Operations Research Letters 47 (2019) 130–132

Contents lists available at ScienceDirect

Operations Research Letters

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

Derivatives and subderivatives of buffered probability of exceedance

Tong Zhang *, Stan Uryasev, Yongpei Guan

Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, United States

ARTICLE INFO

Article history:
Received 21 October 2018
Received in revised form 6 February 2019
Accepted 6 February 2019
Available online 13 February 2019

Keywords:
Buffered probability of exceedance
CVaR
Gradient
Quasigradient

ABSTRACT

In this letter, we study the derivatives and subderivatives of *buffered probability of exceedance* (bPOE), in which we provide the mathematical expressions with rigorous proofs for the case when bPOE is smooth. Furthermore, we extend the study to a general non-smooth case for which a set of quasigradients are explored, under a mild assumption, i.e., the corresponding random function with respect to the decision variable is convex.

1. Introduction

Due to its advantages in convexity and reduction of optimization to linear and convex programming [9], *Conditional Value-at-Risk* (CVaR) has become a popular risk measure in financial optimization and other engineering areas. As an alternative to CVaR, *buffered probability of exceedance* (bPOE) has been recently introduced in [4], mostly as one minus inverse of CVaR. The bPOE concept is an extension of the buffered failure probability suggested in [8]. By definition, bPOE is the probability of a tail such that the average of this tail equals the threshold. The pair of bPOE and CVaR has advantages as compared to the pair of probability of exceedance (POE) and Value-at-Risk (VaR) because the former maintains quasiconvexity (for bPOE as shown in [4]) and convexity (for CVaR as shown in [9]). This advantage leads to efficient algorithms for solving optimization problems. Moreover, according to [4], bPOE is a monotonic function of the random variable; it is a strictly decreasing function of the threshold on the interval between the expectation and the essential supremum. The multiplicative inverse of the bPOE is a convex function of the threshold, and a piecewise-linear function in the case of discretely distributed random variable. A family of bPOE minimization problems and corresponding CVaR minimization problems share the same set of optimal solutions. Derivatives of bPOE are needed for efficient implementations of optimization methods. These derivatives also can be used for sensitivity analysis in various application areas, such as network optimization [6], data mining algorithms [5,7], among others.

Since bPOE derivatives cannot be derived from the quite general results for Risk, Deviation, Regret, and Error considered in [3], this paper derives formulas for derivatives of bPOE for the smooth case and quasigradients for the general nonsmooth case. For the smooth case, we provide conditions for continuous differentiability of bPOE. For the general case, we assume the convexity of the corresponding random function embedded in bPOE and provide a set of quasigradients belonging to subdifferential of quasiconvex function. We finally provide an example illustrating the formula.

Before we describe our main results, we first list the related definitions as described in [9] and [4] for the completeness of this letter.

Definition 1. Let $V(x) = f(x, y)$ be a random variable with randomness represented by $y \in \mathbb{R}^m$, where $x \in X \subseteq \mathbb{R}^n$ can be seen as a decision variable, and the function $f(x, y)$ is called random function.

Definition 2. Cumulative Distribution Function (CDF) of the random variable $V(x)$:

$$\Psi(x, \zeta) = P[V(x) \leqslant \zeta].$$

Definition 3. Value-at-Risk (VaR) of the random variable $V(x)$:

$$\zeta_\alpha(x) = \min_{\zeta\in\mathbb{R}}\{\zeta \,|\, \Psi(x, \zeta) \geqslant \alpha\}.$$

Definition 4. Upper Value-at-Risk (VaR$^+$) of the random variable $V(x)$:

$$\zeta_\alpha^+(x) = \inf_{\zeta\in\mathbb{R}}\{\zeta \,|\, \Psi(x, \zeta) > \alpha\}.$$

* Correspondence to:303 Weil Hall, Gainesville, FL 32611.
E-mail addresses: tongzhang305@ufl.edu (T. Zhang), uryasev@ufl.edu (S. Uryasev), guan@ise.ufl.edu (Y. Guan).

https://doi.org/10.1016/j.orl.2019.02.002
0167-6377/

Definition 5. The CVaR of the random variable $V(x)$:

