Buffered Probability of Exceedance: Mathematical Properties and Optimization

Alexander Mafusalov, Stan Uryasev

RESEARCH REPORT 2014-1

Risk Management and Financial Engineering Lab
Department of Industrial and Systems Engineering
303 Weil Hall, University of Florida, Gainesville, FL 32611.
E-mails: `mafusalov@ufl.edu`, `uryasev@ufl.edu`.

First draft: October 2014, This draft: September 2017
Correspondence should be addressed to: Stan Uryasev

Abstract

This paper studies a probabilistic characteristic called *buffered probability of exceedance* (bPOE). It is a function of a random variable and a real-valued threshold. By definition, bPOE is the probability of a tail such that the average of this tail equals to the threshold. This characteristic is an extension of the so-called *buffered failure probability* and it is equal to one minus inverse of the *Conditional Value-at-Risk (CVaR)*. bPOE is a quasi-convex function of the random variable w.r.t. the regular addition operation and a concave function w.r.t. the mixture operation; it 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. The paper provides an efficient calculation formulas for bPOE. Minimization of bPOE is reduced to a convex program for a convex feasible region and to linear programming (LP) for a polyhedral feasible region and discretely distributed random variables. A family of bPOE minimization problems and corresponding CVaR minimization problems share the same set of optimal solutions.

Keywords: probability of exceedance, buffered probability of exceedance, bPOE, conditional value-at-risk, superquantile, superdistribution

---

1. Introduction

Many optimization problems in an uncertain environment deal with undesired occurrences. Natural hazard, portfolio value fluctuation, server overloading, and large nuclear emission value are examples from different areas. We suppose that a loss function for a system should be minimized or constrained. For instance, we can consider a largest (most worrying) loss, which should be treated carefully. A popular approach is to set a threshold for the loss and estimate the number of instances, for which the threshold is exceeded. Then, in optimization setting or with a risk management policy, this number is bounded in some sense, i.e. forced to be lower than a certain fraction of the overall occurrences. For example, nuclear safety regulations bound the frequency of radiation release exceeding a specified threshold. Alternatively, frequency (or probability) of exceedance of a threshold is minimized. Similarly, in data-mining problems the number of misclassified objects, defined by a score-function exceeding some threshold, is minimized. We want to point out that the number of losses exceeding the threshold does not provide any information on how large these losses are. What if these events, even if they have low probability, destroy the whole system or cause an irrecoverable damage? We advocate for an alternative measure which counts the number of the largest losses, such that the average of these losses is equal to the threshold. This measure is more conservative (more restrictive), compared to the measure counting the exceedances. Also, it has exceptional mathematical properties and it can be easily minimized or constrained.

Let random variable X describe losses, and let $x \in \mathbb{R}$ be a threshold. The fraction of losses exceeding the threshold equals the probability $p_x(X) = P(X > x)$, called the *probability of exceedance* (POE). The POE is also known as the complimentary cumulative distibution function (CCDF), tail distribution, exceedance, survival function, and reliability function.

To explain the alternative to the POE, let us assume that the loss probability distribution is continuous, that is $P(X = x) = 0$ for all x. Then, consider a measure that equals to $P(X > q)$, where q is such that $E[X|X > q] = x$; i.e., the expectation of loss, X, conditioning that it exceeds q, is equal to the threshold x. It is easy to see that $q < x$ and $P(X > q) > P(X > x)$, that is, the considered measure is more conservative than POE and has a "safety buffer", both in loss value, $x - q > 0$, and in probability, $P(X > q) - P(X > x) > 0$. Therefore, this measure is called the *buffered probability of exceedance* (bPOE), denoted by $\bar{p}_x(X)$. In general, we do not require distribution to be continuous, however, it is required that $E|X| < \infty$, that is, we assume $X \in \mathcal{L}^1(\Omega)$. The formal definition of bPOE is given in section 2.

The bPOE concept is an extension of the buffered failure probability suggested by Rockafellar [18] and explored by Rockafellar and Royset [19]. The paper [19] studied two characteristics: the *failure probability*, which is a special case of POE with $x = 0$, i.e., $p_0(X)$, and the *buffered failure probability*, which is a special case of bPOE with $x = 0$, i.e., $\bar{p}_0(X)$. Of course, $p_x(X) = p_0(X-x)$ and $\bar{p}_x(X) = \bar{p}_0(X-x)$, but it can be beneficial to consider $\bar{p}_x(X)$ for a fixed X as a function of x, see sections 3.1 and 4.1.

Further we want to point out that bPOE concept is closely related with the superdistribution function defined in paper [20]. For a random variable X with the distribution function F_X, the *superdistribution* function $\bar{F}_X$ is an inverse of superquantile (CVaR), see (2) for a formal definition. The paper [20] showed that the superdistribution function $\bar{F}_X$ is a distribution function of some other auxiliary random variable $\bar{X}$. bPOE can be defined through the superdistribution function as follows: $\bar{p}_x(X) = 1 - \bar{F}_X(x)$.

Book [27] considers Chebyshev-type family of inequalities with CVaR deviation $\mathrm{CVaR}_\alpha(X - EX)$, where parameter $\alpha \in [0, 1]$, and shows that the tightest inequality in the family is obtained for $\alpha = \bar{p}_x(X)$, and the tightest inequality itself reduces to

$$p_x(X) \leq \bar{p}_x(X). \tag{1}$$

Inequality (1) is one of the motivations for introducing bPOE as an alternative to POE. Paper [19] uses inequality (1) to point out that the buffered failure probability is a conservative estimate of the failure probability. Similarly, bPOE is a conservative estimate of POE.

This paper uses the notation $\mathrm{CVaR}_\alpha(X)$ for *conditional-value-at-risk* (CVaR) for a random variable X and a confidence level $\alpha \in [0,1]$, as defined in [21]. Further, for CVaR we use an alternative name *superquantile*, $\bar{q}_\alpha(X)$, similar to paper [20]. The motivation for this term is the similarity to quantile, $q_\alpha(X)$, which is popular in engineering literature. That is,

$$\bar{q}_\alpha(X) = \mathrm{CVaR}_\alpha(X) = \int_\alpha^1 q_p(X)dp = \min_{c\in\mathbb{R}}\ c + (1-\alpha)^{-1}E[\max\{0, X-c\}].$$

Inequality (1) is similar to the well-known inequality $q_\alpha(X) \le \bar{q}_\alpha(X)$. Tightness of quantile vs. superquantile inequality was studied in [8], where Theorem 4.67 proves that superquantile is the smallest law-invariant[1] convex measure of risk that dominates quantile. Also, quantile vs. superquantile inequality, upper bounds for POE, and approximation of chance constraint $p_x(X) \le \alpha$ were studied in [11], where constraint $\bar{q}_\alpha(X) \le x$ was shown to be the best approximation for $p_x(X) \le \alpha$ in a special class of convex constraints.

This paper considers the class of Schur-convex functions, see [15], i.e., functions consistent with convex stochastic order, see [24]. These functions include both convex law-invariant functions and quasi-convex law-invariant functions. Basically, if the decision maker is risk-averse, his objective should be Schur-convex. Interestingly, there is always a single "optimal" upper bound in this class, see section 3.4. It is proved that superquantile is the smallest Schur-convex upper bound for quantile, see proposition Appendix B.3. Also, bPOE is the smallest Schur-convex upper bound for POE, see proposition 3.9.

There are common traits between bPOE and partial moment function $H(x) = E[X - x]^+$, studied in [4]. Here and further $[x]^+ \equiv \max\{x, 0\}$. Correspondence to Lorenz curve and second order stochastic dominance [5], as well as properties w.r.t. threshold, can make the two functions seem alike, see section 3.1 for details.

This paper studies mathematical properties of bPOE, and there are several publications and working papers studying bPOE and its applications, which refer to the results of this paper. Paper [14] uses probability representation of AUC classification metric to define a new metric, the buffered AUC. Paper [12] considers a projection of bPOE on Euclidean space, so called Cardinality of Upper Average, and utilizes this function to network flow optimization problems. Paper [13] introduces a new Support Vector Machine parametric family of formulations via bPOE minimization, and proves that this family is equivalent to C-SVM problem, which interprets the C-SVM in statistics setting. Paper [26] views bPOE in the perspective of server loads and formulates an efficient program for certain bPOE minimization problem. Paper [2] promotes bPOE estimation for the purposes of environmental engineering and discusses bPOE calculation for finite samples. Paper [25] minimizes bPOE in bond immunization problem.

This paper is organized as follows. Section 2 proves several formulas for efficient calculation of $\bar{p}_x(X)$. In the case of a finite number of scenarios (atoms) in the loss distribution bPOE is easily calculated, similar to POE. Section 3.1 investigates mathematical properties of $\bar{p}_x(X)$ w.r.t. the threshold x. We show that bPOE is continuous with respect to the threshold x for general distributions, while POE is not necessarily continuous and, for discrete distributions, is a step function of x. Section 3.2 establishes mathematical properties of $\bar{p}_x(X)$ w.r.t. the random variable X. We prove that bPOE is a closed quasi-convex function of the random variable X, which is an attractive property

[1] A function is called law-invariant if it takes equal values on identically distributed random variables.

for optimization, while optimization of discontinuous POE for discrete distribution is a combinatorial problem. Quasi-convexity of the objective implies that the set of minima is convex (and a strict local minimum is also a global one). Section 3.3 defines upper bPOE and compares its mathematical properties with lower bPOE. Section 3.4 studies tightness of inequality (1). This paper also proves that bPOE is the tightest (smallest) upper bound for POE among quasi-convex functions defined on atomless $\mathcal{L}^1(\Omega)$ space of random variables. Section 4 studies minimization of $\bar{p}_x(X)$ over a feasible region $X \in \mathcal{X}$. bPOE minimization can be performed by iteratively calling a superquantile minimization algorithm. For a convex feasible region and convex loss function, bPOE minimization is formulated as a convex program.

2. Calculation Formulas for bPOE

This section formally defines buffered probability of exceedance and provides equivalent representations, which can be used for efficient calculation. In particular, proposition 2.1 proves that bPOE can be calculated with a partial moment minimization problem, which has a single scalar variable. Corollary 2.2 connects the optimal solution to a certain quantile value. Another corollary 2.3 demonstrates that in case of finite number of scenarios, minimization can be avoided, and, after establishing bPOE values in several special knot points, all intermediate values are filled in with a linear approximation of bPOE's reciprocal. This section is finalized with an alternative dual formulation for bPOE, see proposition 2.4.

The idea of bPOE comes from taking inverse function of superquantile. The notation $\bar{q}(\alpha; X)$ is used to express superquantile $\bar{q}_\alpha(X) = \bar{q}(\alpha; X)$ as a function of parameter α, hence, $\bar{q}^{-1}(x; X)$ should be interpreted as an inverse function of superquantile as a function of α. As α changes from 0 to 1, $\bar{q}_\alpha(X)$ changes from EX to $\sup X$, hence, the domain for the inverse function is $(EX, \sup X)$, where the inverse is uniquely defined because $\bar{q}(\alpha; X)$ strictly increasing on $[0, P(X < \sup X)]$, see, e.g., proposition Appendix A.1. The superdistribution function was introduced in [20] as an inverse of superquantile (CVaR):

$$\bar{F}_X(x) = \begin{cases} 1, & \text{for } x \geq \sup X; \\ \bar{q}^{-1}(x; X), & \text{for } EX < x < \sup X; \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Let U be a random variable uniformly distributed on $[0, 1]$. The auxiliary random variable $\bar{X} = \bar{q}(U; X)$ has the distribution function $F_{\bar{X}}(x) = \bar{F}_X(x)$. In a similar fashion, bPOE is defined as one minus inverse of superquantile.

