European Journal of Operational Research 000 (2018) 1–11

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Estimation and asymptotics for buffered probability of exceedance

Alexander Mafusalov[a,1,*], Alexander Shapiro[b,2], Stan Uryasev[a,1]

[a] Department of Industrial and Systems Engineering, University of Florida, Gainesville FL 32611, USA
[b] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta GA 30332-0205, USA

ARTICLE INFO

Article history:
Received 14 December 2016
Accepted 8 January 2018
Available online xxx

Keywords:
Stochastic programming
Buffered probability of exceedance
Sample average approximation
Rare events
Minimum volume ellipsoid

ABSTRACT

This paper studies statistical properties of empirical (sample) estimates of the buffered probability of exceedance (bPOE). The estimation procedure is based on one dimensional minimization representation of the bPOE. Convergence rates and asymptotic properties of the suggested estimation procedures are investigated. Theoretical predictions are validated with numerical experiments, including a special case of exponential distribution, and a study proposing bPOE modification of minimum volume ellipsoid problem.

1. Introduction

For a given probability distribution of losses and a threshold value, *Probability of Exceedance* (POE), is defined as the probability that the random variable of loss exceeds the threshold. This is a natural measure of uncertainty in losses. The POE is very popular in various engineering applications. For instance, nuclear engineering considers probability that radiation release will exceed specified level, while structural reliability analysis considers probability that load exceeds some threshold. Although POE is included in government regulations, it has some major shortcomings. From a conceptual point of view, the threshold in POE provides a low bound on tail outcomes exceeding this threshold. However, POE does not provide information about the magnitude of these outcomes. In other words, POE is capable of registering an exceeding outcome, but incapable of measuring its impact on the system. Also POE has troublesome mathematical properties for discretely distributed random variables, which are typically obtained from sample data. For these variables, POE is discontinuous with respect to the threshold, which prevents using standard sensitivity analysis based on derivatives. In addition, POE is difficult to optimize because optimization problems for POE are usually reduced to large-scale Mixed-Integer programming involving binary variables, a problem that may be hard to solve.

Buffered Probability of Exceedance (bPOE) for a random variable is a counterpart of the POE. The notion of bPOE was introduced and studied in Mafusalov and Uryasev (2014) and in Norton and Uryasev (2014). For a specified threshold, bPOE equals the probability of an upper tail of the distribution, such that the average of this tail coincides with the threshold. There is a similarity between POE and bPOE: the values of bPOE and POE are bounded between zero and one, and, for a given random variable and varying threshold, decrease with the threshold increase. However, bPOE is an upper bound for POE because it includes all outcomes exceeding the threshold, as well as some outcomes below the threshold. The outcomes below the threshold form the so called buffer, therefore, bPOE is a buffered POE. In that sense, the estimate of loss uncertainty given by bPOE is more conservative then the one given by POE.

The tail averages for probability distributions were introduced by Rockafellar and Uryasev (2000) by employing the notion of Conditional Value-at-Risk (CVaR). Specifically, $\mathrm{CVaR}_{1-\alpha}$ defines average in the upper α-tail of a probability distribution. Therefore, bPOE for a random variable X at a threshold $x \in \mathbb{R}$ equals α such that

$$\mathrm{CVaR}_{1-\alpha}(X) = x.$$

In this sense it is said that bPOE is an inverse function of CVaR. Some attractive mathematical properties hold for bPOE. It is continuous in threshold x (maybe except at one point) and quasi-convex in X. Furthermore, it was proved that bPOE is the tightest upper bound for POE among functions consistent with convex

* Corresponding author.
E-mail addresses: mafusalov@ufl.edu (A. Mafusalov), ashapiro@isye.gatech.edu (A. Shapiro), uryasev@ufl.edu (S. Uryasev).

[1] Research of these authors was partially supported by the USA AFOSR grants: "Design and Redesign of Engineering Systems", FA9550-12-1-0427, and "New Developments in Uncertainty: Linking Risk Management, Reliability, Statistics and Stochastic Optimization", FA9550-11-1-0258.
[2] Research of this author was partly supported by NSF grant 1633196 and DARPA EQUiPS program, grant SNL 014150709.

https://doi.org/10.1016/j.ejor.2018.01.021
0377-2217/

stochastic dominance, see Section 3.4 in Mafusalov and Uryasev (2014). One way to interpret this result is that when decision making with POE criterion is preferred, but also the decision maker is risk averse, then bPOE provides the closest suitable criterion. Moreover, a problem of bPOE minimization can be reduced to convex and even linear programming.

Mafusalov and Uryasev (2014) provide a detailed description of mathematical properties of bPOE and various optimization problem statements. When it comes to formulating optimization problems, the connection of bPOE and CVaR provides additional insights. In particular, CVaR and bPOE level constraints are equivalent. Furthermore, using CVaR or bPOE as objective leads to two parametric optimization problem families, and these families, with minor exceptions, share frontiers of optimal solutions. That is, CVaR minimization solution is not found directly from a single bPOE minimization solution, but rather it is found from a collection of solutions, and vice versa.

Given the above connections, it might seem that introduction of bPOE in optimization is redundant. However, this is only true when needs of decision maker require establishing the whole frontier of CVaR/bPOE optimal solutions by solving multiple (and, in general, infinitely many) optimization problems. Not all practical problems possess the luxury of investigating entire solution frontiers prior to making a decision. On the contrary, oftentimes computational resources are sufficient for solving a single optimization problem. Then, a choice between bPOE and CVaR comes from a choice of parameter type, which should be dictated by the nature of the problem or the motivation of decision maker. That is, some applications relate best to a specified fraction of worst cases, then CVaR objective is used; other applications relate best to a specific loss value threshold, then bPOE is used.

The bPOE concept is an extension of the *Buffered Failure Probability* suggested by Rockafellar (2009) and explored by Rockafellar and Royset (2010). The connection is such that buffered failure probability equals bPOE at zero threshold value. Buffered failure probability is built to be aligned with *failure probability*, defined as the probability that system fails, which happens when a corresponding random variable takes a positive value.

This paper studies statistical properties of empirical (sample) estimates of bPOE. The estimators are based on one-dimension minimization representation of bPOE suggested in Mafusalov and Uryasev (2014) and in Norton and Uryasev (2014). In particular, the asymptotic convergence of the suggested estimators is studied.

This paper is organised as follows. Section 2 formally discusses bPOE and some of its properties, introduces necessary notations, and proves results on bias, asymptotical variance, and convergence for a sample estimate of bPOE. Section 3 discusses approaches, including the importance sampling method, for estimating bPOE in case of rare events. The theoretical results of Sections 2, 3 are validated with numerical experiments in Section 4, where a special case of exponential distribution is considered. Convergence properties for optimal solutions and optimal values for bPOE minimization problem are derived in Section 5. A modification of minimum volume ellipsoid (MVE) problem is considered in Section 6. There, instead of minimizing a fraction of non-covered samples under a covering ellipsoid volume constraint, it is proposed to minimize bPOE with the same constraint. The resulting problem is convex, and hence can be efficiently solved. In a case of true sample generating distribution being elliptical, solutions for POE and bPOE minimization coincide. Theoretical results on optimal solution and value convergence are validated for the considered bPOE-modification of the MVE problem. This approach is closely related and is alternative to the conditional MVE by Gotoh and Takeda (2006, 2008).

2. Statistical properties of buffered probability estimates

For $\alpha \in [0,\ 1)$ Conditional Value-at-Risk (also called Average Value-at-Risk, Expected Shortfall and Expected Tail Loss) of a random variable X is defined as[3]

$$\mathsf{CVaR}_\alpha(X) := \inf_{t\in\mathbb{R}} \left\{ t + (1-\alpha)^{-1}\mathbb{E}[X-t]_+ \right\}. \tag{2.1}$$

Here and further, we assume that $\mathbb{E}|X| < \infty$, and hence the expectation in (2.1) is well defined and finite valued. For $\alpha = 0$, $\mathsf{CVaR}_0(X) = \mathbb{E}[X]$ and $\mathsf{CVaR}_\alpha(X)$ tends to the essential supremum[4] of X as $\alpha\uparrow 1$, so we define $\mathsf{CVaR}_1(X) := \operatorname{ess\,sup}(X)$. Let $F_X(x) := \mathsf{Prob}(X \le x)$ be the cumulative distribution function (CDF) of X and

$$q^-_\alpha(X) := \inf\{t : F_X(x) \ge \alpha\},\ q^+_\alpha(X) := \sup\{t : F_X(x) \le \alpha\},$$

be the left side and right side quantiles of X. If $q^-_\alpha(X) = q^+_\alpha(X)$ we simply denote it by $q_\alpha(X)$. It is well known that for $\alpha \in (0,\ 1)$ the minimum in the right hand side of (2.1) is attained for any $t \in [q^-_\alpha(X), q^+_\alpha(X)]$.

Denote $\bar{q}_\alpha(X) := \mathsf{CVaR}_\alpha(X)$. For $x \in \mathbb{R}$, consider the equation

$$x = \mathsf{CVaR}_\alpha(X), \tag{2.2}$$

with respect to $\alpha \in [0,\ 1]$. It follows from the representation

$$\mathsf{CVaR}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 q^-_\tau(X)d\tau, \tag{2.3}$$

that $\mathsf{CVaR}_\alpha(X)$ is continuous and monotonically increasing in $\alpha \in [0, 1-\kappa\,]$, where $\kappa := \mathsf{Prob}\{X = \operatorname{ess\,sup}(X)\}$. Hence Eq. (2.2) has unique solution $\alpha = \bar{q}_x^{-1}(X)$ for $\mathbb{E}[X] \le x < \operatorname{ess\,sup}(X)$. The buffered probability of exceedance of a random variable X is defined as

$$\bar{p}_x(X) := \begin{cases} 1 - \bar{q}_x^{-1}(X) & \text{if} \quad \mathbb{E}[X] < x < \operatorname{ess\,sup}(X), \\ 1 & \text{if} \quad x \le \mathbb{E}[X], \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

That is,

$$\mathsf{CVaR}_{1-\bar{p}_x(X)}(X) = x, \text{ when } \mathbb{E}[X] \le x < \operatorname{ess\,sup}(X).$$

Consider the following representation of the buffered probability of exceedance of a random variable X (cf. Mafusalov & Uryasev, 2014, Proposition 1):

$$\bar{p}_x(X) = \begin{cases} \inf_{a\ge 0}\mathbb{E}[a(X-x)+1]_+ & \text{if} \quad x < \operatorname{ess\,sup}(X), \\ 0 & \text{if} \quad x \ge \operatorname{ess\,sup}(X). \end{cases} \tag{2.5}$$

Consider

$$\Psi(a, X) := [a(X-x)+1]_+ \ \text{ and } \ \psi(a) := \mathbb{E}[\Psi(a, X)]. \tag{2.6}$$

Note that $\Psi(a, X)$ and hence $\psi(a)$ are convex functions of a. For $\mathbb{E}[X] < x < \operatorname{ess\,sup}(X)$ the set of minimizers $\arg\min_{a\ge 0}\psi(a)$ forms a closed interval $[a_1, a_2]$, where

$$a_1 = 1/(x - q^-_\alpha(X)) \ \text{ and } \ a_2 = 1/(x - q^+_\alpha(X)), \tag{2.7}$$

with α defined by Eq. (2.2). In particular if the quantile $q_\alpha(X)$ is unique, i.e., $q_\alpha(X) = q^-_\alpha(X) = q^+_\alpha(X)$, then

$$\bar{a} = 1/(x - q_\alpha(X)) \tag{2.8}$$

is the unique minimizer of the right hand side of (2.5).

