=1$ it reduces to the Chebycheff norm $\|\cdot\|_\infty$.

[2] The minimization problem with empirical distribution in Takeda and Kanamori (2014, Section 3.2) is essentially Problem (1).