This paper considers a general probability space $(\Omega, \mathcal{F}, P)$, unless specified otherwise. In some application-focused derivations it is noted that the space is assumed to be finite, while for some theoretical properties it is assumed that the space is non-atomic. We consider the space $\mathcal{L}^1(\Omega)$ of random variables with the finite first moment.

This paper introduces two versions of bPOE, the *lower buffered probability of exceedance* $\bar{p}^-_x(X) = \bar{p}_x(X)$, see definition 1, and the *upper buffered probability of exceedance* $\bar{p}^+_x(X)$, see definition 2. The upper and lower bPOE differ only at the point $x = \sup X$. If it is not specified otherwise, the term bPOE is used for the lower bPOE.

Definition 1. *For a random variable $X \in \mathcal{L}^1(\Omega)$ and $x \in \mathbb{R}$, the buffered probability of* *exceedance is defined as*

$$\bar{p}_x(X) = \begin{cases} 0, & \textit{for } x \geq \sup X; \\ 1 - \bar{q}^{-1}(x; X), & \textit{for } EX < x < \sup X; \\ 1, & \textit{otherwise.} \end{cases} \tag{3}$$

where for $x \in (EX, \sup X)$, $\bar{q}^{-1}(x; X)$ is the inverse of $\bar{q}(\alpha; X)$ as a function of α.

An implicit way to define the bPOE would be through the following level set equivalence:

$$\bar{p}_x(X) \leq p \Leftrightarrow \bar{q}_{1-p}(X) \leq x.$$

Instead, this equivalence is proved as a part of proposition 3.4.

A minimization formula for bPOE was discovered in [14], where Proposition 1 proves equivalent optimization formulation for upper bPOE, $\bar{p}_x^+(X)$. The proposition below shares the same idea, but provides the proof directly for lower bPOE, $\bar{p}_x(X)$, and in a more concise manner.

Proposition 2.1. *For a random variable $X \in \mathcal{L}^1(\Omega)$ and $x \in \mathbb{R}$,*

$$\bar{p}_x(X) = \begin{cases} \min_{a \geq 0} E[a(X - x) + 1]^+, & \textit{if } x \neq \sup X; \\ 0, & \textit{if } x = \sup X. \end{cases} \tag{4}$$

Proof. In the definition of bPOE we have three cases:

1. $\bar{p}_x(X) = 1 - \bar{q}^{-1}(x; X)$ when $EX < x < \sup X$,
2. $\bar{p}_x(X) = 1$ when $x < EX$ or $x = EX < \sup X$,
3. $\bar{p}_x(X) = 0$ when $x \geq \sup X$.

Let us prove the proposition case by case.

1. Let $EX < x < \sup X$, and take $x = 0$. Since $\bar{q}_\alpha(X)$ is a strictly increasing function of α on $\alpha \in [0, 1 - P(X = \sup X)]$, then equation $\bar{q}_p(X) = 0$ has a unique solution $0 < p^* < 1$ for $EX < x < \sup X$. Then, $\bar{p}_0(X) = p^*$ such that $\min_c c + \frac{1}{p^*} E[X - c]^+ = 0$ (note that optimal c is a p^*-quantile of X, hence, is finite). Since $\bar{q}_\alpha(X)$ is an increasing function of parameter α, then we can reformulate $\bar{p}_0(X) = \min_p p$ such that $\min_c c + \frac{1}{p} E[X - c]^+ \leq 0$. Therefore,

$$\begin{aligned} \bar{p}_0(X) = \min_{p,c} \quad & p \\ s.t. \quad & c + \frac{1}{p} E[X - c]^+ \leq 0. \end{aligned}$$

   Optimal $c^* < 0$, since $c^* \leq c^* + \frac{1}{p^*} E[X - c^*]^+ \leq 0$, and $c^* = 0$ implies $\sup X \leq 0$, which is not the case we consider. Therefore,

$$\begin{aligned} \bar{p}_0(X) = \min_{p,c} \quad & p \\ s.t. \quad & p \frac{c}{|c|} + E\left[\frac{1}{|c|} X - \frac{c}{|c|}\right]^+ \leq 0. \end{aligned}$$

Since $c^* < 0$, then $\frac{c^*}{|c^*|} = -1$. Further, denoting $a = \frac{1}{|c|}$ (optimal solution for (4), $a^* = \frac{1}{|c^*|} < +\infty$), we have

$$\begin{aligned} \bar{p}_0(X) = \min_{p,a>0} \quad & p \\ s.t. \quad & E[aX+1]^+ \le p. \end{aligned}$$

and, therefore,

$$\bar{p}_0(X) = \min_{a\ge 0} E[aX+1]^+.$$

Note that change $a > 0$ to $a \ge 0$ includes value 1 as a feasible value of the objective, which does not affect the case considered. Finally, since $\bar{p}_x(X) = \bar{p}_0(X - x)$, then

$$\bar{p}_x(X) = \min_{a\ge 0} E[a(X-x)+1]^+.$$

2. When $EX \ge x$, we have $E[a(X-x)+1]^+ \ge aE(X-x)+1 \ge 1$. Note also that $E[a(X-x)+1]^+ = 1$ for $a = 0$. Therefore, $\min_{a\ge 0} E[a(X-x)+1]^+ = 1$.
3. For $x = \sup X$, by the formula, $\bar{p}_x(X) = 0$. Consider $x > \sup X$, i.e., $X - x \le -\varepsilon < 0$. Taking $a = \frac{1}{\varepsilon}$ makes $a(X-x) \le -1$, therefore, $\min_{a\ge 0} E[a(X-x)+1]^+ = 0$.

□

The one-dimensional optimization program in (4) is convex, and in the case of finite probability space can be reduced to LP. Note that the presented formula gives bPOE a representation similar to the well-known superquantile representation $\bar{q}_\alpha(X) = \min_c c + (1-\alpha)^{-1}E[X-c]^+$, where an optimal solution is the corresponding quantile, $c^* = q_\alpha(X)$. Note also, from the proof of the proposition, that optimal solutions of the two representations (3) and (4) are related, for $x = 0$, with $a^* = (0 - c^*)^{-1}$. This observation is explored in detail with the following corollary.

Corollary 2.2. *For $X \in \mathcal{L}^1(\Omega)$ and $EX < x < \sup X$,*

$$\bar{p}_x(X) = 1 - \bar{q}^{-1}(x; X) = \min_{c<x} \frac{E[X-c]^+}{x-c}. \tag{5}$$

Furthermore, for $x = \bar{q}_\alpha(X)$, where $\alpha \in (0,1)$, it is valid that

$$q_\alpha(X) \in \arg\min_{c<x} \frac{E[X-c]^+}{x-c},$$

and, consequently,

$$\bar{p}_x(X) = \frac{E[X - q_\alpha(X)]^+}{\bar{q}_\alpha(X) - q_\alpha(X)}.$$

Proof. Since $EX < x < \sup X$, then $\bar{q}^{-1}(x; X) \in (0,1)$, therefore, $a = 0$ is not optimal for $\min_{a\ge 0} E[a(X-x)+1]^+$. Therefore, change of variable $a \to \frac{1}{x-c}$ leads to an equivalent program:

$$\min_{a\ge 0} E[a(X-x)+1]^+ = \min_{c<x} E\left[\frac{1}{x-c}(X-x)+1\right]^+ = \min_{c<x} \frac{E[X-c]^+}{x-c}.$$

Note that if $x = \bar{q}_\alpha(X)$, then $\bar{p}_x(X) = 1 - \bar{q}^{-1}(x; X) = 1 - \alpha$. Since $\bar{q}_\alpha(X) = q_\alpha(X) + \frac{1}{1-\alpha} E[X - q_\alpha]^+$, then

$$\bar{p}_x(X) = 1 - \alpha = \frac{E[X - q_\alpha(X)]^+}{\bar{q}_\alpha(X) - q_\alpha(X)},$$

that is, $q_\alpha(X) \in \arg\min_{c<x} \frac{E[X-c]^+}{x-c}$. □

Since an optimal solution c^* in formula (5) is a quantile, in case of discretely distributed random variable, one of optimal solutions coincides with an atom. Further we shown how to find this atom and how to calculate bPOE in this case without solving the optimization problem. Let X be a discretely distributed random variable with atoms $\{x^i\}_{i=1}^N$, and probabilities $\{p^i\}_{i=1}^N$, where $x^i \le x^{i+1}$, $i = 1, \ldots, N-1$, and N is either finite or $N = \infty$. For confidence levels $\alpha^j = \sum_{i=1}^j p^i$, where $j = 0, \ldots, N$, let us denote the corresponding superquantiles $\bar{x}^j = \sum_{i=j+1}^N x^i p^i / (1 - \alpha^j)$, with $\bar{x}^N = x^N$ for finite N and $\bar{x}^N = \lim_{i\to\infty} x^i$ for $N = \infty$. Then, $\bar{p}_x(X) = 1$ for $x \le \bar{x}^0 = EX$, $\bar{p}_x(X) = 0$ for $x \ge \bar{x}^N = \sup X$, and $\bar{p}_{\bar{x}^j}(X) = 1 - \alpha^j$ for $j = 0, \ldots, N-1$.

Corollary 2.3. *For discretely distributed $X \in \mathcal{L}^1(\Omega)$, the function $1/\bar{p}_x(X)$ is a piecewise-linear function of x on $(-\infty, \sup X)$. In particular,*

$$\bar{p}_x(X) = \frac{E[X - x^{j+1}]^+}{x - x^{j+1}} = \frac{\sum_{i=j+1}^N p^i [x^i - x^{j+1}]^+}{x - x^{j+1}}, \tag{6}$$

for $\bar{x}^j < x < \bar{x}^{j+1}$, where $j = 0, \ldots, N-1$.

Proof. Note that for $\bar{x}^j < \bar{q}_\alpha(X) < \bar{x}^{j+1}$ we have $\alpha_j < \alpha < \alpha_{j+1}$, therefore, $q_\alpha(X) = x^{j+1}$. Therefore, formula (6) is implied by corollary 2.2. Equality $\bar{p}_x(X) = 1$ for $x \le EX$, continuity of $\bar{p}_x(X)$ for $x \in [EX, \sup X)$, see proposition 3.1, and (6) imply the piecewise-linearity of $1/\bar{p}_x(X)$. □

Suppose that the probability space $(\Omega, \mathcal{F}, P)$ is non-atomic. Since bPOE is a tail probability where conditional expectation is x, it is clear that for any other event, if conditional probability is at least x, then the event probability is no greater than bPOE. In other words, bPOE equals the maximal probability of an event A such that the expectation of X conditional to A is greater than the threshold x. The following proposition confirms that, and, therefore, provides an alternative way to interpret bPOE.