For $\alpha \in (0,\ 1)$ we have that $\mathsf{CVaR}_\alpha(X) > q^-_\alpha(X)$, and hence the numbers a_1 and a_2 are positive when $\mathbb{E}[X] < x$. When $x < \mathbb{E}[X]$, the minimizer in (2.5) is $\bar{a} = 0$, and $\bar{p}_x(X) = 1$. When $x < \operatorname{ess\,sup}(X)$ we have that $X < x$ w.p.1, and hence $\inf_{a\ge 0}\psi(a) = 0 = \bar{p}_x(X)$. When $x = \operatorname{ess\,sup}(X)$,

$$\inf_{a\ge 0}\psi(a) = \mathsf{Prob}(X = x), \tag{2.9}$$

[3] We use notation $[a]_+ := \max\{0, a\}$ for $a \in \mathbb{R}$.

[4] The essential supremum $\operatorname{ess\,sup}(X)$ can be $+\infty$ if the random variable X is unbounded.

Please cite this article as: A. Mafusalov et al., Estimation and asymptotics for buffered probability of exceedance, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.01.021

and hence $\inf_{a\geq 0}\psi(a) > \bar{p}_x(X)$ if $\mathrm{Prob}(X = x) > 0$.

Let $X^1, \ldots, X^N$ be an iid random sample of X. By $\bar{X} := N^{-1}(X^1 + \ldots + X^N)$ we denote average of the sample. Consider

$$\hat{\psi}_N(a) := \frac{1}{N}\sum_{j=1}^{N}\Psi(a, X^j) = \frac{1}{N}\sum_{j=1}^{N}\big[a(X^j - x) + 1\big]_+.$$

That is, $\hat{\psi}_N(a)$ is the empirical (sample average) estimate of the expectation $\psi(a)$. Then a natural estimator of $\bar{p}_x(X)$ is obtained by replacing the probability distribution of X with its empirical estimate. Hence we consider the following estimator of $\bar{p}_x(X)$:

$$\hat{p}_N(x) = \begin{cases} \inf_{a\geq 0}\hat{\psi}_N(a) & \text{if} \quad x < \max\{X^1, \ldots, X^N\}, \\ 0 & \text{if} \quad x \geq \max\{X^1, \ldots, X^N\}. \end{cases} \tag{2.10}$$

By $\hat{a}_N$ we denote a minimizer of the right hand side of (2.10). This minimizer can be computed as above using the empirical distribution given by the considered sample. That is, $\hat{a}_N$ can be any number in the interval $[\hat{a}_{1N}, \hat{a}_{2N}]$, where $\hat{a}_{1N}$ and $\hat{a}_{2N}$ are obtained by replacing in Eq. (2.7) the respective quantiles by their sample estimates. Note that $\hat{p}_N(x) = 1 - \hat{\alpha}$, where $\hat{\alpha}$ is computed by using Eq. (2.2) with $\mathrm{CVaR}_\alpha(X)$ replaced by its empirical estimate. That is

$$x = \widehat{\mathrm{CVaR}}_{\hat{\alpha},N} \tag{2.11}$$

where

$$\widehat{\mathrm{CVaR}}_{\alpha,N} = \inf_{t\in\mathbb{R}}\left\{t + (1-\alpha)^{-1}N^{-1}\sum_{i=1}^{N}[X_i - t]_+\right\}. \tag{2.12}$$

The minimizer in the right hand side of (2.12) is given by the empirical quantile $\bar{t} = \hat{q}_\alpha$. Note that if $x \leq \bar{X}$, then $\hat{a}_N = 0$ and $\hat{p}_N(x) = 1$.

We can view the right hand side of (2.10) as the Sample Average Approximation (SAA) of the stochastic problem in the right hand side of (2.5). Hence some standard results of the statistical inference of the SAA problems can be applied. In particular we have the following results.

Theorem 2.1. *Suppose that $\mathbb{E}|X| < \infty$ and let $\alpha = \bar{q}_x^{-1}(X)$ for $\mathbb{E}[X] \leq x < \operatorname{ess\,sup}(X)$. Then the following holds. (i) The estimator $\hat{p}_N(x)$ converges to $\bar{p}_x(X)$ w.p.1 uniformly in x on any interval $[c, d]$ such that $\mathbb{E}[X] < c \leq d < \operatorname{ess\,sup}(X)$, that is*

$$\lim_{N\to\infty}\sup_{x\in[c,d]}\big|\hat{p}_N(x) - \bar{p}_x(X)\big| = 0, \ \text{w.p.1}. \tag{2.13}$$

(ii) If $\mathbb{E}[X] \leq x < \operatorname{ess\,sup}(X)$, then the bias $\mathbb{E}\big[\hat{p}_N(x)\big] - \bar{p}_x(X)$ of the estimator $\hat{p}_N(x)$ is negative, *and this bias is monotonically decreasing, i.e.,*

$$\mathbb{E}\big[\hat{p}_N(x)\big] \leq \mathbb{E}\big[\hat{p}_{N+1}(x)\big] \leq \bar{p}_x(X). \tag{2.14}$$

(iii) If $\mathbb{E}[X] < x < \operatorname{ess\,sup}(X)$, the quantile $q_\alpha(X)$ is unique and variance

$$\sigma^2(x) := \mathrm{Var}\big\{[\bar{a}(X - x) + 1]_+\big\} \tag{2.15}$$

is finite ($\bar{a}$ is defined in (2.8)), then

$$\hat{p}_N(x) = \frac{1}{N}\sum_{j=1}^{N}[\bar{a}(X^j - x) + 1]_+ + o_p(N^{-1/2}), \tag{2.16}$$

and $N^{1/2}\big(\hat{p}_N(x) - \bar{p}_x(X)\big)$ converges in distribution to normal $\mathcal{N}(0, \sigma^2(x))$.

Proof. Let us observe that the empirical estimate $\widehat{\mathrm{CVaR}}_{\alpha,N}$ converges to $\mathrm{CVaR}_\alpha(X)$ w.p.1 uniformly in $\alpha \in [\gamma_1, \gamma_2]$, where $0 < \gamma_1 \leq \gamma_2 < 1$. Indeed, by using a uniform Law of Large Numbers (e.g., Shapiro, Dentcheva, & Ruszczyński, 2014, Theorem 7.53) we have that $N^{-1}\sum_{j=1}^{N}[X^j - t]_+$ converges to $\mathbb{E}[X - t]_+$ w.p.1 uniformly in t on any bounded interval. For $\alpha \in [\gamma_1, \gamma_2]$ we can restrict the minimization in t in (2.1) and (2.12) to a bounded interval, and hence the uniform convergence of $\widehat{\mathrm{CVaR}}_{\alpha,N}$ to $\mathrm{CVaR}_\alpha(X)$ follows. Since $\mathrm{CVaR}_\alpha(X)$ is continuous and monotonically increasing in α, this implies uniform convergence w.p.1 in $x \in [c, d]$ of the empirical estimate of the inverse function $\bar{q}_x^{-1}(X)$, and hence (2.13) follows:

For proof of assertions (ii) and (iii) we can refer to (Shapiro et al., 2014, Proposition 5.6) and (Shapiro et al., 2014, Theorem 5.7), respectively, by using representation (2.5). Note that the minimization in (2.5) can be restricted to a bounded interval containing numbers a_1 and a_2 defined in (2.7). □

Under assumption (iii) of the above theorem we have that $\hat{a}_N$ converges w.p.1 to $\bar{a}$, and hence the variance $\sigma^2(x)$ can be consistently estimated by

$$\hat{\sigma}^2(x) := \frac{1}{N-1}\sum_{j=1}^{N}\Big([\hat{a}_N(X^j - x) + 1]_+ - \hat{p}_N(x)\Big)^2. \tag{2.17}$$

Remark 1. Of course it follows from the assertion (i) of the above theorem that $\hat{p}_N(x)$ converges to $\bar{p}_x(X)$ w.p.1 for any $\mathbb{E}[X] < x < \operatorname{ess\,sup}(X)$. For $x \leq \mathbb{E}[X]$ the set of minimizers in (2.5) is bounded (unless X is constant), and includes $\bar{a} = 0$. Note that the corresponding function $\Psi(a, X)$ is convex in a. We can apply Shapiro et al., 2014, Theorem 5.4 to conclude that $\hat{p}_N(x)$ converges to $\bar{p}_x(X)$ w.p.1. If $x \geq \operatorname{ess\,sup}(X)$ and $\mathrm{Prob}(X = x) = 0$, then both $\bar{p}_x(X)$ and $\hat{p}_N(x)$ are zeros. Finally consider the case of $x = \operatorname{ess\,sup}(X)$ and $\mathrm{Prob}(X = x) > 0$. Then probability that at least one of X_i is equal to x, and hence $x = \max\{X_1, \ldots, X_N\}$, tends to one. In that case $\hat{p}_N(x)$ converges to $\bar{p}_x(X)$ in probability rather than w.p.1.

Remark 2. If the quantile $q_\alpha(X)$ is not unique, then the set of minimizers in the right hand side of (2.5) is the interval $[a_1, a_2]$, where a_1 and a_2 are defined in (2.7). In that case the asymptotics of $\hat{p}_N(x)$ is given by the minimum of $\hat{\psi}_N(a)$ over $a \in [a_1, a_2]$, and the distribution of $\hat{p}_N(x)$ is not asymptotically normal (cf. Shapiro et al., 2014, Theorem 5.7, Eqs. (5.24) and (5.25)).

Remark 3. When the (true) $\alpha > 0$ we have that $x > \mathbb{E}[X]$. However if α is close to 0, and hence x is close to $\mathbb{E}[X]$, it can happen that $x \leq \bar{X}$ in which case $\hat{p}_N(x) = 1$. Therefore for α close to 0, a better approximation of the distribution of $\hat{p}_N(x)$ will be the mixture[5] of distributions $\delta(1)$ and $\mathcal{N}(0, \sigma^2(x))$ with the respective weights ρ and $1 - \rho$, where $\rho := \mathrm{Prob}(x \leq \bar{X})$. By the CLT the distribution of $\bar{X}$ can be approximated (under standard regularity conditions) by the normal distribution $\mathcal{N}(\mu, \nu^2/N)$, where $\mu := \mathbb{E}[X]$ and $\nu^2 := \mathrm{Var}(X)$. Consequently the probability ρ can be approximated by[6] $1 - \Phi\big(\sqrt{N}(x - \mu)/\nu\big)$. In particular, if $\sqrt{N}(x - \mu)/\nu$ is greater than, say, three, then the probability ρ can be negligibly small and the normal distribution $\mathcal{N}(0, \sigma^2(x))$ could give a good approximation of the distribution of $\hat{p}_N(x)$.