$\bar{q}_\alpha(V(x))$ = mean of the α-tail distribution of $V(x)$,

where the distribution in question is the one with distribution function $\Psi_\alpha(x,\cdot)$ defined by

$$\Psi_\alpha(x,\zeta)=\begin{cases}0, & \zeta<\zeta_\alpha(x),\\ (\Psi(x,\zeta)-\alpha)/(1-\alpha), & \text{otherwise.}\end{cases}$$

Definition 6. Upper bPOE of the random function $V(x)$, which is called just bPOE in this paper:

$$\bar{p}_z(V(x))=\begin{cases}0, & z>\sup V(x);\\ P(V(x)=\sup V(x)), & z=\sup V(x);\\ 1-\alpha(x,z), & E[V(x)]<z<\sup V(x);\\ 1, & \text{otherwise,}\end{cases}$$

where for $z\in(E[V(x)],\sup V(x))$, $\alpha(x,z)$ is the inverse of $\bar{q}_\alpha(V(x))$ as a function of α.

In the following sections, we prove the claims based on the inverse relationship between CVaR and bPOE, i.e., we only study derivatives and subderivatives of bPOE with the mathematical expression $\bar{p}_z(V(x))=1-\alpha(x,z)$ when $E[V(x)]<z<\sup V(x)$ as described in Definition 6.

2. Derivatives of smooth case

In the interval $z\in(E[V(x)],\sup V(x))$, we solve for α from the following equation to obtain the bPOE

$$\bar{q}_\alpha(V(x))=z,$$

where z is a real-valued threshold, and the solution to bPOE is a function of x and z. We only need to derive the derivatives of $\alpha(x,z)$ with regard to x and z. Accordingly, to show this, we first let

$$F(x,z,\alpha)=\bar{q}_\alpha(V(x))-z, \tag{1}$$

where $x\in\mathbb{R}^n$ and $z,\alpha\in\mathbb{R}$. In addition, for each $w=(x,z,\alpha)\in\mathbb{R}^{n+2}$, we denote $U(w)$ to an open neighborhood of the point w. Also, for notation convenience, we define the following three sets: $\mathrm{A}=\{\alpha: w=(x,z,\alpha)\in U(w_0)\}$, $\mathrm{B}=\{(\alpha,x): w=(x,z,\alpha)\in U(w_0)\}$, and $\mathrm{C}=\{(x,z): w=(x,z,\alpha)\in U(w_0)\}$.

Theorem 1. *Considering the function $F(x,z,\alpha)$ defined in Eq. (1) and a point $w_0=(x_0,z_0,\alpha_0)\in\mathbb{R}^{n+2}$, if the following conditions are satisfied:*

1. $0<\alpha<1$, $\forall\alpha\in A$,
2. $\zeta_\alpha(x)=\zeta_\alpha^+(x)$, $\forall(\alpha,x)\in B$,
3. *$\zeta_\alpha(x)$ is continuous in both α and x, $\forall(\alpha,x)\in B$,*
4. $z_0\neq\zeta_{\alpha_0}(x_0)$,
5. *The gradient of CVaR w.r.t x, i.e., $\nabla_x\bar{q}_\alpha(V(x))$ is continuous in both α and x, $\forall(\alpha,x)\in B$,*

we have $\alpha=\alpha(x,z)$ continuously differentiable at all the points in C and

$$\nabla_x\,\alpha(x,z)=-\left(\frac{1-\alpha}{z-\zeta_\alpha(x)}\cdot\nabla_x\,\bar{q}_\alpha(V(x))\right)\Bigg|_{\alpha=\alpha(x,z)},$$

$$\frac{\partial}{\partial z}\alpha(x,z)=\frac{1-\alpha}{z-\zeta_\alpha(x)}\Bigg|_{\alpha=\alpha(x,z)}.$$

Proof. We prove the results based on Implicit Function Theorem [2]. We need to prove that all conditions of Implicit Function Theorem are satisfied under our assumptions, and then use the conclusions of Implicit Function Theorem to get the gradient.