Proposition 2.4. *For a random variable $X \in \mathcal{L}^1(\Omega)$, defined on a non-atomic probability space $(\Omega, \mathcal{F}, P)$, and for $x < \sup X$,*

$$\bar{p}_x(X) = \sup_{A \in \mathcal{F}} P(A) \qquad s.t. \quad E(X|A) \ge x. \tag{7}$$

Proof. Define $I_A(\omega) = \{1, \text{ if } \omega \in A;\ 0, \text{ otherwise}\}$ and $W = E(I_A|X)$. Note that $0 \le W \le 1$ and, since $(\Omega, \mathcal{F}, P)$ is non-atomic, for each random variable $0 \le W \le 1$ there exists $A \in \mathcal{F}$ such that $W = E(I_A|X)$. Then, due to $P(A) = EW$ and $E(X|A) = EXW/EW$, the (7) is equivalent to $\sup_W\{EW | EXW \ge xEW, 0 \le W \le 1\}$. Taking a dual problem gives $\inf_{M\ge0,a\ge0} \sup_{W\ge0}\{EW + EM(1-W) + aEW(X-x)\}$, which results in $\inf_{M\ge0,a\ge0}\{EM | M \ge a(X-x)+1\}$, for which $M = [a(X-x)+1]^+$ is optimal. Finally, the equivalent problem is $\inf_{a\ge0} E[a(X-x)+1]^+$, which equals $\bar{p}_x(X)$ when $x < \sup X$. □

3. Mathematical Properties of bPOE

bPOE is a function of real-valued threshold and random variable. bPOE properties for these two arguments are discussed in sections 3.1 and 3.2, respectively. Section 3.3 introduces and studies *upper* bPOE as a counterpart to lower bPOE, which was considered so far. The upper bPOE should be considered when probability of exceedance $P(X \geq x)$ is more suitable than $P(X > x)$. A special attention is given to the tightness of POE $\leq$ bPOE inequality in section 3.4. In particular, it is shown that bPOE is the tightest upper bound for POE, consistent with convex stochastic dominance.

3.1. Properties of bPOE w.r.t. Parameter x

This section studies bPOE behavior with respect to threshold. A comparison with cumulative distribution function (CDF) of a random variable is appropriate: probability of exceedance ($1-$CDF) is right-continuous, non-increasing, and bounded between 0 and 1. Being $1-$CDF of auxiliary variable $\bar{X} = \bar{q}(U; X)$, U is uniformly distributed on $[0,1]$, bPOE possesses the above properties, but also, as was pointed by [20] and proved in detail in proposition 3.1, superdistribution is continuous almost everywhere. Moreover, by proposition 3.3, reciprocal of bPOE is a convex function.

Proposition 3.1. *For $X \in \mathcal{L}^1(\Omega)$ the distribution $\bar{F}_X$ has at most one atom at* $\sup X$*, and the atom probability is* $P(\bar{X} = \sup X) = P(X = \sup X)$.

Proof. proposition Appendix A.1 and (2) imply that if superdistribution has an atom, then there are two possible locations: EX and $\sup X$. Since $\bar{F}_X$ is right-continuous, the continuity at x follows if $\lim_{y \uparrow x} \bar{F}_X(y) = \bar{F}(x)$. For $x = EX$, $\lim_{y \uparrow x} \bar{F}_X(y) = 0$, and since $\bar{q}_0(X) = EX$, then $\bar{F}_X(x) = 0$, hence, EX is a continuity point of $\bar{F}_X(x)$ and there is no atom at EX. For $x = \sup X$, $\bar{F}_X(x) = 1$, and by proposition Appendix A.1, $\lim_{y \uparrow x} \bar{F}_X(y) = 1 - P(X = \sup X)$, hence, the atom probability equals $\bar{F}_X(x) - \lim_{y \uparrow x} \bar{F}_X(y) = P(X = \sup X)$. □

Since bPOE equals one minus superdistribution, continuity of the latter is transferred to bPOE as well.

Corollary 3.2. *For $X \in \mathcal{L}^1(\Omega)$, buffered probability of exceedance $\bar{p}_x(X)$ is a non-increasing right-continuous function of x on $\mathbb{R}$ with no more than one point of discontinuity. Furthermore, $\bar{p}_x(X)$ is a continuous strictly decreasing function of x on the interval* $[EX, \sup X)$.

Proof. proposition 3.1 immediately implies the first part of this corollary. The second part follows from proposition Appendix A.1. □

Continuity of bPOE almost everywhere allows to present superdistribution as a mixture of continuous distribution and, perhaps, a constant (one-atomic discrete distribution), which is a more special case than general distribution. A more advanced property states that multiplicative inverse (reciprocal) of bPOE is not only continuous and increasing, as implied by corollary 3.2, but also convex. Moreover, there is piecewise-linearity for discrete distributions. Unlike corollary 2.3, here atoms of discrete distribution are not required to be ordered.

Proposition 3.3. *For $X \in \mathcal{L}^1(\Omega)$, the reciprocal of buffered probability of exceedance,*

$$\frac{1}{1-\bar{F}_X(x)} = \frac{1}{\bar{p}_x(X)},$$

is a convex function of x on $\mathbb{R}$, piecewise-linear for discretely distributed X.

Proof. Consider interval $EX < x < \sup X$, where formula (5) is valid. Then,

$$\frac{1}{\bar{p}_x(X)} = 1/\min_{c<x} \frac{E[X-c]^+}{x-c} = \max_{c<x} \frac{x-c}{E[X-c]^+}.$$

Note that since $\max_{c<x}(x-c)/E[X-c]^+ > 0$, then

$$\max_{c<x} \frac{x-c}{E[X-c]^+} = \max_{c<x} \frac{[x-c]^+}{E[X-c]^+} = \max_{c} \frac{[x-c]^+}{E[X-c]^+}.$$

The last expression $\max_c \{[x-c]^+/E[X-c]^+\}$ is convex in x as a maximum over the family of convex functions of x. $\bar{p}_x(X)$ is a continuous non-increasing function on $x \in (-\infty, \sup X)$, therefore, $1/\bar{p}_x(X)$ is a continuous non-decreasing function on $x \in (-\infty, \sup X)$. Then, extending the interval from $(EX, \sup X)$ to $(-\infty, \sup X)$ does not violate convexity of $1/\bar{p}_x(X)$, since $1/\bar{p}_x(X) = 1$, i.e., constant, for $x \in (-\infty, EX]$. Further extending of the interval from $(-\infty, \sup X)$ to $(-\infty, +\infty)$, i.e., $\mathbb{R}$, will not violate convexity either, since $1/\bar{p}_x(X) = +\infty$ for $x \geq \sup X$. That is, $1/\bar{p}_x(X)$ is a convex function of x.

Suppose that X is discretely distributed. Again, $1/\bar{p}_x(X) = 1$ for $x \in (-\infty, EX]$, and that is the first interval of linearity. Consider probability atom of random variable X at point x_* with probability p_*. Denote $\alpha^1 = F_X^-(x_*) = P(X < x_*)$, $\alpha^2 = F_X(x) = P(X \leq x_*) = \alpha^1 + p_*$ and $\bar{x}^i = \bar{q}_{\alpha^i}(X)$ for $i = 1, 2$. Then for $\bar{x}^1 < x < \bar{x}^2$ we have $x = \bar{q}_\alpha(X)$ with $\alpha \in (\alpha_1, \alpha_2)$, therefore, $q_\alpha(X) = x_*$. With corollary 2.2 we find that $1/\bar{p}_x(X) = (x - x_*)/E[X - x_*]^+$ for $\bar{x}^1 < x < \bar{x}^2$. Therefore, $1/\bar{p}_x(X)$ is linear on $\bar{x}^1 < x < \bar{x}^2$. This way, all the atom probability intervals of type $(F_X^-(x_*), F_X(x_*)) \subseteq [0, 1]$ will project into the intervals of type $(\bar{x}^1; \bar{x}^2) \subseteq (EX, \sup X)$ between corresponding superquantiles, covering all the interval $(EX, \sup X)$. Therefore, $1/\bar{p}_x(X)$ is a piecewise-linear function on $x \in (-\infty, \sup X)$, and $1/\bar{p}_x(X) = +\infty$ on $x \in [\sup X, +\infty)$. □

Note that the reciprocal of bPOE has a lot in common with partial moment function, $H_X(x) = E[X - x]^+$ (also called integrated survival function and excess function). First, both functions are convex and monotone (although former is increasing and latter is decreasing). Second, both functions can be used to formulate the second-order stochastic dominance relation[2]:

$$\begin{aligned} Y \leq_2 Z &\Leftrightarrow H_Y(x) \leq H_Z(x) \forall x \in \mathbb{R} \Leftrightarrow \bar{q}_\alpha(Y) \leq \bar{q}_\alpha(Z) \forall \alpha \in (0,1) \Leftrightarrow \\ &\Leftrightarrow \bar{p}_x(Y) \leq \bar{p}_x(Z) \forall x \in \mathbb{R} \Leftrightarrow 1/\bar{p}_x(Y) \geq 1/\bar{p}_x(Z) \forall x \in \mathbb{R}. \end{aligned}$$

In the case of discretely distributed X with a finite number of atoms, both convex functions are also picewise-linear, which allows to reformulate the dominance constraint $Y \leq_2 X$ with just a finite number of constraints on H, and reformulate $X \leq_2 Y$ with just a

[2]For details see [5, 20] or section 3.4. Also, third equivalence uses proposition 3.4.

finite number of constraints on $1/\bar{p}$. Interestingly, both functions are derived from the Lorenz function $L(\alpha) = \int_\alpha^1 q_p(X)dp$. Function H, as proved in [4], is a convex conjugate of $-L(\cdot + 1)$. Graph of $1/\bar{p}$ can be obtained from graph of L through composition of transformations: $(x, y) \to (x, y/(1-x))$ transforms, under constraint $x \in [0, 1)$, the hypograph of L into the hypograph of $\bar{q}$; $(x, y) \to (y, 1-x)$ transforms the constrained hypograph of $\bar{q}$ into the hypograph of $\bar{p}$, constrained by $y > 0$; $(x, y) \to (x, 1/y)$ transforms the constrained hypograph of $\bar{p}$ into the epigraph of $1/\bar{p}$; combining the mappings, $(x, y) \to (y/(1-x), 1/(1-x))$ transforms the hypograph of L, constrained by $x \in [0, 1)$, into the epigraph of $1/\bar{p}$.

3.2. Properties of bPOE w.r.t. Random Variable

Behavior of bPOE with respect to random variable is especially important for optimization. In particular, a lot of attention was paid to convex risk and uncertainty measures, see e.g. [7, 22]. Implications of convexity for optimization are well known, however, bPOE is not a convex function. As proved in proposition 3.4, bPOE is quasi-convex, that is, it has convex level-sets. For advantages of quasi-convex optimization in general, see, for example, paper [9] on interior methods and paper [10] on multiobjective optimization. Among other properties, frequently desired for risk measures, bPOE possesses monotonicity, see proposition 3.6. Instead of positive homogeneity, $f(\lambda X) = \lambda f(X)$ for $\lambda > 0$, bPOE satisfies $\bar{p}_{\lambda x}(\lambda X) = \bar{p}_x(X)$ for $\lambda > 0$. Instead of translation invariance, $f(X + c) = f(X) + c$ for $c \in \mathbb{R}$, bPOE satisfies $\bar{p}_{x+c}(X + c) = \bar{p}_x(X)$ for $c \in \mathbb{R}$. These two properties can be united into invariance under monotonic linear transformation: if h is monotonic linear function (that is, h is linear and increasing, $h(x) = \lambda x + c$ and $\lambda > 0$), then $\bar{p}_{h(x)}(h(X)) = \bar{p}_x(X)$. For constant variable $X \equiv c$, bPOE is the indicator function $\bar{p}_x(X) = I(c < x) \equiv \{1, \text{ if } c < x;\ 0, \text{ otherwise}\}$. Instead of aversion, $f(X) > EX$, bPOE satisfies $\bar{p}_x(X) < 1$ for $x < EX$.