When α is close to one, and hence x is close to[7] $\operatorname{ess\,sup}(X)$ and $F_X(x)$ is close to one, it can happen that $x \geq \max\{X^1, \ldots, X^N\}$ in which case $\hat{p}_N(x) = 0$. The probability

$$\varrho := \mathrm{Prob}\big(x \geq \max\{X^1, \ldots, X^N\}\big) = \prod_{j=1}^{N}\mathrm{Prob}(X^j \leq x)$$
$$= (1 - p)^N \approx e^{-Np}, \tag{2.18}$$

where $p := 1 - F_X(x) = \mathrm{Prob}(X > x)$. Therefore we will need a sample size of order $p^{-1}\ln\varepsilon^{-1}$ to make this probability ϱ less than

[5] $\delta(y)$ denotes probability measure of mass 1 at the point y.

[6] $\Phi(\cdot)$ denotes the cumulative distribution function of standard normal distribution.

[7] If $\operatorname{ess\,sup}(X) = +\infty$, this means that x is large.

Please cite this article as: A. Mafusalov et al., Estimation and asymptotics for buffered probability of exceedance, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.01.021

$\varepsilon > 0$. If the probability ϱ is not small, then we can use the mixture of distributions $\delta(0)$ and $\mathcal{N}(0, \sigma^2(x))$, with the respective weights ϱ and $1-\varrho$, as an approximation of the distribution of $\hat{p}_N(x)$.

Remark 4. The right hand side of (2.16) gives a first order expansion of the estimator $\hat{p}_N(x)$. Under stronger assumptions it is possible to derive a second order term in the corresponding expansion. That is, under appropriate regularity conditions[8],

$$\hat{p}_N(x) - \hat{\psi}_N(\bar{a}) = N^{-1} \inf_{\tau \in \mathbb{R}} \left\{ \tau V + \tfrac{1}{2}\tau^2 \psi''(\bar{a}) \right\} + o_p(N^{-1}) \tag{2.19}$$

$$= -\frac{V^2}{2N\psi''(\bar{a})} + o_p(N^{-1}). \tag{2.20}$$

where $V \sim N(0, \gamma^2)$ with

$$\gamma^2 = \mathrm{Var}\left(\frac{\partial \Psi(\bar{a}, X)}{\partial a} \right) = \mathbb{E}\left[\frac{\partial \Psi(\bar{a}, X)}{\partial a} \right]^2. \tag{2.21}$$

Note that $\psi'(\bar{a}) = \mathbb{E}\left[\frac{\partial \Psi(\bar{a}, X)}{\partial a} \right] = 0$ by optimality of $\bar{a}$.

We have that

$$\mathbb{E}[\hat{\psi}_N(\bar{a})] = \psi(\bar{a}) = \bar{p}_x(X),$$

and hence the bias can be approximated as

$$\mathbb{E}[\hat{p}_N(x)] - \bar{p}_x(X) = -\frac{\gamma^2}{2N\psi''(\bar{a})} + o(N^{-1}). \tag{2.22}$$

Assuming that random variable X has continuous probability density function $f(\cdot)$, we have

$$\psi'(a) = \mathbb{E}\left[\frac{\partial \Psi(a, X)}{\partial a} \right] = \int_{-1/a}^{+\infty} t f(t + x) dt,$$

and hence

$$\psi''(\bar{a}) = \frac{f(x - 1/\bar{a})}{\bar{a}^3} = (x - q_\alpha(X))^3 f(q_\alpha(X)).$$

The assumption of existence of continuous probability density function $f(\cdot)$ is an essential condition needed to ensure existence of the second order derivative $\psi''(\bar{a})$.

3. Rare events

Consider now the case where probability $p := P(X \geq x)$ is very small, say of order 10^{-5} or smaller. In that case, given a sample $X^1, \ldots, X^N$, it will be difficult to employ the corresponding estimator $\hat{p}_N(x)$ in a straightforward way. Recall that $\hat{p}_N(x)$ is zero if x is greater than $\max\{X^1, \ldots, X^N\}$ (see (2.10)). The probability that at least one of the samples $X^1, \ldots, X^N$ is greater than x is $1 - (1-p)^N \approx 1 - e^{-Np}$ (see (2.18)). That is, in order to have a reasonable probability for $\max\{X^1, \ldots, X^N\}$ to be less than x, i.e., for the estimator $\hat{p}_N(x)$ to be greater than zero, we will need a sample size of order of millions. This, of course, could be practically infeasible.

Suppose that X has normal distribution $\mathcal{N}(\mu, \nu^2)$ and $x = \mu + k\nu$, where $k \geq 4$, say. Then the probability $P(X \geq x) = 1 - \Phi(x)$ is very small. For example for $x = \mu + 4\nu$ the probability $p = 1 - \Phi(4) = 3 \times 10^{-5}$. Nevertheless we can proceed as follows. It is known that in the case of normal distribution,

$$\mathsf{CVaR}_\alpha(X) = \mu + \frac{\nu}{(1-\alpha)\sqrt{2\pi}} e^{-z_\alpha^2/2}, \tag{3.1}$$

where $z_\alpha = \Phi^{-1}(\alpha)$. In that case first we can compute the estimates $\hat{\mu} = \bar{X}$ and $\hat{\nu}^2 = \frac{1}{N-1}\sum_{i=1}^N (X_i - \bar{X})^2$, and then to estimate $\bar{p}_x(X)$ by solving (numerically) the equation

$$x = \hat{\mu} + \frac{\hat{\nu}}{(1-\alpha)\sqrt{2\pi}} e^{-z_\alpha^2/2}, \tag{3.2}$$

and setting the estimate $\hat{p}_X(x) = 1 - \hat{\alpha}$. For the estimates $\hat{\mu}$ and $\hat{\nu}$ their confidence intervals can be computed from the same sample. Consequently these confidence intervals can be used to construct a confidence interval for $\bar{p}_x(X)$.

Suppose now that the probability distribution of X is contaminated by another distribution. That is, for some $\gamma \in (0, 1)$ and CDFs F_1 and F_2, the CDF F of X is given as convex combination $F(\cdot) = \gamma F_1(\cdot) + (1-\gamma)F_2(\cdot)$. We have then

$$\mathsf{CVaR}_\alpha(F) \geq \gamma \mathsf{CVaR}_\alpha(F_1) + (1-\gamma)\mathsf{CVaR}_\alpha(F_2), \tag{3.3}$$

as $\mathsf{CVaR}_\alpha(F) = \inf_{t \in \mathbb{R}} \left\{ t + (1-\alpha)^{-1}\mathbb{E}_F[X - t]_+ \right\}$, and $\mathbb{E}_F[\cdot]_+$ is linear in F, and an infimum over a collection of linear functions is concave. It follows that solution of Eq. (2.2) is smaller than solution of equation

$$x = \gamma \mathsf{CVaR}_\alpha(F_1) + (1-\gamma)\mathsf{CVaR}_\alpha(F_2). \tag{3.4}$$

Hence by computing solution $\tilde{\alpha}$ of Eq. (3.4) we obtain $\bar{p}_x(X) \geq 1 - \tilde{\alpha}$, i.e., $1 - \tilde{\alpha}$ gives a lower bound for $\bar{p}_x(X)$. In particular, if F_1 and F_2 are respective normal distributions $\mathcal{N}(\mu_1, \nu_1^2)$ and $\mathcal{N}(\mu_2, \nu_2^2)$, then Eq. (3.4) can be solved numerically using formula (3.1). This approach works universally for those distributions, where analytical formula for CVaR exists, while considering mixtures of distributions allows for close approximations of general distributions. For example, distributions representable as mixtures of normal distributions are particularly popular. Even though the proposed approach is imprecise, the high value of the resulting lower bound may be used as an alarming signal. Notice that since $\bar{p}$, as CVaR, is mixture-concave (Mafusalov & Uryasev, 2014), another lower bound is obtained from $\bar{p}_x(F) \geq \gamma \bar{p}_x(F_1) + (1-\gamma)\bar{p}_x(F_2)$. Finally, it might, in some cases, be efficient to achieve the precise value of the buffered probability, that is, to solve

$$\bar{p}_x(F) = \inf_{a>0} \sum_{i=1}^m \gamma_i \int [a(\xi - x) + 1]_+ dF_i(\xi),$$

where $F = \sum_{i=1}^m \gamma_i F_i$, $\gamma_i \geq 0$, and $\sum_{i=1}^m \gamma_i = 1$. If all distributions F_i are such that partial moment functions $\varphi(F_i, t) := \int [\xi - t]_+ dF_i(\xi)$ are easy to calculate, the resulting expression of $\bar{p}_x(F)$ is a minimum of a convex function (the right-hand side equals $\mathbb{E}[a(X - x) + 1]_+$, which is convex in a), that is

$$\bar{p}_x(F) = \inf_{a>0} a \sum_{i=1}^m \gamma_i \varphi\left(F_i, x - a^{-1}\right).$$

If functions F_i are continuously differentiable with derivatives f_i, the gradient methods should find the solution efficiently, with the use of the first two derivatives of the objective function (recall definition (2.6) of function $\psi(\cdot)$):

$$\frac{d\psi}{da} = \sum_{i=1}^m \gamma_i \left\{ \varphi\left(F_i, x - a^{-1}\right) + a^{-1} F_i\left(x - a^{-1}\right) \right\},$$

$$\frac{d^2\psi}{da^2} = \sum_{i=1}^m \frac{\gamma_i}{a^3} f_i\left(x - a^{-1}\right).$$

We can also approach estimation of $\bar{p}_x(X)$ by using the importance sampling method. That is, suppose that we have a reasonable estimate $F(\cdot)$ of the distribution of X, with respective density $f(\cdot) = F'(\cdot)$. Consider the transformation $Y^j := X^j + c$, $i = 1, \ldots, N$, of the sample, where c is a chosen constant. Then $\mathsf{CVaR}_\alpha(X)$ can

[8] Expansion (2.19) is based on a second order Delta Theorem. Rigourous derivation of (2.19) is not trivial and is beyond the scope of this paper. We refer to Shapiro et al., 2014, p.338 for a discussion of this type results. Of course for (2.19) to make sense the function $\psi(\cdot)$, defined in (2.6), should be at least two times differentiable.