First, we prove that the function $F(x,z,\alpha)$ defined in Eq. (1) has continuous partial derivatives at $U(w_0)$. Combining the conditions 1 and 2, according to [9] (Proposition 13 on page 1458), we know that the partial derivative of $F(x,z,\alpha)$ w.r.t α exists and

$$\frac{\partial}{\partial\alpha}F(x,z,\alpha)=\frac{1}{(1-\alpha)^2}E\{[V(x)-\zeta_\alpha(x)]^+\}. \tag{2}$$

Equivalently, we have

$$(1-\alpha)\cdot\frac{\partial}{\partial\alpha}\bar{q}_\alpha(V(x))=\frac{1}{1-\alpha}E\{[V(x)-\zeta_\alpha(x)]^+\}. \tag{3}$$

In [9] (Theorem 10 on page 1454), we know that CVaR can also be rewritten as

$$\bar{q}_\alpha(V(x))=z=\frac{1}{1-\alpha}E\{[V(x)-\zeta_\alpha(x)]^+\}+\zeta_\alpha(x). \tag{4}$$

Eqs. (3) and (4) imply that

$$\frac{\partial}{\partial\alpha}\bar{q}_\alpha(V(x))=\frac{z-\zeta_\alpha(x)}{1-\alpha}. \tag{5}$$

Combining Eqs. (2) and (5), we get

$$\frac{\partial}{\partial\alpha}F(x,z,\alpha)=\frac{z-\zeta_\alpha(x)}{1-\alpha}. \tag{6}$$

From Eq. (6) and condition 3, we get $\frac{\partial F}{\partial\alpha}(x,z,\alpha)$ continuous at $U(w_0)$. Also $\nabla_xF(x,z,\alpha)=\nabla_x\bar{q}_\alpha(V(x))$ is continuous at $U(w_0)$ according to condition 5. And it is easy to observe that $\frac{\partial}{\partial z}F(x,z,\alpha)=-1$ is continuous at $U(w_0)$. So we have proved that the function $F(x,z,\alpha)$ has continuous partial derivatives at $U(w_0)$.

Next, it is obvious that $F(w_0)=0$.

Finally, we have $\frac{\partial F}{\partial\alpha}(x_0,z_0,\alpha_0)\neq0$ due to condition 4.

Thus, all conditions of Implicit Function Theorem [2] are satisfied. Based on conclusions of Implicit Function Theorem, we have (i) $\alpha=\alpha(x,z)$ continuously differentiable everywhere in the set C, (ii) for each component x_i of the vector x,

$$\frac{\partial}{\partial x_i}\alpha(x,z)=-\frac{\partial F}{\partial x_i}(x,z,\alpha)\Big/\frac{\partial F}{\partial\alpha}(x,z,\alpha)=\left(-\frac{1-\alpha}{z-\zeta_\alpha(x)}\cdot\frac{\partial}{\partial x_i}\bar{q}_\alpha(V(x))\right)\Bigg|_{\alpha=\alpha(x,z)},$$

and (iii) $$\frac{\partial}{\partial z}\alpha(x,z)=-\frac{\partial F}{\partial z}(x,z,\alpha)\Big/\frac{\partial F}{\partial\alpha}(x,z,\alpha)=\frac{1-\alpha}{z-\zeta_\alpha(x)}\Bigg|_{\alpha=\alpha(x,z)}. \qquad\square$$

Corollary 1. *Suppose that all conditions of Theorem 1 are satisfied and $E[V(x)]<z<\sup V(x)$. Then, gradient of bPOE, i.e., $\bar{p}_z(V(x))$, w.r.t. x is continuous in both x and z, $\forall(x,z)\in C$ and is presented with the formula*

$$\bar{p}_z(V(x))=\frac{\bar{p}_z(V(x))}{z-\zeta_\alpha(x)}\cdot\nabla_x\,\bar{q}_\alpha(V(x))\Big|_{\alpha=1-\bar{p}_z(V(x))}.$$

Proof. The formula directly follows from the definition of bPOE, since $\bar{p}_z(V(x))=1-\alpha(x,z)$ for $E[V(x)]<z<\sup V(x)$. $\square$

3. Subderivatives of general case

We first define quasigradient of a quasiconvex function as follows.

Definition 7 ([1]). If $f(x)$ is quasiconvex, then g is a quasigradient at the point x_0 if

$$g^T(x-x_0) \geq 0 \Rightarrow f(x) \geq f(x_0).$$

As compared to Section 2, we assume the random function $f(x,y)$ is convex w.r.t. $x \in X$ for every $y \in \mathbb{R}^m$. In addition, we do not require the random variable $V(x) = f(x,y)$ following a continuous distribution. We still keep the restriction that $E|V(x)| < \infty$, i.e., for each $x \in X$, $V(x) = f(x,y) \in \mathcal{L}^1(\Omega)$.