Another type of convexity, convexity with respect to distribution mixtures (mixture convexity), is usually overlooked in risk measure literature. As [3] suggests, this type of convexity is important in such applications as schedule optimization, and many popular risk measures are mixture quasi-concave. It is proved by proposition 3.5 that bPOE is mixture concave.

Proposition 3.4. *Buffered probability is a closed quasi-convex function of random variable, i.e., for $x \in \mathbb{R}$ the set $\{X \in \mathcal{L}^1(\Omega) | \bar{p}_x(X) \le p\}$ is closed*[3] *convex for any $p \in \mathbb{R}$. Furthermore, for $p \in [0, 1)$,*

$$\bar{p}_x(X) \le p \Leftrightarrow \bar{q}_{1-p}(X) \le x. \tag{8}$$

Proof. Since $\bar{p}_x(X) = 0 \Leftrightarrow \bar{q}_1(X) \le x$ and $\bar{p}_x(X) = p \Leftrightarrow \bar{q}_{1-p}(X) = x$ for $p \in (0, 1)$, then, for $p \in [0, 1)$, (8) holds. Since superquantile is a closed convex function, then the set $\{X | \bar{q}_{1-p}(X) \le x\}$ is closed convex, and by (8), the set $\{X | \bar{p}_x(X) \le p\}$ is closed convex for $p \in [0, 1)$. For $p \ge 1$, the inequality $\bar{p}_x(X) \le p$ holds for any x and X, while $\{X | \bar{p}_x(X) \le p\} = \emptyset$ for $p < 0$. Combining the three intervals proves that $\{X | \bar{p}_x(X) \le p\}$ is closed convex for all $p \in \mathbb{R}$. □

[3] Closure here is understood in terms of convergence in $\mathcal{L}^1$, that is, a set S is closed if for any $\{X^i\}_{i=1}^\infty \subseteq S$ and $X \in \mathcal{L}^1(\Omega)$ such that $E|X^i - X| \to 0$, as $i \to \infty$, holds $X \in S$.

Example 1. *Buffered probability of exceedance is not a convex function of random variable, i.e. for r.v. X, Y, $x \in \mathbb{R}$, $\lambda \in (0,1)$, in general, $\bar{p}_x(\lambda X + (1-\lambda)Y) \not\leq \lambda \bar{p}_x(X) + (1-\lambda)\bar{p}_x(Y)$. For counterexample, take $x = 0$, $\lambda = 0.5$, $X \equiv 2$ and $Y \equiv -1$: $1 = \bar{p}_0(0.5) = \bar{p}_0((2-1)/2) > (\bar{p}_0(2) + \bar{p}_0(-1))/2 = (1+0)/2 = 0.5$.*

We call Z a mixture of random variables X and Y with coefficient $\lambda \in (0,1)$ and denote $Z = \lambda X \oplus (1-\lambda)Y$ if $F_Z(z) = F_{\lambda X \oplus (1-\lambda)Y}(z) \equiv \lambda F_X(z) + (1-\lambda)F_Y(z)$ for all $z \in \mathbb{R}$, where F_Z is a cumulative distribution function of the random variable Z. Taking a mixture of random variables corresponds to picking randomly between two (random) outcomes. Hence, it may be preferred for a risk-averse decision maker to consider objectives that are concave w.r.t. mixtures or at least quasi-concave. The study [3] shows that, in fact, most of the commonly used risk measures are mixture quasi-concave, or, as they call it, *randomization-proof*. A proof for mixture-concavity of superquantile is provided in proposition Appendix A.2. Below we show that buffered probability of exceedance is mixture-concave as well, hence, randomization-proof.

Proposition 3.5. *Buffered probability of exceedance is a concave function of random variable w.r.t. mixture operation, i.e.,*

$$\bar{p}_x(\lambda X \oplus (1-\lambda)Y) \geq \lambda \bar{p}_x(X) + (1-\lambda)\bar{p}_x(Y),$$

for $x \in \mathbb{R}$, $X, Y \in \mathcal{L}^1(\Omega)$, and $\lambda \in (0,1)$.

Proof. Without loss of generality, the concavity holds if proved for the following three cases: $x \neq \sup X$, $x \neq \sup Y$; $x = \sup X$, $x < \sup Y$; $x = \sup X$, $x > \sup Y$.

If $x \neq \sup X$ and $x \neq \sup Y$, then $x \neq \sup Z$, so the minimization formula $\min_{a \geq 0} E[a(\cdot - x)+1]^+$ is used to calculate bPOE. Since expectation $E[a(\cdot - x)+1]^+$ is linear with respect to mixture operation, then taking minimum over a collection of linear functions results in a concave function.

If $x = \sup X$ and $x < \sup Y$, then $\bar{p}_x(X) = 0$, $x < \sup Z = \sup Y$, and

$$\begin{aligned}
\bar{p}_x(Z) &= \min_{a \geq 0} E[a(\lambda X \oplus (1-\lambda)Y - x) + 1]^+ = \\
&= \min_{a \geq 0} \lambda E[a(X - x) + 1]^+ + (1-\lambda)E[a(Y - x) + 1]^+ \geq \\
&\geq \min_{a \geq 0} (1-\lambda)E[a(Y - x) + 1]^+ = \lambda \bar{p}_x(X) + (1-\lambda)\bar{p}_x(Y).
\end{aligned}$$

If $x = \sup X$ and $x > \sup Y$, then $x = \sup Z = \sup X$ and bPOE values for X, Y, Z are zero, hence, the concavity inequality holds. □

Distribution functions are linear w.r.t. mixture operation: $F_{\lambda X \oplus (1-\lambda)Y}(x) = \lambda F_X(x) + (1-\lambda)F_Y(x)$. Note that proposition 3.5 proves that superdistribution functions are convex w.r.t. mixture operation: $\bar{F}_{\lambda X \oplus (1-\lambda)Y}(x) \leq \lambda \bar{F}_X(x) + (1-\lambda)\bar{F}_Y(x)$.

Another important property for risk and uncertainty measures is monotonicity: greater outcomes imply greated function values. POE is monotone in random variable. The next proposition shows that bPOE is also a monotone function.

Proposition 3.6. *Buffered probability of exceedance is a monotonic function of random variable, i.e., for $x \in \mathbb{R}$, $Y, Z \in \mathcal{L}^1(\Omega)$, $\bar{p}_x(Y) \leq \bar{p}_x(Z)$ if $Y \leq Z$ almost surely.*

Proof. Since $Y \leq Z$ and $\bar{q}_\alpha(X)$ is monotonic, then $\bar{q}_{1-p}(Z) \leq x$ implies $\bar{q}_{1-p}(Y) \leq x$, which, by proposition 3.4, for $p \in [0,1)$, is equivalent to $\bar{p}_x(Z) \leq p$ implying $\bar{p}_x(Y) \leq p$. The latter implication holds automatically for $p < 0$ and $p \geq 1$, therefore, it holds for $p \in \mathbb{R}$, hence, $\bar{p}_x(Z) \geq \bar{p}_x(Y)$. □

3.3. Upper bPOE and Lower bPOE

The probability of exceedance can be defined in two ways, depending on whether to include or not the boundary threshold value. The lower POE is defined with a strict inequality $p_x^-(X) = P(X > x) = 1 - F_X(x)$ and the upper POE with a non-strict inequality $p_x^+(X) = P(X \geq x) = F_{-X}(-x)$. Note that terms "lower" and "upper" correspond to the inequality $p_x^-(X) \leq p_x^+(X)$.

Example 2. *Difference between upper and lower POE can be illustrated with random variable $X \equiv 0$, for which $P(X > 0) = 0$ and $P(X \geq 0) = 1$. Non-strict inequality as an indicator of exceedance can be preferred for some applications. Consider an example of binary classification problem. For every object $\mathbf{x}$, a class label $y \in \{-1, +1\}$ needs to be assigned, which is done by taking a sign of difference of corresponding scoring functions $S_+(\mathbf{x}) - S_-(\mathbf{x})$. Misclassification is then described by $-y(S_+(\mathbf{x}) - S_-(\mathbf{x})) \geq 0$. Note that minimizing misclassification losses, defined from strict inequality $-y(S_+(\mathbf{x}) - S_-(\mathbf{x})) > 0$, would result in trivial classifier $S_+(\mathbf{x}) = S_-(\mathbf{x}) \equiv 0$ being optimal. Such considerations led [14] to using of upper bPOE.*

Recall the definition of the auxiliary variable $\bar{X} = \bar{q}(U; X)$, where U is uniformly distributed on $[0, 1]$. With $\bar{X}$ the definition 1 of lower bPOE can be reformulated as $\bar{p}_x(X) = P\left(\bar{X} > x\right)$, which is similar to the definition of lower POE. We use notation, $\bar{p}_x(X) = \bar{p}_x^-(X)$ for the lower bPOE. Similarly, the *upper bPOE* is defined.

Definition 2. *For a random variable $X \in \mathcal{L}^1(\Omega)$ and $x \in \mathbb{R}$, the upper bPOE is defined as*

$$\bar{p}_x(X) = \begin{cases} 0, & \text{for } x < \sup X; \\ P(X = \sup X), & \text{for } x = \sup X; \\ 1 - \bar{q}^{-1}(x; X), & \text{for } EX < x < \sup X; \\ 1, & \text{otherwise.} \end{cases} \tag{9}$$

or equivalently

$$\bar{p}_x^+(X) = P\left(\bar{X} \geq x\right) = F_{-\bar{X}}(-x),$$

where for $x \in (EX, \sup X)$, $\bar{q}^{-1}(x; X)$ is the inverse of $\bar{q}(\alpha; X)$ as a function of α.

The following proposition provides the calculation formula for the upper bPOE, which is slightly simpler than formula (4) for the lower bPOE.

Proposition 3.7. *For a random variable $X \in \mathcal{L}^1(\Omega)$ and $x \in \mathbb{R}$, the upper buffered probability of exceedance equals*

$$\bar{p}_x^+(X) = \inf_{a \geq 0} E[a(X - x) + 1]^+. \tag{10}$$

In particular, $\bar{p}_x^+(X) = P(X = \sup X)$ for $x = \sup X$, otherwise, $\bar{p}_x^+(X) = \bar{p}_x^-(X)$.

Proof. Let us first prove the second part of the proposition. Proposition 3.1 implies that for $x \neq \sup X$ upper and lower bPOE must coincide, since $P\left(\bar{X} > x\right) = P\left(\bar{X} \geq x\right)$. Also, $P\left(\bar{X} \geq \sup X\right) = P\left(\bar{X} = \sup X\right) = P(X = \sup X)$.