Please cite this article as: A. Mafusalov et al., Estimation and asymptotics for buffered probability of exceedance, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.01.021

be estimated by

$$\widehat{\mathrm{CVaR}}^{c}_{\alpha,N} := \inf_{t\in\mathbb{R}}\left\{t + (1-\alpha)^{-1}N^{-1}\sum_{j=1}^{N} L(Y^j)\big[Y^j - t\big]_+\right\}, \qquad (3.5)$$

where $L(y) := f(y)/f(y-c)$ is the corresponding likelihood ratio. By computing solution $\tilde{\alpha}_N$ of equation

$$x = \widehat{\mathrm{CVaR}}^{c}_{\alpha,N},$$

we obtain an estimate $1-\tilde{\alpha}_N$ of $\bar{p}_x(X)$. Intuitively "good choice" of constant c should be such that $Y^j - t^*$ is positive for many of Y^j values (here t^* is the minimizer in the right hand side of (3.5)). That happens, for example, when α is close to 1, so x is close to the expected value of Y, which, if the estimate f is accurate, is close to the expected value of X plus the shift constant c. With that being said, a potential rule of thumb for constant choice is $c = x - \bar{X}$. This would require a further investigation, see additional illustration in Section 4.

4. Numerical illustration for exponentially distributed variables

As an example we discuss in this section the case of exponential distribution. Consider an exponentially distributed random variable X with parameter $\lambda > 0$, and pdf $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ and $f(x) = 0$ for $x < 0$. Note that $\mathbb{E}[X] = \lambda^{-1}$. The respective quantile function is $q_\alpha(X) = -\lambda^{-1}\ln(1-\alpha)$, and the Conditional Value-at-Risk is

$$\bar{q}_\alpha(X) = -\lambda^{-1}\big[\ln(1-\alpha) - 1\big],\ \alpha \in [0,1).$$

It is easy to see that the solution of equation $-\lambda^{-1}[\ln(1-\alpha) - 1] = x$ yields

$$\bar{p}_x(X) = 1-\alpha = e^{-\lambda(x-\lambda^{-1})} = 1 - F_X(x-\lambda^{-1}),\ \text{for } x > \mathbb{E}[X],$$

while for $x \leq \mathbb{E}[X]$ the bPOE value is $\bar{p}_x(X) = 1$. For the further calculations let us assume that $x \geq \mathbb{E}[X]$. Denote $\alpha = 1 - \bar{p}_x(X)$, then

$$\bar{a} = \frac{1}{x - q_\alpha(X)} = \frac{1}{x + \lambda^{-1}(-\lambda x + 1)} = \lambda,$$

i.e., the optimal value of a is independent of x for exponential distributions. Let us calculate the asymptotic variance of the bPOE estimator $\hat{p}_N(x)$. Since $\bar{a} = \lambda$ we have that asymptotic variance $\sigma^2(x)$, given in (2.15), can be written as

$$\sigma^2(x) = \mathbb{E}\big([\lambda(X-x)+1]_+\big)^2 - \bar{p}_x(X)^2.$$

Denote $b := -\lambda x + 1$, then $\bar{p}_x(X) = e^b$, and

$$\begin{aligned}\sigma^2(x) &= \int_{-b/\lambda}^{\infty}(\lambda\xi + b)^2 f(\xi)d\xi - e^{2b} = e^b\int_0^\infty \zeta^2 e^{-\zeta}d\zeta - e^{2b}\\ &= e^{-\lambda x+1}(2 - e^{-\lambda x+1}).\end{aligned}$$

Let us illustrate Theorem 2.1. Namely, that $N^{1/2}\big(\hat{p}_N(x) - \bar{p}_x(X)\big)$ converges in distribution to normal $\mathcal{N}(0,\sigma^2(x))$. For $\lambda = 1$ and $x = 2$ the bPOE value is $\bar{p}_2(X) \approx 0.368$, while the asymptotic variance is $\sigma^2(2) \approx 0.6$. The convergence in distribution of $N^{1/2}\big(\hat{p}_N(2) - 0.368\big)$ to the asymptotic distribution $\mathcal{N}(0, 0.6)$ is illustrated in Fig. 1. For $x = 5$ the bPOE value is $\bar{p}_5(X) \approx 0.018$ and the asymptotic variance is $\sigma^2(5) \approx 0.036$. It is expected that convergence will be slower for the larger x, see Fig. 2 illustrating $N^{1/2}\big(\hat{p}_N(5) - 0.018\big)$ converges in distribution to $\mathcal{N}(0, 0.036)$. Similarly to the case of large x, convergence to asymptotically normal distribution can be slower for the values x close to $\mathbb{E}[X]$, see Fig. 3 for an illustration. Figs. 1–3 show that for considered example and reasonably large N, theoretical asymptotics were in reasonable agreement with experimental data.

The convergence to the asymptotic distribution, for general distributions, should behave similarly to the illustrated case. In terms of Kolmogorov-Smirnov distance between the distribution of empirical estimate and the limiting distribution, it seems the most likely candidates to attain the maximal CDF difference are the extreme points: cases when the empirical estimate is either 0 or 1. Then, the distance is at least $\frac{1}{2}\max\{\mathrm{Prob}(\hat{p}_N(x) = 0), \mathrm{Prob}(\hat{p}_N(x) = 1)\}$ and, likely, is less than $\max\{\mathrm{Prob}(\hat{p}_N(x) = 0), \mathrm{Prob}(\hat{p}_N(x) = 1)\}$. The $\mathrm{Prob}(\hat{p}_N(x) = 0)$ has appeared in (2.18); it decreases exponentially with N, although the multiplying factor under the exponent could be rather small, which slows down the convergence (primarily at large values of x). As for the other probability, $\mathrm{Prob}(\hat{p}_N(x) = 1) = \mathrm{Prob}(\bar{X} \geq x)$. For $N = 1$, $\mathrm{Prob}(\hat{p}_N(x) = 1) = \mathrm{Prob}(X < \mathbb{E}X)$. For $N > 1$, if $\bar{X}$ is approximated with $\mathcal{N}(\mathbb{E}X, N^{-1}\sigma^2(X))$ and $x = \mathbb{E}X + \kappa\sigma(X)$, then $P(\hat{p} = 1) \approx \Phi(-\kappa N)$. That is, the convergence of $\mathrm{Prob}(\hat{p}_N(x) = 0)$ to 0 starts at $\mathrm{Prob}(X < \mathbb{E}X)$ and is (asymptotically) faster than exponential. Note that choosing x close to $\mathbb{E}X$, which is equivalent to choosing small κ, slows down the convergence.

Let us illustrate bias estimate (2.22). Since $\gamma^2 = \mathbb{E}\Big[\frac{\partial\Psi(\bar{a},X)}{\partial a}\Big]^2$ and

$$\frac{\partial\Psi(\bar{a},X)}{\partial a} = \begin{cases} X - x, & \text{if } X \geq x - \frac{1}{\bar{a}};\\ 0, & \text{otherwise,}\end{cases}$$

we have

$$\begin{aligned}\gamma^2 &= \int_{x-1/\lambda}^{\infty}(\xi - x)^2\lambda e^{-\lambda\xi}d\xi\\ &= \frac{1}{\lambda^2}e^{-\lambda x}\int_{x-1/\lambda}^{\infty}(\lambda(\xi-x))^2 e^{-\lambda(\xi-x)}d\lambda(\xi-x)\\ &= \frac{1}{\lambda^2}e^{-\lambda x}\int_{-1}^{\infty}\zeta^2 e^{-\zeta}d\zeta = \frac{1}{\lambda^2}e^{-\lambda x+1}.\end{aligned}$$

Note further that $\psi''(\bar{a}) = \frac{f(x-1/\bar{a})}{\bar{a}^3} = \frac{\lambda e^{-\lambda x+1}}{\lambda^3} = \gamma^2$. Therefore,

$$\mathbb{E}[\hat{p}_N(x)] - \bar{p}_x(X) = -\frac{1}{2N} + o(N^{-1}).$$

Remark 5. It could be noted that for larger N the theoretical asymptotic CDF is consistently below the corresponding estimated CDF. This behavior is consistent with the fact that the second order term (2.20), in the respective asymptotic expansion, is negative.

Now following the method of Section 3 aimed at rare events, denote $L := \sum_{j=1}^N L(Y^j)$ and rewrite (3.5) as

$$x = \inf_{t\in\mathbb{R}}\left\{t + \frac{N^{-1}L}{(1-\hat{\alpha})}\sum_{j=1}^{N}\frac{L(Y^j)}{L}[Y^j - t]_+\right\}.$$

Denote by $\tilde{p}^c_N(x)$ the importance sampling estimate, namely, $1-\hat{\alpha}$. If $\sum_{j=1}^N \frac{L(Y^j)}{L}Y^j < x < \max_j Y^j$, then, following bPOE calculation formula,

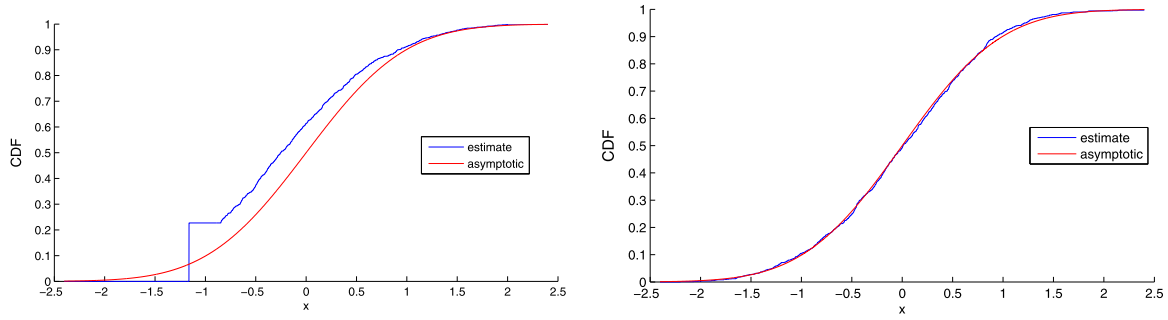
$$\begin{aligned}\tilde{p}^c_N(x) &= N^{-1}L\min_{a\geq 0}\sum_{j=1}^{N}\frac{L(Y^j)}{L}[a(Y^j - x) + 1]_+\\ &= N^{-1}L\hat{p}_{N|L(X+c)}(x-c),\end{aligned}$$

where $N|L(X+c)$ implies that the empirical distribution has probabilities $L(X^j+c)/L$ instead of equal probabilities N^{-1}. That is, the importance sampling estimate equals to a scaled regular estimate calculated under modified sample weights.