We describe our results in the following Section 3.1 and illustrate an example in Section 3.2.

3.1. The formula and proof

Proposition 1 ([4] *Proposition 4.8 on page 1097*)**.** *For some random vector y, let $f(x,y)$ be a convex function of x on a convex set $X \subseteq \mathbb{R}^n$ and $V(x) = f(x,y) \in \mathcal{L}^1(\Omega)$ for all $x \in X$. Then, for a given $z \in \mathbb{R}$, $\bar{p}_z(V(x))$ is a quasiconvex function of x on X.*

Proof. Convexity of f implies

$$f(x^M, y) \leq \lambda f(x^1, y) + (1-\lambda) f(x^2, y)$$

for $x^M = \lambda x^1 + (1-\lambda)x^2$. Then, based on the monotonicity property of $\bar{p}_z(V)$ as described in [4] (Proposition 3.6 on page 11), we have

$$\bar{p}_z(f(x^M, y)) \leq \bar{p}_z(\lambda f(x^1, y) + (1-\lambda) f(x^2, y)). \tag{7}$$

In addition, because $\bar{p}_z(V)$ is a quasiconvex function of V [4] (Proposition 3.4 on page 10), (i.e., $\bar{p}_z(\lambda V^1 + (1-\lambda)V^2)) \leq \max\{\bar{p}_z(V^1), \bar{p}_z(V^2)\}$,) we have

$$\bar{p}_z(\lambda f(x^1,y) + (1-\lambda) f(x^2,y)) \leq \max\{\bar{p}_z(f(x^1,y)), \bar{p}_z(f(x^2,y))\}. \tag{8}$$

Therefore, combining Eqs. (7) and (8), we have

$$\bar{p}_z(f(x^M, y)) \leq \max\{\bar{p}_z(f(x^1, y)), \bar{p}_z(f(x^2, y))\},$$

i.e., $\bar{p}_z(f(x,y))$ is a quasiconvex function of x. □

Theorem 2. *When the random function $f(x,y)$ is convex w.r.t x, for bPOE's mathematical expression $\bar{p}_z(V(x)) = 1-\alpha(x,z)$, where $z \in (E[V(x)], \sup V(x))$, any subgradient of CVaR at a given point x_0 is also a quasigradient of the corresponding bPOE.*

Proof. First, based on our assumption that $f(x,y)$ is convex in x, according to [9] (Corollary 11 on page 1457), $\bar{q}_\alpha(V(x))$ is convex w.r.t. x as well. Now let g_0 be any subgradient of CVaR at the point x_0, we have

$$\bar{q}_{\alpha_0}(V(x)) \geq \bar{q}_{\alpha_0}(V(x_0)) + g_0^T(x-x_0), \tag{9}$$

for all $x \in X$.

Note here that for any fixed z, we have $\bar{q}_\alpha(V(x)) = z$ is a constant. This implies

$$\bar{q}_{\alpha_0}(V(x_0)) = \bar{q}_\alpha(V(x)) = z \tag{10}$$

for any given z. Then, if $g_0^T(x-x_0) \geq 0$, combining Eqs. (9) and (10), we have

$$\bar{q}_{\alpha_0}(V(x)) \geq \bar{q}_{\alpha_0}(V(x_0)) = \bar{q}_\alpha(V(x)).$$

Due to the monotonically increasing property of CVaR regarding α, we have $\alpha(x_0,z) \geq \alpha(x,z)$, i.e., $1-\alpha(x,z) \geq 1-\alpha(x_0,z)$.

We have proved that

$$\begin{aligned} g_0^T(x-x_0) \geq 0 &\Rightarrow 1-\alpha(x,z) \geq 1-\alpha(x_0,z) \\ &\Rightarrow \bar{p}_z(V(x)) \geq \bar{p}_z(V(x_0)) \end{aligned}$$

for any given z, where the last one follows from the fact that $\bar{p}_z(V(x)) = 1-\alpha(x,z)$ when $z \in (E[V(x)], \sup V(x))$. According to the quasiconvexity of bPOE and Definition 7, we proved that g_0 is also a quasigradient of the corresponding bPOE, and thus the theorem holds. □

3.2. An example

Here we provide an example to help illustrate the formula in Theorem 2 by deriving subderivatives of bPOE with linear random function and discrete distribution.