For $x \neq \sup X$, since $\bar{p}_x^+(X) = \bar{p}_x^-(X)$, then $\bar{p}_x^+(X) = \bar{p}_x^-(X) = \min_{a \geq 0} E[a(X - x) + 1]^+ = \inf_{a \geq 0} E[a(X - x) + 1]^+$. Consider $x = \sup X$. Then, $E[a(X - \sup X) + 1]^+ \geq P(X = \sup X)$. Also, since $a(X - \sup X) + 1 > 0$ only for $X > \sup X - 1/a$, and

$a(X-\sup X)+1 \le 1$, then $E[a(X-\sup X)+1]^+ \le 1 \cdot P(X > \sup X - 1/a)$. For $a \to \infty$, $P(X > \sup X - 1/a) \to P(X = \sup X)$. Therefore,

$$\inf_{a\ge 0} E[a(X-\sup X)+1]^+ = \lim_{a\to\infty} E[a(X-\sup X)+1]^+ = P(X = \sup X) =$$
$$= \bar{p}^+_{\sup X}(X).$$

□

The formula (10) was used in [25] as a definitions of bPOE. The following corollary states mathematical properties for upper bPOE, similar to the properties specified in sections 3.1 and 3.2 for lower bPOE.

Corollary 3.8. *Properties of upper bPOE for $X \in \mathcal{L}^1(\Omega)$ and $x \in \mathbb{R}$:*

1. $\bar{p}^+_x(X) \ge p^+_x(X)$*;*
2. $\bar{p}^+_x(X)$ *is a non-increasing left-continuous function of x with no more than one point of discontinuity, and it is a strictly decreasing function of x on the interval $[EX, \sup X]$;*
3. *Multiplicative inverse of $\bar{p}^+_x(X)$ is a convex function of x and is piecewise-linear for discretely distributed X;*
4. $\bar{p}^+_x(X)$ *is a quasi-convex function of X;*
5. $\bar{p}^+_x(X)$ *is not lower semi-continuous, and $\bar{p}^+_x(X) \le p \not\Leftrightarrow \bar{q}_{1-p}(X) \le x$, but*
$$\bar{q}_{1-p}(X) < x \Rightarrow \bar{p}^+_x(X) \le p \Rightarrow \bar{q}_{1-p}(X) \le x.$$
6. $\bar{p}^+_x(X)$ *is concave with respect to mixture operation;*
7. $\bar{p}^+_x(X)$ *is monotonic with respect to random variable X.*

Proof. Let us prove the corollary item by item.

1. At $x = \sup X$, $\bar{p}^+_x(X) = p^+_x(X)$, otherwise $\bar{p}^+_x(X) = \bar{p}^-_x(X) \ge p^+_x(X)$, since $\bar{p}^-_x(X)$ is continuous w.r.t. x and inequality (1) holds.
2. Follows from corollaries 3.2 and 3.2.
3. Follows from proposition 3.3.
4. Note that
$$\bar{p}^+_x(X) = \lim_{\varepsilon\downarrow 0} \bar{p}^-_{x-\varepsilon}(X) = \inf_{\varepsilon>0} \bar{p}^-_{x-\varepsilon}(X).$$
Therefore,
$$\{X|\bar{p}^+_x(X) \le p\} = \bigcap_{\varepsilon>0}\{X|\bar{p}^-_{x-\varepsilon}(X) \le p\}.$$
Since sets $\{X|\bar{p}^-_{x-\varepsilon}(X) \le p\}$ are convex, then their intersection, the set $\{X|\bar{p}^+_x(X) \le p\}$, is also convex. Therefore, $\bar{p}^+_x(X)$ is a quasi-convex function of X.
5. Note that if X_μ is exponentially distributed with a mean value μ, then $\sup\{-X_\mu\} = 0$ and $\bar{p}^+_0(-X_\mu) = 0$. When $\mu \to 0$, $-X_\mu$ converge in $\mathcal{L}^1$ to $X_0 \equiv 0$. For $X_0 \equiv 0$, upper bPOE $\bar{p}^+_0(X_0) = 1$, which finishes the counterexample to lower-semicontinuity. For $X \equiv 0$, $x = 0$, $p = 0$, $\bar{q}_{1-p}(X) = x$, but $\bar{p}^+_x(X) > x$. The two-sided implication follows from $\bar{p}^-_x(X) \le \bar{p}^+_x(X) \le \bar{p}^-_{x-\varepsilon}(X)$ for all $\varepsilon > 0$.

6. For concavity with respect to mixture operation, note that

$$
\begin{aligned}
&\bar{p}_x^+(\lambda Y \oplus (1-\lambda)Z)) - \left(\lambda \bar{p}_x^+(Y) + (1-\lambda)\bar{p}_x^+(Z)\right) = \\
= \lim_{\varepsilon \downarrow 0} &\left[\bar{p}_{x-\varepsilon}^-(\lambda Y \oplus (1-\lambda)Z)) - \left(\lambda \bar{p}_{x-\varepsilon}^-(Y) + (1-\lambda)\bar{p}_{x-\varepsilon}^-(Z)\right)\right] \le 0,
\end{aligned}
$$

since a limit of non-positive values is non-positive.

7. If $Y \le Z$ a.s., then $\inf_{a\ge 0} E[a(Y-x)+1]^+ \le \inf_{a\ge 0} E[a(Z-x)+1]^+$.

□

3.4. Tightness of the POE $\le$ bPOE Inequality

The inequality (1) implies that buffered probability of exceedance is an upper bound for the probability of exceedance, but the open question is how tight the bPOE as a bound for the POE. The inequality itself is tight, that is, for each x exists X such that bPOE equals POE. Indeed, take $X \equiv x$, then $p_x(X) = \bar{p}_x(X) = 0$. This section studies the tightness of the inequality in the sense of finding the smallest upper bounds for the POE in certain classes of functions.

Definition 3. *For a function $f : \mathcal{X} \to \mathbb{R}$ and a class of functions $\mathcal{C}$, a function $g \in \mathcal{C} : \mathcal{X} \to \mathbb{R}$ is called a minimal upper bound (maximal lower bound) for f in $\mathcal{C}$ if*

1. *For f, g is an upper bound (lower bound), i.e. for all $x \in \mathcal{X}$ holds $f(x) \le g(x)$ ($f(x) \ge g(x)$).*
2. *For any $h \in \mathcal{C} : \mathcal{X} \to \mathbb{R}$ there exists $x \in \mathcal{X}$ such that $g(x) < h(x)$ ($g(x) > h(x)$).*

Furthermore, g is unique if for any $h \in \mathcal{C} : \mathcal{X} \to \mathbb{R}$ and for any $x \in \mathcal{X}$ holds $g(x) \le h(x)$ ($g(x) \ge h(x)$).

Certain class of functions and optimality within were discussed in [11], and used as an argument for CVaR vs. VaR tightness. A straightforward result was obtained in [8], showing that CVaR is the only minimal convex upper bound for VaR. When compared to convex functions, quasi-convex functions may better approximate bounded functions, such as POE. Therefore, it would make sense to study minimal quasi-convex upper bounds instead of convex ones. However, these results will appear as a corollary for the study of bounds in a wider class of *Schur-convex* functions.

To define Schur-convexity, we need the concept of the convex stochastic dominance. The random variable $X \in \mathcal{L}^1(\Omega)$ dominates $Y \in \mathcal{L}^1(\Omega)$ in the convex order if for any convex function $\phi : \mathbb{R} \to \mathbb{R}$ holds $E\phi(X) \ge E\phi(Y)$, provided that the expectations exist. We denote the convex dominance by $Y \le_{cx} X$. Generally, convex dominance implies that X is more "variable" than Y.

The convex dominance is closely related with the the second-order dominance, denoted by $Y \le_2 X$. The dominance can be used to describe preferences of risk-averse decision makers, see, e.g. [5]. The second-order and convex dominance are related as follows: $Y \le_{cx} X \Leftrightarrow EY = EX, Y \le_2 X$. As mentioned in section 3.1, the second-order dominance can be equivalently defined using inequalities for partial moment, $E[Y-z]^+ \le E[Z-z]^+$ for all $z \in \mathbb{R}$, or superquantile, $\bar{q}_\alpha(Y) \le \bar{q}_\alpha(X)$ for all $\alpha \in [0,1)$. For the review of stochastic dominance relations and their properties, see [24].

A function is called *law-invariant* if it takes equal values on identically distributed random variables, that is, $F_X(z) = F_Y(z)$ for all $z \in \mathbb{R}$ implies $f(X) = f(Y)$ for the law invariant f. A law-invariant function is called Schur-convex if it is consistent with the convex order, that is, $f : \mathcal{L}^1(\Omega) \to \mathbb{R}$ is Schur-convex if $f(Y) \le f(X)$ for any $Y \le_{cx} X$. The properties of Schur-convex functions have a more extensive representation in the Euclidean projection, where random variables are substituted with vectors, law-invariance with symmetry, and convex order with majorization, see [15].

Note that the consistency with the convex order is necessary for the consistency with the second order. With the convex and the second order dominances being closely related to the risk-aversion, for a risk-averse decision maker, it is reasonable to minimize a Schur-convex function. That is, the risk-averseness, in the sense of the convex stochastic dominance, contradicts using a non Schur-convex objective function. Therefore, upper bounds in the class of Schur-convex functions correspond to the largest class of functions appropriate for a risk-averse decision maker. It is easy to see from the definition of the Schur-convexity that for an arbitrary function f, a minimal Schur-convex upper bound on $\mathcal{L}^1(\Omega)$ is unique and it is given by

$$\bar{f}(X) \equiv \sup_{Y \le_{cx} X} f(Y),$$

for details see proposition Appendix B.1. A function is called *Schur-concave* if its negative is Schur-convex. It is easy to see that the unique maximal Schur-concave lower bound for a function f is given by $\inf_{Y \le_{cx} X} f(Y)$. The proposition below studies bounds for POE.

Proposition 3.9. *Let the probability space be non-atomic, then for p_x on $\mathcal{L}^1(\Omega)$, where $x \in \mathbb{R}$, $\bar{p}_x$ is the minimal Schur-convex upper bound, while the maximal Schur-concave lower bound is given by $1 - \bar{p}^+_{-x}(-X)$ for $X \in \mathcal{L}^1(\Omega)$.*

Proof. Since $\bar{p}_x(X)$ is closed quasi-convex and law-invariant, it is Schur-convex, see proposition Appendix B.2. To prove that the bound is minimal let us construct for any x and X a sequence $\{Y^i\}_{i=1}^{\infty}$ such that $Y^i \le_{cx} X$ and $\lim_{i\to\infty} p_x(Y^i) \ge \bar{p}_x(X)$.

Define Y_β, a random variable such that $\bar{q}_\alpha(Y_\beta) = \bar{q}_\alpha(X)$ for $\alpha \in [0, \beta]$ and $\bar{q}_\alpha(Y_\beta) = \bar{q}_\beta(X)$ for $\alpha \in [\beta, 1]$. Note that $EY_\beta = EX$, hence, $Y_\beta \le_{cx} X$, and also that $P(Y_\beta = \bar{q}_\beta(X)) = 1 - \beta$.