Note further that for an exponential variable the value $f(Y)/f(Y-c) = e^{-\lambda c}$, therefore, new weights are still uniform, and

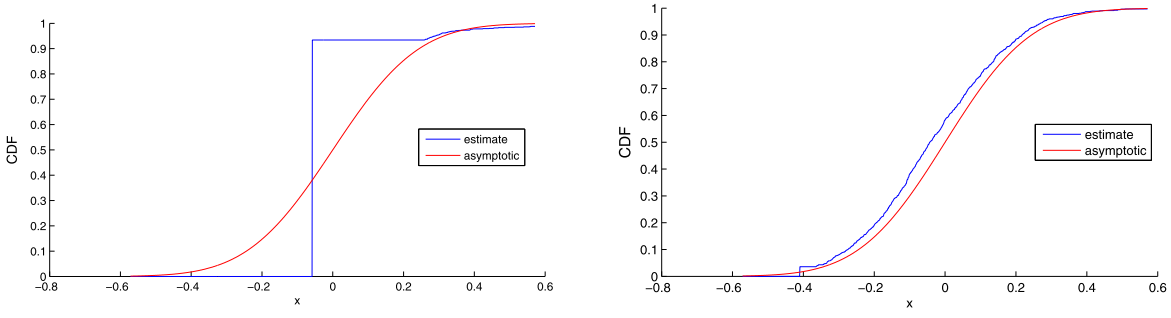
$$\tilde{p}^c_N(x) = e^{-\lambda c}\hat{p}_N(x-c).$$

If $\hat{p}_N(x-c)$ estimates $\bar{p}_{x-c}(X) = e^{-\lambda(x-c)+1}$, then $\tilde{p}^c_N(x)$ estimates $e^{-\lambda c}e^{-\lambda(x-c)+1} = e^{-\lambda x+1} = \bar{p}_x(X)$, the true bPOE value. Note that



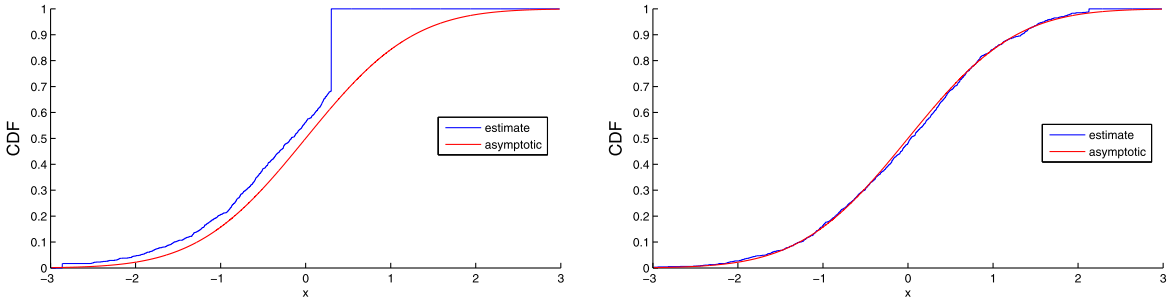
(a) $N = 10$ (b) $N = 500$

Fig. 1. Asymptotic CDF and empirical CDF of $N^{1/2}\big(\hat{p}_N(x) - \bar{p}_x(X)\big)$ for an estimator $\hat{p}_N(x)$ for $x = 2$.



(a) $N = 10$ (b) $N = 500$

Fig. 2. Asymptotic CDF and empirical CDF of $N^{1/2}\big(\hat{p}_N(x) - \bar{p}_x(X)\big)$ for an estimator $\hat{p}_N(x)$ for $x = 5$.



(a) $N = 10$ (b) $N = 500$

Fig. 3. Asymptotic CDF and empirical CDF of $N^{1/2}\big(\hat{p}_N(x) - \bar{p}_x(X)\big)$ for an estimator $\hat{p}_N(x)$ for $x = 1.1$.

this will only work when $x - c > \mathbb{E}[X]$, that is, $c < x - \frac{1}{\lambda}$. The variable $N^{1/2}(\tilde{p}^c_N(x) - \bar{p}_x(X)) = e^{-\lambda c} N^{1/2}(\hat{p}_N(x-c) - \bar{p}_{x-c}(X))$, therefore, converges in distribution to $\mathcal{N}(0, \sigma^2_c(x))$, where

$$\sigma^2_c(x) = (e^{-\lambda c})^2 \sigma^2(x-c) = 2e^{-2\lambda c - \lambda(x-c)+1} - e^{-2\lambda c - 2\lambda(x-c)+2}$$
$$= 2e^{-\lambda(x+c)+1} - e^{2(-\lambda x+1)},$$

i.e., the variance decreases with increase of c and is minimal when $c \to x - \frac{1}{\lambda} \equiv c^*$, then $\sigma^2_c(x) \to (e^{-\lambda x+1})^2$ as opposed to $N^{1/2}(\hat{p}_N(x) - \bar{p}_x(X))$ converging to $\mathcal{N}(0, \sigma^2(x))$ with $\sigma^2(x) = e^{-\lambda x+1}(2 - e^{-\lambda x+1})$. The larger x, the smaller the ratio

$$\frac{\sigma^2_{c^*}(x)}{\sigma^2(x)} = \frac{e^{-\lambda x+1}}{2 - e^{-\lambda x+1}}.$$

5. Optimization of bPOE

Consider the following optimization problem:

$$\min_{y \in \mathcal{Y}} \bar{p}_x(G(y, \xi)). \tag{5.1}$$

Here $\mathcal{Y}$ is a nonempty closed subset of $\mathbb{R}^n$, $\xi \in \mathbb{R}^d$ is a random vector and $G : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ is a *Carathéodory function*, i.e., $G(y, \cdot)$ is measurable for every y and $G(\cdot, \xi)$ is continuous for a.e. ξ (cf., (Rockafellar, 1998, Example 14.29)). We assume that $\mathbb{E}|G(y, \xi)| < \infty$ for every $y \in \mathcal{Y}$, and hence the function $\bar{p}_x(G(y, \xi))$ is well defined and finite valued on $\mathcal{Y}$. We also assume that the set $\mathcal{Y}$ is *convex* and $G(\cdot, \xi)$ is *convex* for a.e. ξ. In particular we consider piecewise affine functions of the form[9]

$$G(y, \xi) := \max_{1 \le i \le m} \big\langle b_i(\xi), y \big\rangle + c_i(\xi), \tag{5.2}$$

with $b_i : \mathbb{R}^d \to \mathbb{R}^n$ and $c_i : \mathbb{R}^d \to \mathbb{R}$, $i = 1, \ldots, m$, being measurable. Note that by (2.4) we have that if $x \ge G(\bar{y}, \xi)$ for some $\bar{y} \in \mathcal{Y}$ and a.e. ξ, then $\bar{p}_x(G(\bar{y}, \xi)) = 0$ and hence $\bar{y}$ is an optimal solution of problem (5.1).

[9] By $\langle x, y \rangle$ we denote the standard scalar product of vectors $x, y \in \mathbb{R}^n$.

Please cite this article as: A. Mafusalov et al., Estimation and asymptotics for buffered probability of exceedance, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.01.021

By (2.5) we can write problem (5.1) in the form

$$\min_{a\geq 0,\, y\in\mathcal{Y}} \psi(a,y), \tag{5.3}$$

where $\Psi(a,y,\xi) := [a(G(y,\xi)-x)+1]^+$ and $\psi(a,y) := \mathbb{E}[\Psi(a,y,\xi)]$. The Sample Average Approximation (SAA) of problem (5.3) is the problem

$$\min_{a\geq 0,\, y\in\mathcal{Y}} \hat{\psi}_N(a,y), \tag{5.4}$$

where $\hat{\psi}_N(a,y) := N^{-1}\sum_{j=1}^N \Psi(a,y,\xi^j)$ with $\xi^1,\ldots,\xi^N$ being an iid sample of the random vector ξ. That is, the SAA problem is obtained by replacing the probability distribution of ξ with its empirical estimate based on the generated random sample.

Let us consider the following reformulation of problems (5.3) and (5.4) (cf., Mafusalov & Uryasev, 2014). By Fenchel–Moreau Theorem we have

$$G(y,\xi) = \sup_{y^*\in\mathbb{R}^n} \langle y^*, y\rangle - G^*(y^*,\xi), \tag{5.5}$$

where $G^*(y^*,\xi)$ is the conjugate of $G(y,\xi)$ given by

$$G^*(y^*,\xi) = \sup_{y\in\mathbb{R}^n} \langle y^*, y\rangle - G(y,\xi). \tag{5.6}$$

Note that $G^*(y^*,\xi)$ can take value $+\infty$ and the maximization in (5.5) is performed over domain

$$D(\xi) := \{y^*\in\mathbb{R}^n : G^*(y^*,\xi) < +\infty\} \tag{5.7}$$

of $G^*(\cdot,\xi)$.

By replacing $y\in\mathbb{R}^n$ with $(y,1)\in\mathbb{R}^{n+1}$ in (5.5), we can write

$$G(y,\xi) - x = \sup_{y^*\in D(\xi)} \big\langle \big(y^*, -G^*(y^*,\xi)-x\big), (y,1)\big\rangle. \tag{5.8}$$

Consequently we can formulate problem (5.3) in the following equivalent form:

$$\min_{z\in\mathcal{Z}} \mathbb{E}[\bar{G}(z,\xi)+1]^+, \tag{5.9}$$

where

$$\mathcal{Z} := \big\{z\in\mathbb{R}^{n+1} : z=(ay,a),\ y\in\mathcal{Y},\ a\geq 0\big\} \tag{5.10}$$

and

$$\bar{G}(z,\xi) := \sup_{y^*\in D(\xi)} \big\langle \big(y^*, -G^*(y^*,\xi)-x\big), z\big\rangle. \tag{5.11}$$

In particular, if $G(y,\xi)$ is of the form (5.2), then

$$\bar{G}(z,\xi) = \max_{1\leq i\leq m} \big\langle \big(b_i(\xi), c_i(\xi)-x\big), z\big\rangle. \tag{5.12}$$

The corresponding SAA problem is

$$\min_{z\in\mathcal{Z}} \frac{1}{N}\sum_{j=1}^N \big[\bar{G}(z,\xi^j)+1\big]_+. \tag{5.13}$$

Convexity of $\mathcal{Y}$ implies that the set (cone) $\mathcal{Z}$ is convex. If the set $\mathcal{Y}$ is polyhedral, defined by a finite number of linear constraints, then the cone $\mathcal{Z}$ is also polyhedral. If, moreover, $G(x,\xi)$ is of the form (5.2), and hence $\bar{G}(x,\xi)$ is of the form (5.12), then the SAA problem (5.13) can be written as a linear programming problem (cf., Mafusalov & Uryasev, 2014). In general closedness of the set $\mathcal{Y}$ does not imply that the cone $\mathcal{Z}$ is closed. The cone $\mathcal{Z}$ is closed in two important cases, namely when $\mathcal{Y}$ is polyhedral or compact. Anyway by continuity arguments the optimal values of problems (5.9) and (5.13) are not changed if the set $\mathcal{Z}$ is replaced by its topological closure $\mathrm{cl}(\mathcal{Z})$.