Assuming there are m scenarios for y each with the equal probability $1/m$ and $f(x,y) = x^T y$, here we describe the procedure to obtain a set of quasigradients of bPOE.

First, for a given x_0, we rank the m values: $f(x_0, y^i)$, $i = 1, \ldots, m$, from the largest to the smallest and get the corresponding list: $y_{(1)}, y_{(2)}, \ldots, y_{(m)}$, where

$$f(x_0, y_{(1)}) \geq f(x_0, y_{(2)}) \geq, \ldots, \geq f(x_0, y_{(m)}).$$

This list is not unique if the set $\{y \in \mathbb{R}^m | f(x_0,y) = f(x_0, y_{(i)})\}$ contains more than one element for some i. But it will not affect our procedure to get the quasigradients.

Then for a given x_0 and α_0, based on Definitions 1 and 5, $\bar{q}_{\alpha_0}(V(x_0)) = \bar{q}_{\alpha_0}(f(x_0,y))$ is just the inner product of x_0 and the "average" of some of the largest $y_{(i)}$'s. The "average" is a subgradient of CVaR at the point x_0. This fact can be derived from the general formula for the gradient of the CVaR deviation (which is the difference of CVaR and mean value), see in [3] formula (9). From Theorem 2, we know that it is also a quasigradient of bPOE at the point x_0.

Finally, we derive the specific mathematical expression for this set of quasigradients of bPOE at the point x_0. Let $k^* = \min\{k \in \mathbb{Z} | k/m < 1-\alpha_0, (k+1)/m \geq 1-\alpha_0\}$, i.e., $k^* = \lceil (1-\alpha_0)m \rceil - 1$, where $\lceil x \rceil = \min\{n \in \mathbb{Z} | n \geq x\}$. Then, this set of "averages" is

$$\left\{ g_0 \in \mathcal{L}^1(\Omega) \middle| g_0 = \sum_{i=1}^{k^*} \frac{1}{(1-\alpha_0)m} y_{(i)} + \frac{1-\alpha_0 - \sum_{i=1}^{k^*} \frac{1}{m}}{1-\alpha_0} y_{(e)} \right\},$$

where $y_{(e)} \in \left\{ y \in \mathcal{L}^1(\Omega) | f(x_0,y) = f(x_0, y_{(k^*+1)}) \right\}$.

Acknowledgments

Research of Stan Uryasev was partially supported by AFOSR, United States under award FA9550-18-1-0391. Research of Yongpei Guan was partially supported by NSF, United States under award 1436749.

References

[1] S. Boyd, L. Vandenberghe, Subgradients, notes for EE364b, Stanford University, 2008.
[2] W. Chen, Multivariable and Vector Analysis, Macquarie University, 2008.
[3] B. Grechuk, M. Zabarankin, Sensitivity analysis in applications with deviation, risk, regret, and error measures, SIAM J. Optim. 27 (4) (2017) 2481–2507.
[4] A. Mafusalov, S. Uryasev, Buffered probability of exceedance: Mathematical properties and optimization, SIAM J. Optim. 28 (2) (2018) 1077–1103.
[5] M. Norton, A. Mafusalov, S. Uryasev, Soft margin support vector classification as buffered probability minimization, J. Mach. Learn. Res. (18) (2017) 1–43.
[6] M. Norton, A. Mafusalov, S. Uryasev, Cardinality of upper average and its application to network optimization, SIAM J. Optim. 28 (2) (2018) 1726–1750.
[7] M. Norton, S. Uryasev, Maximization of auc and buffered auc in binary classification, Math. Program. (2018).
[8] R. Rockafellar, Safeguarding strategies in risky optimization, in: Presentation at the International Workshop on Engineering Risk Control and Optimization, Gainesville, FL, 2009.
[9] R. Rockafellar, S. Uryasev, Conditional value-at-risk for general loss distributions, J. Bank. Finance 26 (7) (2002) 1443–1471.