Now, to construct $\{Y^i\}_{i=1}^{\infty}$ consider one of the three possible cases. First, if $\sup X \le x$, then $p_x(X) = \bar{p}_x(X) = 0$, define $\{Y^i\}_{i=1}^{\infty}$ by $Y^i = X$. Second, if $EX > x$, take $Y \equiv EX$, then $Y \le_{cx} X$ and $p_x(Y) = \bar{p}_x(X) = 1$. Third, if $\sup X > x$, $EX \le x$, hence, $\bar{q}_\alpha(X) = x$ for some $\alpha \in [0, 1)$ and $\bar{p}_x(X) = 1 - \alpha$. Since $\bar{q}_\alpha(X) = x < \sup X$, then $\bar{q}_\beta(X) > \bar{q}_\alpha(X)$ for $\beta \in (\alpha, 1]$. For a sequence $\{\beta^i\}_{i=1}^{\infty}$ such that $\beta^i \downarrow \alpha$ and $\beta \in (\alpha, 1]$, define $\{Y^i\}_{i=1}^{\infty}$ by $Y^i = Y_{\beta^i}$. Then $1 - \beta^i = P(Y_{\beta^i} = \bar{q}_{\beta^i}(X)) \le P(Y_{\beta^i} > \bar{q}_\alpha(X)) = p_x(Y_{\beta^i})$, hence, $\lim_{i\to\infty} p_x(Y^i) \ge \lim_{i\to\infty} 1 - \beta_i = 1 - \alpha = \bar{p}_x(X)$.

The proof goes similarly for the second part of the proposition. Since $\bar{p}^+_x(X) = \inf_{a\ge 0} E[a(X - x) + 1]^+$ and $E[a(X - x) + 1]^+$ is Schur-convex for all $a, x \in \mathbb{R}$, the $1 - \bar{p}^+_{-x}(-X)$ is Schur-concave. It is therefore sufficient to provide a Y such that $Y \le_{cx} X$ and $p_x(Y) = 1 - \bar{p}^+_{-x}(-X)$.

Again, three cases are considered. First, if $\inf X > x$, then $p_x(X) = 1 - \bar{p}^+_{-x}(-X) = 1$. Second, if $EX \le x$, take $Y \equiv EX$, then $Y \le_{cx} X$ and $p_x(Y) = 1 - \bar{p}^+_{-x}(-X) = 0$. Third, if $\inf X \le x$, $EX > x$, hence, $\bar{q}_\alpha(-X) = -x$ for some $\alpha \in [0, 1]$ and $1 - \bar{p}^+_{-x}(-X) = \alpha$. For Y_α, corresponding to $-X$, $P(Y_\alpha = \bar{q}_\alpha(-X)) = 1 - \alpha$. Take $Y = -Y_\alpha$, then $Y \le_{cx} X$ is implied by $Y_\alpha \le_{cx} -X$, and $p_x(Y) = \alpha = 1 - \bar{p}^+_{-x}(-X)$ is implied by $\sup Y_\alpha = \bar{q}_\alpha(-X) = -x$. □

For the corollary below, recall $p_x^+(X) \equiv P(X \geq x)$ notation.

Corollary 3.10. *Let the probability space be non-atomic, then for p_x^+ on $\mathcal{L}^1(\Omega)$, where $x \in \mathbb{R}$, $\bar{p}_x^+$ is the minimal Schur-convex upper bound, while the maximal Schur-concave lower bound is given by $1 - \bar{p}_{-x}(-X)$ for $X \in \mathcal{L}^1(\Omega)$.*

Proof. Equation

$$\sup_{Y \leq_{cx} X} P(Y \geq x) = 1 - \inf_{Y \leq_{cx} X} P(-Y > -x) = 1 - \inf_{-Y \leq_{cx} -X} P(-Y > -x) =$$
$$= 1 - (1 - \bar{p}^+_{-(-x)}(-(-X)) = \bar{p}_x^+(X),$$

proves the first part of the proposition, while equation

$$\inf_{Y \leq_{cx} X} P(Y \geq x) = 1 - \sup_{Y \leq_{cx} X} P(-Y > -x) = 1 - \sup_{-Y \leq_{cx} -X} P(-Y > -x) =$$
$$= 1 - \bar{p}_{-x}(-X),$$

proves the second part of the proposition. □

A generalized result of CVaR being the unique minimal Schur-convex upper bound of VaR is presented in proposition Appendix B.3. Since bPOE is law-invariant and quasi-convex, and the whole class of law-invariant quasi-convex functions belongs to the class of Schur-convex functions, see proposition Appendix B.2, then bPOE is a unique minimal quasi-convex upper bound for POE.

Corollary 3.11. *For the non-atomic probability space, the class of closed quasi-convex law-invariant functions on $\mathcal{L}^1(\Omega)$, and $x \in \mathbb{R}$, buffered probability of exceedance $\bar{p}_x$ is the unique minimal upper bound for for probability of exceedance p_x.*

Note that for the class of quasiconvex functions obtaining maximal lower bound is straightforward, and such bound is unique. For a function f on $\mathcal{L}^1(\Omega)$, consider its level sets $L_c = \{X \in \mathcal{L}^1(\Omega) | f(X) \leq c\}$, and take convex hulls of these sets. Let g be a function for which the corresponding level sets coincide with $conv(L_c)$, then $g(X) = \sup\{c | X \in conv(L_c)\}$. However, such bound may not be unique, as demonstrated in the following example.

Example 3. *Consider function $f(x) = \cos(x)$, its quasi-convex upper bounds are $g_k(x) = \{\cos(x),\ \text{for } x \in [2\pi k, 2\pi(k+1)]; 1,\ \text{otherwise}\}$, where $k \in \mathbb{Z}$, see fig. 1.*

In the case of POE, uniqueness of minimal quasi-convex upper bound is due to the fact that bPOE, unique minimal upper Schur-convex bound, happens to be quasi-convex.

4. Optimization Problems with bPOE

This section provides some results on bPOE minimization. In some cases, bPOE can be very efficiently minimized. There are two groups of results. Section 4.1 considers the parametric family of bPOE minimization formulations without any assumptions about feasible region. We proved that the bPOE minimization and the superquantile minimization share the same sets of optimal solutions and parameter-objective pairs, with minor exceptions. This means, in particular, that with a superquantile minimization solver it is possible to minimize bPOE with several solver calls, see algorithm 1. Section 4.2 provides convex and LP equivalent reformulations of bPOE minimization with a convex feasible region.

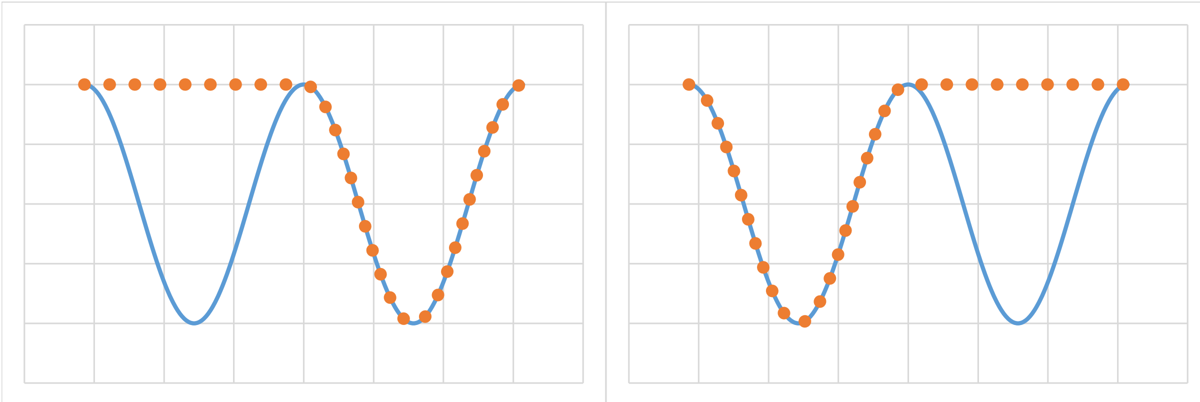

Figure 1: Function $\cos(x)$ is presented with solid lines; minimal upper bounds among quasi-convex functions are presented with dashed lines.

4.1. Two Families of Optimization Problems

This section considers two parametric families of optimization problems, for buffered probability of exceedance and superquantile, over an arbitrary feasible region $\mathcal{X}$. An indirect equivalence between the two families is shown. Propositions 4.1 and 4.2 and corollary 4.3 prove that, with minor exceptions, an optimal solution for bPOE minimization is also optimal for superquantile optimization at some parameter value, and vice versa. Each of the two families provides the optimal objective as a function of parameter value. Proposition 4.4 shows that graphs of these functions, if aligned properly, have a significant intersection, which coincides with Pareto frontier of a certain "feasible" set of pairs, induced by the problem's feasible set $\mathcal{X}$. The results of this section rely heavily on the level set equivalence between $\bar{p}_x$ and $\bar{q}_\alpha$, (8).

Program $\mathcal{P}(x)$, for $x \in \mathbb{R}$,

$$\mathcal{P}(x): \qquad \begin{array}{rl} \min & \bar{p}_x(X) \\ s.t. & X \in \mathcal{X}. \end{array}$$

Program $\mathcal{Q}(\alpha)$, for $\alpha \in [0,1]$,

$$\mathcal{Q}(\alpha): \qquad \begin{array}{rl} \min & \bar{q}_\alpha(X) \\ s.t. & X \in \mathcal{X}. \end{array}$$

Here and further it is assumed $\mathcal{X} \subseteq \mathcal{L}^1(\Omega)$. For a set of random variables $\mathcal{X}$, define

$$e_\mathcal{X} = \inf_{X \in \mathcal{X}} EX, \qquad s_\mathcal{X} = \inf_{X \in \mathcal{X}} \sup X.$$

Proposition 4.1. *Let $X_0 \in \mathcal{X} \subseteq \mathcal{L}^1(\Omega)$ be an optimal solution to $\mathcal{P}(x_0)$, $x_0 \in \mathbb{R}$. Then, X_0 is an optimal solution to $\mathcal{Q}(1 - \bar{p}_{x_0}(X_0))$ for $e_\mathcal{X} < x_0 \le s_\mathcal{X}$, while for $x_0 \notin (e_\mathcal{X}, s_\mathcal{X}]$, in general, it is not.*

Proof. Let $p^* = \bar{p}_{x_0}(X_0)$. Since $x_0 > e_\mathcal{X}$, then $\bar{p}_{x_0}(X_0) < 1$. If $\bar{p}_{x_0}(X_0) = 0$, then $\sup X_0 \le x_0$, but $x_0 \le s_\mathcal{X}$, and $\sup X_0 \ge s_\mathcal{X}$ by definition, therefore, $\sup X_0 = x_0 = s_\mathcal{X}$. If $0 < \bar{p}_{x_0}(X_0) < 1$, then, by definition 1 of bPOE, $\bar{q}_{1-p^*}(X_0) = x_0$.