Denote by ϑ^* and $\hat{\vartheta}_N$ the optimal values of problems (5.9) and (5.13), respectively. Note again that ϑ^* is the optimal value of the "true" problem (5.1) and $\hat{\vartheta}_N$ is the optimal value of its SAA counterpart. Also let $\mathcal{Z}^*$ be the set of optimal solutions of problem (5.9), and $\hat{\mathcal{Z}}_N$ be the set of optimal solutions of the SAA problem (5.13) with $\mathcal{Z}$ replaced by its closure $\mathrm{cl}(\mathcal{Z})$. It could be noted that $(\bar{a},\bar{y})$ is an optimal solution of problem (5.3) iff $\bar{z}=\bar{a}(\bar{y},1)$ is an optimal solution of problem (5.9), and similarly for the corresponding SAA problems. Hence if the cone $\mathcal{Z}$ is closed, then the set $\mathcal{Z}^*$ consists of points $\bar{z}=\bar{a}(\bar{y},1)$ with $\bar{a},\bar{y}$ being an optimal solution of problem (5.3).

Denote $\mathbb{D}(A,B) := \sup_{y\in A}\mathrm{dist}(y,B)$ the deviation of set $A\subset\mathbb{R}^n$ from set $B\subset\mathbb{R}^n$.

Theorem 5.1. *Suppose that the optimal set $\mathcal{Z}^*$ is nonempty and bounded. Then $\hat{\vartheta}_N$ converges w.p.1 to ϑ^* and $\mathbb{D}(\hat{\mathcal{Z}}_N,\mathcal{Z}^*)$ converges w.p.1 to 0 as $N\to\infty$.*

Proof. The function $[\bar{G}(\cdot,\xi)+1]_+$ is convex and the set $\mathrm{cl}(\mathcal{Z})$ is convex. Therefore we can apply Shapiro et al., 2014, Theorem 5.4 to conclude the proof. □

By using Shapiro et al. (2014, Theorem 5.7), together with representation (5.9), we can write the following asymptotics of the SAA estimator $\hat{\vartheta}_N$.

Theorem 5.2. *Suppose that: (i) the optimal set $\mathcal{Z}^*=\{\bar{z}\}$ is a singleton, (ii) variance*

$$\sigma^2 = \mathrm{Var}\big([\bar{G}(\bar{z},\xi)+1]_+\big)$$

is finite, (iii) there exists a measurable function $C(\xi)$ such that $\mathbb{E}[C(\xi)^2]<\infty$ and

$$|\bar{G}(z,\xi)-\bar{G}(z',\xi)| \leq C(\xi)\|z-z'\|$$

for all $z,z'\in\mathcal{Z}$ and a.e. ξ.

Then $N^{1/2}(\hat{\vartheta}_N-\vartheta^*)$ converges in distribution to normal $\mathcal{N}(0,\sigma^2)$.

6. Minimum volume ellipsoid problem: modification with bPOE

The *Minimum Volume Ellipsoid* (MVE) problem is a problem of covering K out of N data points $\xi^1,\ldots,\xi^N$ with an ellipsoid while minimizing the volume of such ellipsoid. If $K=N$, then the problem is called minimum *covering* ellipsoid problem, since all data points must lie inside the ellipsoid. This problem can be formulated as a convex problem and solved very efficiently (Sun & Freund, 2004). The general MVE problem, with $K<N$, is a hard nonconvex problem. A lot of efforts were directed towards finding an approximate solutions, or towards finding good heuristics to solve a similar problem and achieve a similar solution.

Minimum volume ellipsoid problems were studied extensively from perspectives of optimization (Cook, Hawkins, & Weisberg, 1993; Kumar & Yildirim, 2005; Sun & Freund, 2004), statistics (Davies, 1992; Van Aelst & Rousseeuw, 2009), and machine learning (Abou-Moustafa & Ferrie, 2008; Abou-Moustafa & Ferrie, 2007; Hadi, 1992; Shivaswamy & Jebara, 2007; Wei, Löfberg, Feng, Li, & Li, 2007). Current results include various optimization algorithms and applications, primarily in machine learning, as well as statistical properties of the MVE estimator. The general MVE is known for its high resistance to outliers and high (up to 0.5) breakdown value (Davies, 1992). Below we present a new bPOE-MVE estimator. Similar to covering-MVE estimator, it can be computed efficiently as a convex optimization problem. Unlike covering-MVE and similar to MVE, the bPOE-MVE allows data points to lie outside the ellipsoid, but, unlike MVE, it accounts for actual positions of data points lying outside. This is why bPOE-MVE might not be as good as MVE in identifying outliers. Rather than that, we position this new estimator as a good tool to treat "problematic" rare points, which, even if may seem like outliers, appear in datasets on a regular basis.

Suppose that ξ is a random vector of dimension n. Let us parameterize an ellipsoid in $\mathbb{R}^n$ with its center $c\in\mathbb{R}^n$ and its

positive definite shape matrix[10] $Q \in \mathbb{S}^{n\times n}_{++}$, such that the set $\{z \in \mathbb{R}^n : (z-c)^{\mathsf{T}}Q(z-c) = 1\}$ corresponds to the surface of the ellipsoid. Then covering the distribution ξ with an ellipsoid means satisfying the inequality $(\xi - c)^{\mathsf{T}}Q(\xi - c) \le 1$ almost surely. Note that this problem has a solution only when the support $\Xi \subset \mathbb{R}^n$, of the distribution of random vector ξ, is bounded.

In order to make the corresponding problem convex let us introduce the following change of variables, $A := Q^{1/2}$ and $b := Ac$. Then, $(\xi - c)^{\mathsf{T}}Q(\xi - c) = \|A\xi - b\|_2$. Volume of the ellipsoid, parameterized by Q, equals $V = \det(Q^{-1})$ and minimization of ellipsoid volume is equivalent to maximization of $V^{-1/2} = \det(A)$, or to minimization of $-V^{-1/2n} = -(\det A)^{1/n}$, which is a convex function of A[11]. Note that ellipsoid covering of points $\xi^1, \dots, \xi^N$ means that $\|A\xi^i - b\|_2 \le 1$ for $1 \le i \le N$, and the covering-MVE problem is reduced to a finite-dimensional convex programming problem. General MVE problem for the random vector ξ can be written as the following chance constrained problem

$$\begin{array}{ll}\min\limits_{A\in\mathbb{S}^{n\times n}_{++},\, b\in\mathbb{R}^n} & -(\det A)^{1/n} \\ \text{s.t.} & \text{Prob}\big(\|A\xi - b\|_2 \ge 1\big) \le \alpha.\end{array} \tag{6.14}$$

Note that if random vector ξ has finite support $\Xi = \{\xi^1, \dots, \xi^N\}$, with assigned equal probabilities, and $\alpha = (N-K)/N$, then problem (6.14) corresponds to the problem of covering at least K points from the set $\{\xi^1, \dots, \xi^N\}$ with an ellipsoid, while minimizing volume of this ellipsoid. If $\alpha = 0$, then MVE problem becomes covering-MVE and convex. The considered MVE problem has two parameters, the ellipsoid volume V and a measure of the area outside the ellipsoid parameterized by α. The problem (6.14) constraints a measure outside of ellipsoid with parameter α and minimizes volume. It is clear that if value of parameter α increases from 0 to 1, then the optimal volume decreases from $+\infty$ to 0. An alternative problem statement would require the volume to be no greater than V, and minimize the measure of the outside area. That is, another view of the problem (6.14) is obtained when the objective and the constraint switch places:

$$\begin{array}{ll}\min\limits_{A\succ 0, b} & p_1(\|A\xi - b\|_2) \\ \text{s.t.} & (\det A)^{1/n} \ge V^{-1/2n},\end{array} \tag{6.15}$$

where

$$p_1(\|A\xi - b\|_2) := \text{Prob}\big(\|A\xi - b\|_2 \ge 1\big).$$

It is easy to see that the two parametric problem families with parameters α and V share the same frontier of optimal solutions. Consider set

$$S := \Big\{(V, \alpha) : V = \det A^{-1/2}, \alpha = p_1(\|A\xi - b\|_2), A \succ 0\Big\},$$

and suppose that a certain pair $(V_0,\ \alpha_0) \in S$ is dominant, i.e. $S \cap [0, V_0] \times [0, \alpha_0] = (V_0, \alpha_0)$. Then problem (6.14) with parameter α_0 and problem (6.15) with parameter V_0 will have the same optimal solutions. Furthermore, the probability of exceedance $p_1(\|A\xi - b\|_2)$ can be changed to the bPOE $\bar{p}_1(\|A\xi - b\|_2)$, which is the smallest quasi-convex upper bound for POE, and allows for the following formulation:

$$\begin{array}{ll}\min\limits_{A\succ 0, b} & \bar{p}_1(\|A\xi - b\|_2) \\ \text{s.t.} & (\det A)^{1/n} \ge V^{-1/2n}.\end{array} \tag{6.16}$$

This problem, as the previous one, has an upper bound on an ellipsoid volume and minimizes a measure of the furthest from the ellipsoid center points such that "on average" these points lie on the surface, i.e., conditionally expected value of $\|A\xi - b\|_2$ for these points is 1.

An alternative way to convexify the chance constrained problem (6.14) is to substitute the constraint with $\text{CVaR}_\alpha(\|A\xi - b\|_2) \le 1$. In fact, this is the approach of Gotoh and Takeda (2006, 2008), except the squared norm $\|A\xi - b\|_2^2$ is used. These papers show that achieved generalization of MVE problem is also connected to maximum likelihood estimations for normal distributions, that the computation time for the corresponding convex programming problem can be significantly reduced, and that this approach may be adopted to solve multiclass discrimination problem.

Note now that CVaR constraint is equivalent to $\bar{p}_x(\|A\xi - b\|_2) \le \alpha$. Therefore, the correspondence between CVaR-MVE problem family and (6.16) bPOE-MVE problem family is exactly the same as the one between (6.14) and (6.15). That is, when problem parameters V and α are varied, resulting sets of optimal solutions coincide. The formulation (6.16) allows for a convex reformulation, as shown in Section 5. It is easy to see that the function $G(y, \xi) := \|A\xi - b\|_2$, where $y = (A, b) \in \mathcal{Y} := \{(A, b) : A \in \mathbb{S}^{n\times n}_{++}, b \in \mathbb{R}^n\}$, is a Caratheodory function. Since $G(\cdot, \xi)$ is positive homogeneous, i.e., $G(\lambda y, \xi) = \lambda G(y, \xi)$ for $\lambda \ge 0$, its convex conjugate $G^*(y^*, \xi)$, defined in (5.6), is the indicator function of its domain $D(\xi)$ (see (5.7) for definition of $D(\xi)$). Rewriting (5.10) we get

$$\begin{aligned}\mathcal{Z} &= \{z = (ay, a) : y \in \mathcal{Y}, a \ge 0\} \\ &= \Big\{z = (aA, ab, a) : A \in \mathbb{S}^{n\times n}_{++}, b \in \mathbb{R}^n, a \ge 0\Big\} \\ &= \Big\{z = (B, d, a) : B \in \mathbb{S}^{n\times n}_{++}, d \in \mathbb{R}^n, a \ge 0\Big\}.\end{aligned}$$