Suppose that X_0 is not an optimal solution to $\mathcal{Q}(1 - \bar{p}_{x_0}(X_0))$, then there exists $X^* \in \mathcal{X}$ such that $\bar{q}_{1-p^*}(X^*) < x_0$. Since $x_0 \leq s_{\mathcal{X}}$, then $p^* > 0$ and $\sup X^* \geq x_0$. Therefore, there exists $p < p^*$ such that $\bar{q}_{1-p}(X^*) = x_0$, since $\bar{q}_{1-p}(X)$ is a continuous non-increasing function of p. There are two possible cases. First, if $\sup X^* = x_0$, then $\bar{p}_{x_0}(X^*) = 0 < p^*$, X_0 is not an optimal solution to $\mathcal{P}(x_0)$, leading to a contradiction. Second, if $\sup X^* > x_0$, then $\bar{p}_{x_0}(X^*) = p < p^*$, X_0 is not an optimal solution to $\mathcal{P}(x_0)$, leading to a contradiction.

For $x_0 \leq e_{\mathcal{X}}$ the optimal value for $\mathcal{P}(x_0)$ is 1, therefore, any feasible solution is an optimal solution. For $x > s_{\mathcal{X}}$, the optimal value for $\mathcal{P}(x_0)$ is 0. If $s_{\mathcal{X}} < \sup X_0 \leq x_0$, then $\bar{p}_{x_0}(X_0) = 0$, and it is optimal for $\mathcal{P}(x_0)$, but $\bar{q}_1(X_0) > s_{\mathcal{X}}$, and it is not optimal for $\mathcal{Q}(1)$. $\square$

Proposition 4.2. *Let $X_0 \in \mathcal{X} \subseteq \mathcal{L}^1(\Omega)$ be an optimal solution to $\mathcal{Q}(\alpha_0)$, $\alpha_0 \in [0,1]$. Then X_0 is an optimal solution to $\mathcal{P}(\bar{q}_{\alpha_0}(X_0))$, unless $\sup X_0 > \bar{q}_{\alpha_0}(X_0)$ and there exists $X^* \in \mathcal{X}$ such that*

1. $\sup X^* = \bar{q}_{\alpha_0}(X_0)$,
2. $P(X^* = \sup X^*) \geq 1 - \alpha_0$.

Proof. Denote $x_0 = \bar{q}_{\alpha_0}(X_0)$. First, suppose $\sup X_0 = x_0$. Then $\bar{p}_{x_0}(X_0) = 0$, and X_0 is an optimal solution to $\mathcal{P}(x_0)$.

Second, suppose that $\sup X_0 > x_0$ and that exists $X^* \in \mathcal{X}$ such that $\bar{p}_{x_0}(X^*) < \bar{p}_{x_0}(X_0)$. Since $x_0 = \bar{q}_{\alpha_0}(X_0)$ and $\sup X_0 > x_0$, then $\bar{p}_{x_0}(X_0) = 1 - \alpha_0$.

Suppose $\sup X^* > x_0$, then $\bar{q}_{\alpha}(X^*)$ is strictly increasing on $[0, 1 - \bar{p}_{x_0}(X^*)]$. Therefore, $\bar{q}_{\alpha_0}(X^*) < \bar{q}_{1-\bar{p}_{x_0}(X^*)}(X^*) = x_0$, which implies that X_0 is not an optimal solution to $\mathcal{Q}(\alpha_0)$, which is a contradiction. Consequently, $\sup X^* = x_0$.

Suppose $P(X^* = x_0) < 1 - \alpha_0$. Then, $\bar{q}_{\alpha}(X^*)$ is strictly increasing on $[0, 1 - P(X^* = x_0)]$, and $\bar{q}_{\alpha_0}(X^*) < x_0$. Therefore, X_0 is not an optimal solution to $\mathcal{Q}(\alpha_0)$, which is a contradiction. Therefore, $P(X^* = x_0) \geq 1 - \alpha_0$. $\square$

Intuition behind proposition 4.2 is as follows. Note that X^* is also an optimal solution to $\mathcal{Q}(\alpha_0)$. Therefore, we have two optimal solutions for the right tail expectation minimization problem. The difference between optimal solutions X^* and X_0 is that X^* is constant in its right $1 - \alpha_0$ tail, and X_0 is not, since $\bar{q}_{\alpha_0}(X_0) < \sup X_0$. Proposition 4.2 implies that X^* is an optimal solution to $\mathcal{P}(\bar{q}_{\alpha_0}(X_0))$, while X_0 is not. This is a very natural risk-averse decision. This implies that, for certain problems, formulations of type $\mathcal{P}(x)$ provide more reasonable solutions than formulations of type $\mathcal{Q}(\alpha)$.

Corollary 4.3. *Let $\mathcal{X} \subseteq \mathcal{L}^1(\Omega)$ be a set of random variables, such that $\sup X = \infty$ for all $X \in \mathcal{X}$. Then, program families $\mathcal{P}(x)$, for $x > e_{\mathcal{X}}$, and $\mathcal{Q}(\alpha)$, for $0 < \alpha < 1$, have the same set of optimal solutions. That is, if X_0 is optimal for $\mathcal{P}(x_0)$, then X_0 is optimal for $\mathcal{Q}(1 - \bar{p}_{x_0}(X_0))$. Conversely, if X_0 is optimal for $\mathcal{Q}(\alpha_0)$, then X_0 is optimal for $\mathcal{P}(\bar{q}_{\alpha_0}(X_0))$.*

Proof. proposition 4.1 implies that if $e_{\mathcal{X}} < x_0 \leq s_{\mathcal{X}} = \infty$, then if X_0 is optimal for $\mathcal{P}(x_0)$, then X_0 is optimal for $\mathcal{Q}(1 - \bar{p}_{x_0}(X_0))$. Note that since $e_{\mathcal{X}} < x_0 < s_{\mathcal{X}}$, then $\bar{p}_{x_0}(X_0) \in (0,1)$.

proposition 4.2 implies that if X_0 is optimal for $\mathcal{Q}(\alpha_0)$, then X_0 is optimal for $\mathcal{P}(\bar{q}_{\alpha_0}(X_0))$, unless exists $X^* \in \mathcal{X}$ such that $\sup X^* = \bar{q}_{\alpha_0}(X_0)$. Which is impossible since $\sup X^* = \infty > \bar{q}_{\alpha_0}(X_0)$. Note that since $\alpha_0 \in (0,1)$, then $e_{\mathcal{X}} < \bar{q}_{\alpha_0}(X_0) < \infty$. $\square$

Assumption $\sup X = +\infty$ for all $X \in \mathcal{X}$ in corollary 4.3 might be too strong for some practical problems, where it is a common practice for all random variables to be defined on a finite probability space generated by system observations.

Optimization problem families $\mathcal{P}(x)$ and $\mathcal{Q}(\alpha)$ map parameter values from $\mathbb{R}$ to $[0,1]$ and from $[0,1]$ to $\mathbb{R}$ correspondingly:

$$f_{\mathcal{P}}(x) = \min_{X\in\mathcal{X}} \bar{p}_x(X), \quad f_{\mathcal{Q}}(\alpha) = \min_{X\in\mathcal{X}} \bar{q}_\alpha(X).$$

Graphs of these functions can be aligned into the space of pairs $(x,\alpha) \in \mathbb{R}\times[0,1]$ to produce the following sets:

$$S_{\mathcal{P}} = \{(x,\alpha) | f_{\mathcal{P}}(x) = 1-\alpha\}, \quad S_{\mathcal{Q}} = \{(x,\alpha)|f_{\mathcal{Q}}(\alpha) = x\}.$$

These sets of pairs have significant intersection, which, as will be proved in proposition 4.4, coincides with the defined below reductions of sets $S_{\mathcal{P}}$ and $S_{\mathcal{Q}}$,

$$S_{\mathcal{P}}^- = \{(x,\alpha) \in S_{\mathcal{P}} | e_{\mathcal{X}} \le x \le s_{\mathcal{X}}\}, \quad S_{\mathcal{Q}}^- = \{(x,\alpha) \in S_{\mathcal{Q}} | x < s_{\mathcal{X}}\} \cup \{(s_{\mathcal{X}}, 1)\}.$$

Alternatively, the feasible set $\mathcal{X}$ can be used to define a "feasible" set of pairs in $\mathbb{R}\times[0,1]$. For any random variable $X \in \mathcal{X}$ there is a set $S_X = \{(x,\alpha)|\bar{q}_\alpha(X) = x\}$. Let us define union of such sets for $X \in \mathcal{X}$ as

$$S_{\mathcal{X}} = \bigcup_{X\in\mathcal{X}} S_X = \{(x,\alpha)| \text{ exists } X : \bar{q}_\alpha(X) = x\}.$$

Naturally, since feasible solutions represent loss distributions, preferred random variables have superquantile as small as possible for a fixed confidence level, or have confidence level as big as possible for a fixed superquantile value. Therefore, for the set $S_{\mathcal{X}}$ we define a Pareto frontier, which is often called an efficient frontier in finance, as follows:

$$S_{\mathcal{X}}^- = \{(x,\alpha) \in S_{\mathcal{X}} | \{(x',\alpha') \in S_{\mathcal{X}} | x' \le x, \alpha' \ge \alpha, (x',\alpha') \ne (x,\alpha)\} = \emptyset\}.$$

From the perspective of minimizing superquantile and bPOE, only pairs from $S_{\mathcal{X}}^-$ are of interest for decision making. The proposition below provides connection between the sets defined.

Proposition 4.4.

$$S_{\mathcal{P}} \cap S_{\mathcal{Q}} = S_{\mathcal{P}}^- = S_{\mathcal{Q}}^- = S_{\mathcal{X}}^-.$$

Proof. Let us start with $S_{\mathcal{P}} \cap S_{\mathcal{Q}} = S_{\mathcal{X}}^-$. Notice that

$$(x,\alpha) \in S_{\mathcal{P}} \Leftrightarrow (x,\alpha') \in S_{\mathcal{X}} \to \alpha' \le \alpha, \tag{11}$$
$$(x,\alpha) \in S_{\mathcal{Q}} \Leftrightarrow (x',\alpha) \in S_{\mathcal{X}} \to x' \ge x. \tag{12}$$

Clearly, right sides of (11),(12) hold for $S_{\mathcal{X}}^-$, which implies $S_{\mathcal{X}}^- \subseteq S_{\mathcal{P}} \cap S_{\mathcal{Q}}$. Suppose $S_{\mathcal{X}}^- \subset S_{\mathcal{P}} \cap S_{\mathcal{Q}}$, i.e., for some $(x,\alpha) \in S_{\mathcal{P}} \cap S_{\mathcal{Q}}$ there exists $(x',\alpha') \in S_{\mathcal{X}}$ such that $x' \le x$, $\alpha' \ge \alpha$ and $(x',\alpha') \ne (x,\alpha)$. Notice that if $(x,\alpha) \in S_{\mathcal{P}} \cap S_{\mathcal{Q}}$, then $(x,\alpha') \in S_{\mathcal{X}} \to \alpha' \le \alpha$ and $(x',\alpha) \in S_{\mathcal{X}} \to x' \ge x$. Then, $x' < x$ and $\alpha' > \alpha$. Consider random variable X^* which has generated point (x',α'). Since $\bar{q}_{X^*}(\alpha') < x$, then $\bar{q}_{X^*}(\alpha) < x$. Therefore, there exists $(\bar{q}_{X^*}(\alpha), \alpha) \in S_{\mathcal{X}}$ with $\bar{q}_{X^*}(\alpha) < x$, while $(x, \alpha) \in S_{\mathcal{Q}}$ and (12) holds, leading to a contradiction.