When optimal z is obtained, optimal solution to the original problem (6.16) can be obtained from $z = (B, d, a)$ and equations $B = aA$ and $d = ab$. Using $G^*(y^*, \xi) = 0$ for $y^* \in D(\xi)$, we rewrite (5.11):

$$\begin{aligned}\bar{G}(z, \xi) = \bar{G}((B, d, a), \xi) &= \sup_{y^*\in D(\xi)} \langle (y^*, -G^*(y^*, \xi) - 1), (B, d, a)\rangle \\ &= -a + \sup_{y^*\in D(\xi)} \langle y^*, (B, d)\rangle - G^*(y^*, \xi) \\ &= G((B, d), \xi) - a = \|B\xi - d\|_2 - a.\end{aligned}$$

Therefore, rewriting (5.13), we can write the corresponding SAA problem

$$\begin{array}{ll}\min\limits_{B,d,a} & \sum_{i=1}^N \big[\|B\xi^i - d\|_2 - a + 1\big]_+ \\ \text{s.t.} & (\det B)^{1/n} \ge aV^{-1/2n}, \\ & B \succ 0,\ \ a \ge 0.\end{array} \tag{6.17}$$

The MVE problems are suited perfectly for multidimensional distributions coming from the elliptical class, that is, when probability density function (pdf) has the form $f(x) = g\big(\sqrt{(x-\mu)^T\Sigma^{-1}(x-\mu)}\big)$, where g is a one-dimensional density function. For bPOE-MVE problem, we show below that the solution for arbitrary Σ and μ may be obtained from the solution to the problem with identity shape matrix $\Sigma = I$ and zero mean $\mu = 0$. Suppose that random vector ξ has pdf $g\big(\sqrt{(x-\mu)^T\Sigma^{-1}(x-\mu)}\big)$, and the function g is decreasing on $[0, +\infty)$, then $\nu := \Sigma^{-1/2}(\xi - \mu) \propto g(\sqrt{y^Ty})$. Denote by B_Σ, d_Σ, a_Σ the optimal solution to the problem for the original random vector ξ with volume constraint parameter V_Σ:

$$\begin{array}{ll}\min\limits_{B,d,a} & \mathbb{E}\big[\|B\xi - d\|_2 - a + 1\big]_+ \\ \text{s.t.} & (\det B)^{1/n} \ge aV_\Sigma^{-1/2n}, \\ & B \succ 0,\ \ a \ge 0.\end{array} \tag{6.18}$$

For the "standardized" random vector ν, take parameter value V_I and denote the optimal solution by B_I, d_I, a_I, and apply $\nu = \Sigma^{-1/2}(\xi - \mu)$:

$$\min_{B,d,a} \quad \mathbb{E}[\|B\nu - d\|_2 - a + 1]_+$$

[10] By $\mathbb{S}^{n\times n}$ we denote the linear space of $n \times n$ symmetric matrices, and by $\mathbb{S}^{n\times n}_{++}$ its subset of positive definite matrices.

[11] To minimize volume, we minimize a convex function $-(\det A)^{1/n}$ rather than a commonly used and also convex function $-\ln\det A$, mostly for the convenience of numerical experiments, since our optimization program code has started from http://cvxr.com/cvx/examples/cvxbook/Ch08_geometric_probs/html/min_vol_elp_finite_set.html

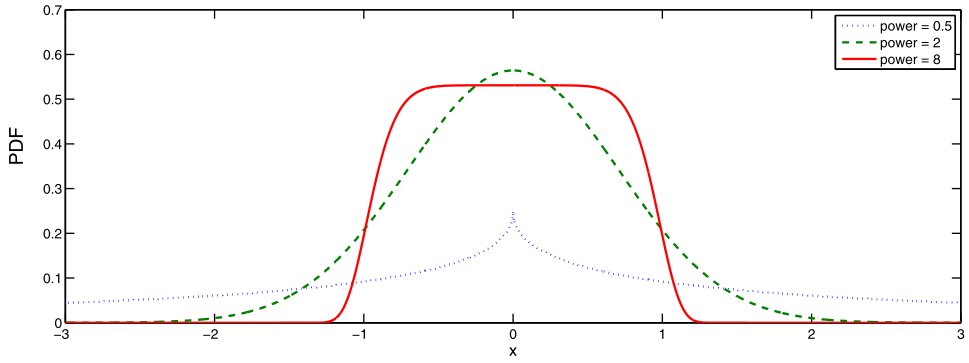


Fig. 4. Probability density functions of exponential power distributions for power $\beta \in \{0.5, 2, 8\}$ and scale $\alpha = 1$.

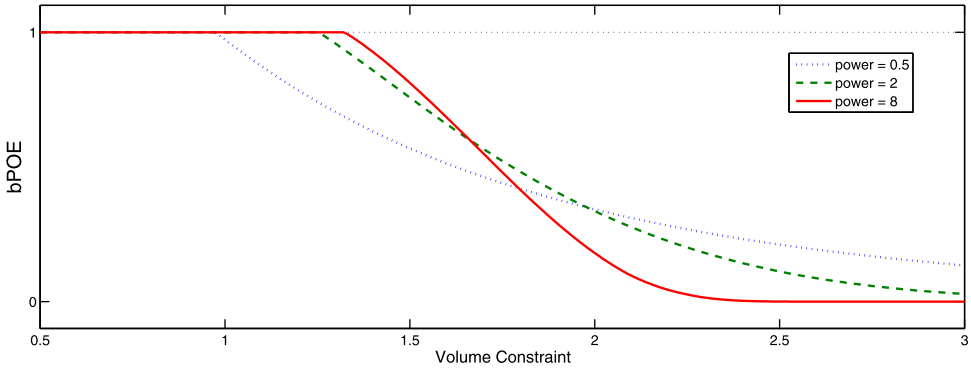


Fig. 5. Optimal bPOE value ϑ^* (vertical axis) for the problem (6.18) as a function of the volume constraint parameter V (horizontal axis), for elliptical exponential power distribution with power values $\beta \in \{0.5, 2, 8\}$ and scale values α such that all covariation matrices are equal to identity matrix.

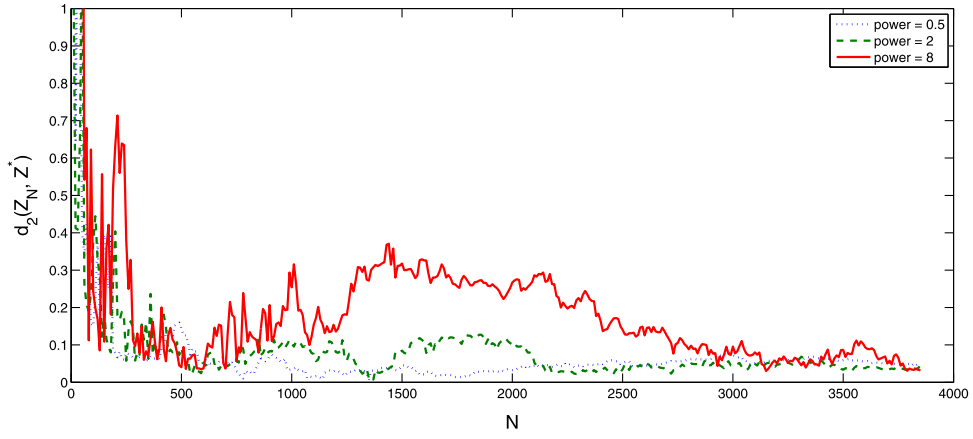


Fig. 6. Euclidean norm of error $d_2(\hat{z}_N, \bar{z}) \equiv \|\hat{z}_N - \bar{z}\|_2$ for optimal to (6.17) solution $\hat{z}_N = (\hat{\beta}_N, \hat{d}_N, \hat{a}_N)$ by the sample size N for elliptical exponential power distribution with power $\beta \in \{0.5, 2, 8\}$. It can be seen that errors converge to 0 with N increasing for all power values.

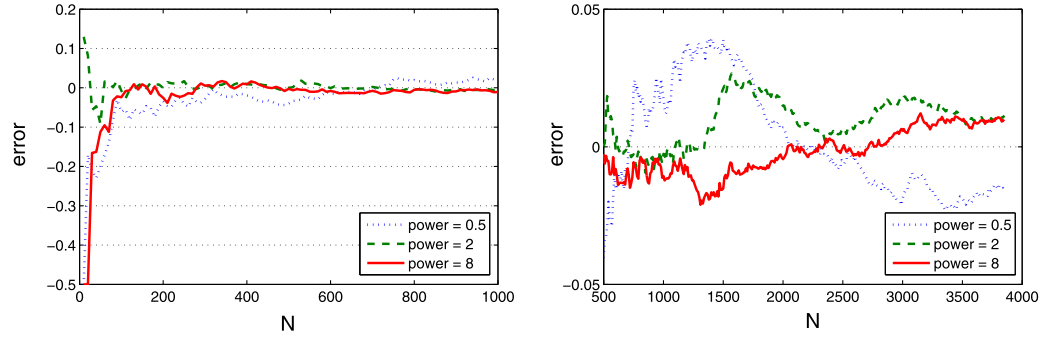
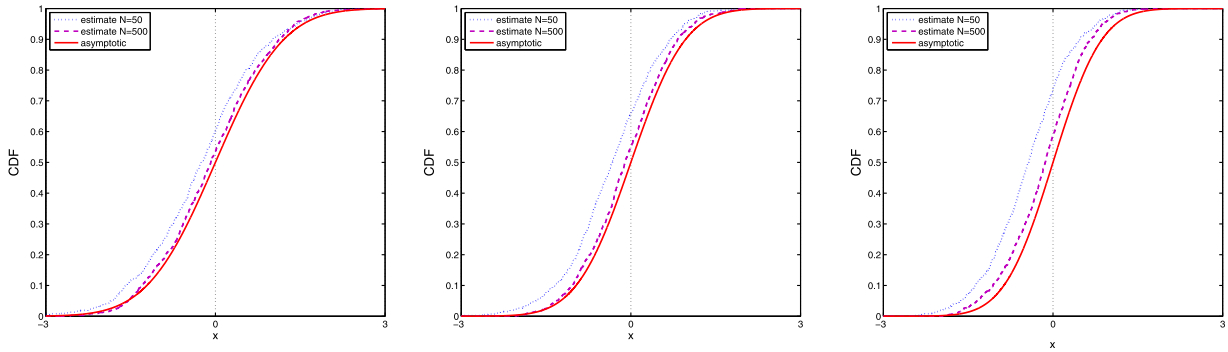


Fig. 7. Error $\hat{\vartheta}_N - \vartheta^* = \hat{\vartheta}_N - 1/2$ of the optimal to (6.17) bPOE value $\hat{\vartheta}_N$ by the sample size N for elliptical exponential power distribution with $\beta \in \{0.5, 2, 8\}$. It can be noted, see left figure presenting interval $N \in [1, 1000]$, that errors converge to 0 with N increasing for all power values. Right figure presents interval $N \in [500, 4000]$. Note that the scale required to plot error values (< 0.05) for $N \in [500, 4000]$ is much smaller than the one required for $N \in [1, 1000]$, with error values ≤ 0.5.

Please cite this article as: A. Mafusalov et al., Estimation and asymptotics for buffered probability of exceedance, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.01.021



(a) $\beta = 0.5$ (b) $\beta = 2$ (c) $\beta = 8$

Fig. 8. Asymptotic normal $\mathcal{N}(0, \sigma^2)$ CDF and empirical CDFs of $N^{1/2}\left(\hat{\vartheta}_N - \vartheta^*\right)$ for elliptical power distribution with power β.

Table 1
Setup of the experiment on weak convergence to the asymptotic normal distribution and obtained estimates. Values β, α are the power and the scale parameters of the power distribution; V is the upper volume constraint; ϑ^* is the optimal value of (6.18); σ^2 is the asymptotic variance for the estimator $\hat{\vartheta}_N$; $\eta_N = N^{1/2}(\hat{\vartheta}_N - \vartheta^*)$; $\hat{\mathbb{E}}$ is an average taken over $M = 1000$ generated samples; $\hat{\sigma}^2$ is an empirical variance calculated over M generated samples. Note that $0 > \hat{\mathbb{E}}\eta_N \to 0$ and $\hat{\sigma}^2(\eta_N) \to \sigma^2$ as $N \to \infty$, as predicted.

power, β	scale, α	V	ϑ^*	σ^2	$\hat{\mathbb{E}}\eta_{50}$	$\hat{\sigma}^2(\eta_{50})$	$\hat{\mathbb{E}}\eta_{500}$	$\hat{\sigma}^2(\eta_{500})$
0.5	1	$1.2 \cdot 10^6$	0.5	0.9	−0.26	1	−0.07	0.83
2	1	2.5	0.5	0.57	−0.33	0.64	−0.09	0.54
8	1	0.54	0.5	0.46	−0.44	0.51	−0.17	0.44

$$= \mathbb{E}[\|B\Sigma^{-1/2}\xi - (B\Sigma^{-1/2}\mu + d)\|_2 - a + 1]_+$$
$$\text{s.t.} \quad (\det B)^{1/n} \geq aV_I^{-1/2n} \Leftrightarrow \left(\det(B\Sigma^{-1/2})\right)^{1/n}$$
$$\geq a(\det\Sigma)^{-1/2n}V_I^{-1/2n}, B \succ 0, \ a \geq 0.$$

Assume that $V_\Sigma = V_I \det\Sigma$, then optimal solutions are connected as follows: $B_\Sigma = B_I\Sigma^{-1/2}$, $d_\Sigma = d_I + B_\Sigma\mu$, $a_\Sigma = a_I$. Therefore, for our purposes, it can be assumed without loss of generality that the elliptical distribution is "standard": it has a zero mean and an identity shape matrix. Further on, because of a symmetry of such distribution, which is a spherical distribution, the optimal ellipsoid is a sphere with a center at zero. Hence $d_I = 0$ and $B_I = a_I V_I^{-1/2n} I$. Noting that $B = aA$ and $d = ab$, we get that in the original problem $A_I = V_I^{-1/2n} I$ and $b_I = 0$. Note also that the same calculations are valid for probability of exceedance minimization with a volume constraint. Therefore, for a class of elliptical distributions, POE minimization and bPOE minimization provide the same optimal solution, but different objective values.

To test the new covering ellipsoid problem behavior, we are varying the tail fatness. One of the ways to do that is to consider a function g from exponential power distribution, $g(x) = C \cdot e^{-\frac{|x|^\beta}{\alpha}}$, see Fig. 4 for an illustration. Normal distribution is a special case of the exponential power distribution with $\beta = 2$.

We generate samples from elliptical exponential power distribution in a following way. We generate a uniformly distributed on a unit sphere random vector $U = X/\|X\|_2$, where $X \propto \mathcal{N}(0, I)$ is a standard normal vector. We generate variable $R \geq 0$ from radial distribution, such that variable UR has a density proportional to $g(\|x\|_2) = C \cdot e^{-\|x\|_2^\beta/\alpha}$. Hence, density for R must be proportional to $x^{n-1}e^{-x^\beta/\alpha}$, therefore, CDF for R is proportional to

$$\int_0^x \zeta^{n-1}e^{-\zeta^\beta/\alpha}d\zeta = C \cdot \int_0^{x^\beta/\alpha} t^{n/\beta-1}e^{-t}dt = C \cdot \gamma(n/\beta, x^\beta/\alpha),$$

where $\gamma(a, x) \equiv \int_0^x t^{a-1}e^{-t}dt$ is an incomplete Gamma-function. Note that CDF of Gamma distribution with parameters k, θ is $C \cdot \gamma(k, \theta x)$. Therefore, if $G \propto \text{Gamma}(n/\beta, 1)$ and $R = \alpha Y^{1/\beta}$, then R has the required distribution. Finally, $\xi = RU\Sigma^{-1/2} + \mu$ has the requested elliptical exponential power distribution.

To compare how the optimal bPOE value decreases for distributions with different power, we fit scaling factors to make covariance matrices equal to identity matrix. Then we vary the upper bound V for the ellipsoid volume and measure optimal bPOE, see Fig. 5 for an illustration. It can be seen that the optimal objective value ϑ^* decreases slower for the distributions with heavier tails.

Further on, we will compare convergence of solutions to the true optimum, as Theorem 5.1 predicts. Since convergence to the true optimal solution and to the asymptotic distribution is slower for both large and small values of bPOE, we take such values of volume constraints for different power values that true optimal bPOE values are equal to 1/2. For the large enough sample size $N = 10^6$ we estimate the true optimal solution $\bar{z} = (\bar{B}, \bar{d}, \bar{a})$ to (6.18). First, we test convergence almost surely for optimal to (6.17) solutions $\hat{z}_N = (\hat{B}_N, \hat{d}_N, \hat{a}_N)$ by measuring $\|\bar{z} - \hat{z}_N\|_2$ and varying N, see Fig. 6. We also measure error $\hat{\vartheta}_N - \vartheta^*$ of optimal values, where $\hat{\vartheta}_N$ is an optimal value to (6.17), and $\vartheta^* = 1/2$ is an optimal value to (6.18). See Fig. 7 for an illustration. It can be noted that both optimal solutions and optimal values converge to the true ones, and that the fluctuations from the true optimum are higher for the lower power values, i.e., for distributions with heavier tails.

Theorem 5.2 shows that the scaled optimal objective values $N^{1/2}(\hat{\vartheta}_N - \vartheta^*)$ converge in distribution to the normal distribution $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 = \text{Var}([\bar{G}(\bar{z}, \xi) + 1]^+)$. With a large sample ($10^6$ observations) we get estimates for σ^2. For the values of $N = 50$ and $N = 500$ we generate $M = 1000$ samples of size N to estimate an empirical distribution of $\hat{\vartheta}_N$. We denote $\eta_N = N^{1/2}(\hat{\vartheta}_N - \vartheta^*)$, and by $\hat{\mathbb{E}}\eta_N$ and $\hat{\sigma}^2(\eta_N)$ we denote average and standard deviation of η_N among M generated samples. Table 1 contains experiment setup and measurements. Note that distributions with heavier tails have larger value of asymptotic variance, but smaller bias. That is, distributions can not be easily ranked on their convergence speed based on their tail heaviness. Note also

Please cite this article as: A. Mafusalov et al., Estimation and asymptotics for buffered probability of exceedance, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.01.021

that with N increasing both bias and estimated variance converge to theoretically predicted values, which supports the result of the theorem. While absolute values of variance are smaller for distributions with lighter tails, values of variance relative to corresponding asymptotic variance are approximately the same among the distributions. For empirical CDFs of η_{50}, η_{500}, and the asymptotic normal distribution, see Fig. 8.

References

Abou-Moustafa, K., & Ferrie, F. (2008). Regularized minimum volume ellipsoid metric for query-based learning. In *Proceedings of seventh international conference on Machine learning and applications, ICMLA'08.* (pp. 188–193). IEEE.

Abou-Moustafa, K. T., & Ferrie, F. P. (2007). The minimum volume ellipsoid metric. In *Pattern recognition* (pp. 335–344). Springer.

Cook, R., Hawkins, D., & Weisberg, S. (1993). Exact iterative computation of the robust multivariate minimum volume ellipsoid estimator. *Statistics & probability letters, 16*(3), 213–218.

Davies, L. (1992). The asymptotics of Rousseeuw's minimum volume ellipsoid estimator. *The Annals of Statistics, 20*(4), 1828–1843.

Gotoh, J.-y., & Takeda, A. (2006). Conditional minimum volume ellipsoid with applications to subset selection for MVE estimator and multiclass discrimination. *Research Reports on Mathematical and Computing Sciences, SERIES B: Operations Research* available at http://www.is.titech.ac.jp/research/research-report/B/B-423.pdf.

Gotoh, J.-y., & Takeda, A. (2008). Conditional minimum volume ellipsoid with application to multiclass discrimination. *Computational Optimization and Applications, 41*(1), 27–51.

Hadi, A. S. (1992). Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological), 54*(3), 761–771.

Kumar, P., & Yildirim, E. A. (2005). Minimum-volume enclosing ellipsoids and core sets. *Journal of Optimization Theory and Applications, 126*(1), 1–21.

Mafusalov, A., & Uryasev, S. (2014). Buffered Probability of Exceedance: Mathematical Properties and Optimization. Technical Report Research Report 2014-1, ISE Dept., University of Florida.

Norton, M., & Uryasev, S. (2014). Maximization of AUC and Buffered AUC in Classification. Technical Report Research Report 2014-2, ISE Dept., University of Florida.

Rockafellar, R. (2009). Safeguarding strategies in risky optimization. In *Proceedings of the international workshop on engineering risk control and optimization, Gainesville, FL.*

Rockafellar, R., & Wets, R. (1998). *Variational Analysis*. Berlin: Springer.

Rockafellar, R. T., & Royset, J. O. (2010). On buffered failure probability in design and optimization of structures. *Reliability Engineering & System Safety, 95*(5), 499–510.

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of risk, 2*, 21–42.

Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2014). *Lectures on stochastic programming: modeling and theory*: 16. SIAM.

Shivaswamy, P. K., & Jebara, T. (2007). Ellipsoidal machines. In *Proceedings of International conference on artificial intelligence and statistics* (pp. 484–491).

Sun, P., & Freund, R. M. (2004). Computation of minimum-volume covering ellipsoids. *Operations Research, 52*(5), 690–706.

Van Aelst, S., & Rousseeuw, P. (2009). Minimum volume ellipsoid. *Wiley Interdisciplinary Reviews: Computational Statistics, 1*(1), 71–82.

Wei, X., Löfberg, J., Feng, Y., Li, Y., & Li, Y. (2007). Enclosing machine learning for class description. In *Proceedings of advances in neural networks–ISNN 2007* (pp. 424–433). Springer.

Please cite this article as: A. Mafusalov et al., Estimation and asymptotics for buffered probability of exceedance, European Journal of Operational Research (2018), https://doi.org/10.1016/j.ejor.2018.01.021