Let us prove $S^-_{\mathcal{P}} = S^-_{\mathcal{Q}}$. Suppose $(x, \alpha) \in S^-_{\mathcal{P}}$ and $x \neq e_{\mathcal{X}}$, then we can use proposition 4.1 to conclude that $(x, \alpha) \in S^-_{\mathcal{Q}}$. If $x = e_{\mathcal{X}}$, then $\bar{p}_x(X) = 0$, therefore, $(x, \alpha) = (e_{\mathcal{X}}, 0) \in S^-_{\mathcal{Q}}$. Let $(x, \alpha) \in S^-_{\mathcal{Q}}$. If $x < s_{\mathcal{X}}$, then there is no X^* such that $\sup X^* = x$, therefore, we can use proposition 4.2 and conclude that $(x, \alpha) \in S^-_{\mathcal{P}}$.

Finally, let us prove $S_{\mathcal{P}} \cap S_{\mathcal{Q}} = S^-_{\mathcal{P}}$. Since $S^-_{\mathcal{P}} \subseteq S_{\mathcal{P}}$, $S^-_{\mathcal{Q}} \subseteq S_{\mathcal{Q}}$ and $S^-_{\mathcal{P}} = S^-_{\mathcal{Q}}$, then $S^-_{\mathcal{P}} \subseteq S_{\mathcal{P}} \cap S_{\mathcal{Q}}$. Suppose $(x, \alpha) \in S_{\mathcal{P}} \cap S_{\mathcal{Q}}$. If $x < s_{\mathcal{X}}$, then $(x, \alpha) \in S^-_{\mathcal{P}}$. If $x = s_{\mathcal{X}}$, then $\alpha = 1$, because there is only one α for any x in $S_{\mathcal{P}}$. Then $(x, \alpha) = (s_{\mathcal{X}}, 1) \in S^-_{\mathcal{P}}$. Point with $x > s_{\mathcal{X}}$ cannot be in $S_{\mathcal{Q}}$ since $f_{\mathcal{Q}}(\alpha) = \min_{X \in \mathcal{X}} \bar{q}_\alpha(X) \leq \min_{X \in \mathcal{X}} \sup X = s_{\mathcal{X}}$. Therefore, $S_{\mathcal{P}} \cap S_{\mathcal{Q}} \subseteq S^-_{\mathcal{P}}$, which finalizes the proof. □

Let us show that proposition 4.2 allows for an algorithm that solves $\mathcal{P}(x)$ via multiple calls to a solver of $\mathcal{Q}(\alpha)$. Algorithm 1 takes in functions $q_{\mathcal{X}}(\alpha) = \min_{X \in \mathcal{X}} \bar{q}_\alpha(X)$ and $X^*(\alpha) = \arg\min_{X \in \mathcal{X}} \bar{q}_\alpha(X)$ as arguments.

Algorithm 1 Solve $\mathcal{P}(x)$ with a solver of $\mathcal{Q}(\alpha)$

Require: $\mathcal{X} \subseteq \mathcal{L}^1(\Omega)$, $x \in \mathbb{R}$, $\varepsilon > 0$, $q_{\mathcal{X}} \colon [0,1] \to \mathbb{R}$, $X^* \colon [0,1] \to \mathcal{X}$
Ensure: $(\min_{X \in \mathcal{X}} \bar{p}_x(X),\ \arg\min_{X \in \mathcal{X}} \bar{p}_x(X))$
  if $x \geq q_{\mathcal{X}}(1)$ **then**
    return $(0,\ X^*(1))$
  else if $x \leq q_{\mathcal{X}}(0)$ **then**
    return $(1,\ \mathcal{X})$
  end if
  $p :=$ BinarySearch$([0,\ 1],\ q_{\mathcal{X}} - x,\ \varepsilon)$ {BinarySearch$([a, b],\ f,\ \delta)$ finds a root of function f on interval $[a, b]$ with precision δ via binary search.}
  return $(p,\ X^*(p))$

The correctness of algorithm 1 is as follows. The $q_{\mathcal{X}}(\alpha)$ is a continuous function of α since minimum in $\mathcal{Q}(\alpha)$ is assumed to be attained and $\bar{q}_\alpha(X)$ is continuous for any $X \in \mathcal{X}$. Therefore, if $q_{\mathcal{X}}(0) < x < q_{\mathcal{X}}(1)$, then there exists $\alpha^* \in (0, 1)$ such that $q_{\mathcal{X}}(\alpha^*) = x$. This α^* is found in the algorithm via binary search. Since $q_{\mathcal{X}}(\alpha^*) = x$ and $x < q_{\mathcal{X}}(1)$, then the exception part of proposition 4.2 is not valid, and the optimal solution to $\mathcal{Q}(\alpha^*)$ is also optimal to $\mathcal{P}(x)$. If $x \geq q_{\mathcal{X}}(1)$, then $\bar{p}_x(X^*) = 0$ for any solution that is optimal to $\mathcal{Q}(1)$. If $x \leq q_{\mathcal{X}}(0) \neq q_{\mathcal{X}}(1)$, then $\bar{p}_x(X) = 1$ for all $X \in \mathcal{X}$.

4.2. Convex Reformulations of bPOE Minimization

This section considers two major types of the bPOE minimization formulations. The first formulation, $\min_{X \in \mathcal{X}} \bar{p}_x(X)$, where X is a random variable and $\mathcal{X} \subseteq \mathcal{L}^1(\Omega)$ is a convex feasible set of random variables. The second formulation, $\min_{\mathbf{w} \in W} \bar{p}_x(f(\mathbf{w}; \mathbf{X}))$, where $\mathbf{X}$ is a random vector with a given distribution, $\mathbf{w} \in \mathbb{R}^n$ is a vector of control variables, $W \subseteq \mathbb{R}^n$ is a convex feasible set for the control variables, and $f(\mathbf{w}; \mathbf{X})$ is a convex function of $\mathbf{w}$, while $f(\mathbf{w}; \mathbf{X}) \in \mathcal{L}^1(\Omega)$ for each $\mathbf{w} \in W$. The bPOE minimization for the case of positive homogeneous function f was first considered in [14].

In this section, for the sake of simplicity of the obtained reformulations, we use notation $\bar{p} := \bar{p}^+$. That is, along the section, whenever we say bPOE we refer to upper bPOE. Since $\bar{p}_x^+(X) \neq \bar{p}_x^-(X)$ only if $x = \sup X$, then, when compared to the minimization of upper bPOE, the minimization of lower bPOE might provide a different optimal objective value (equal to 0) and different optimal solution only if $x = \inf_{X\in\mathcal{X}} \sup X$ or $x = \inf_{\mathbf{w}\in W} \sup f(\mathbf{w};\mathbf{X})$. The likelihood of such coincidence should be rather small for most practical problems.

We start with the formulation $\min_{X\in\mathcal{X}} \bar{p}_x(X)$, assuming that $X \in \mathcal{X} \subseteq \mathcal{L}^1(\Omega)$. For a set $\mathcal{Z}$, denote its conical hull (the smallest pointed cone containing $\mathcal{Z}$) by $cone(\mathcal{Z})$, and denote its closure in $\mathcal{L}^1$ by $cl(\mathcal{Z})$. The proposition below provides a convex reformulation for the considered problem.

Proposition 4.5. *Let $\mathcal{X} \subseteq \mathcal{L}^1(\Omega)$ be a convex set of random variables. Then, for $x \in \mathbb{R}$,*

$$\inf_{X\in\mathcal{X}} \bar{p}_x(X) = \inf_{Y\in\mathcal{Y}} E[Y+1]^+,$$

where $\mathcal{Y} = cl\ cone(\mathcal{X} - x)$ is a closed convex cone.

Proof. proposition 3.7 implies that $\inf_{X\in\mathcal{X}} \bar{p}_x(X) = \inf_{X\in\mathcal{X}, a\geq 0} E[a(X-x)+1]^+$. Denote $Y = a(X-x)$, then $Y \in \cup_{a>0} a(\mathcal{X}-x) \cup \{0\}$. Since $\mathcal{X}$ is convex, then constraints $X \in \mathcal{X}$, $a \geq 0$ are equivalent to $Y \in cone(\mathcal{X}-x)$. Suppose that the sequence $\{Y^i\}_{i=1}^\infty \subset cone(\mathcal{X} - x)$ converges in $\mathcal{L}^1$ to Y, then $\lim_{i\to\infty} E[Y_i+1]^+ = E[Y+1]^+$. Hence, the feasible region can be extended to L_1-closure of $cone(\mathcal{X}-x)$, so finally, $\inf_{X\in\mathcal{X}} \bar{p}_x(X) = \inf_{Y\in\mathcal{Y}} E[Y+1]^+$. □

Note that proposition 4.5 holds for non-closed $\mathcal{Y} = cone(\mathcal{X} - x)$ as well, but, among specific cases of the problem, there is a better chance that inf can be substituted by min for the formulation with the closed feasible region. Note also that, due to change of variables, $Y = a(X - x)$, optimal solutions $\mathcal{X}^*$ to the original problem are found from optimal soluitons $\mathcal{Y}^*$ of the new problem as $\mathcal{X} \cap (cone(\mathcal{Y}^*) + x)$. If $0 \in \mathcal{Y}^*$, then optimal bPOE is 1, hence, every feasible solution is optimal, $\mathcal{X}^* = \mathcal{X}$.

To make proposition 4.5 practical, the abstract set of random variables $\mathcal{X}$ need to be parameterized with a finite-dimension set of decision variables. Let us consider a special case allowing for such parameterization. Suppose that the probability space is finite: $\Omega = \{\omega_1, \ldots, \omega_m\}$, with probabilities $P(\omega_i) = p_i$, $\mathbf{p} := (p_1, \ldots, p_m)$. Denote $\Pi^m := \{\mathbf{q} = (q_1, \ldots, q_m)^T \in \mathbb{R}^m | q_i \geq 0, i = 1, \ldots, m; \sum_{i=1}^m q_i = 1\}$ and $\Pi^m_+ := \{\mathbf{q} = (q_1, \ldots, q_m)^T \in \mathbb{R}^m | q_i > 0, i = 1, \ldots, m; \sum_{i=1}^m q_i = 1\}$. Assume that $\mathbf{p} \in \Pi^m_+$, i.e., it is indeed a vector of probabilities, and there are no degenerate events in the probability space, i.e., $P(\omega_i) = p_i > 0$. Then each random variable X from the feasible set is discretely distributed over m atoms and takes value x_i with probability p_i for $i = 1, \ldots, m$. Therefore, X can be parameterized with a vector of its values: $\mathbf{x} = (x_1, \ldots, x_m)$, where $x_i = X(\omega_i)$. Define

$$S = \bigcup_{X\in\mathcal{X}} (X(\omega_1), \ldots, X(\omega_m)). \tag{13}$$

Note that convexity of $\mathcal{X}$ implies convexity of $S \subseteq \mathbb{R}^m$, and choosing $X \in \mathcal{X}$ is equivalent to choosing $\mathbf{x} \in S$. Transformation $\mathcal{Y} = cl\ cone(\mathcal{X} - x)$ is therefore replaced with $C = cl\ cone(S - x\mathbf{e}_m)$, where $\mathbf{e}_m = (1, \ldots, 1) \in \mathbb{R}^m$ is the vector of ones, and $E[Y + 1]^+$ corresponds to $\mathbf{p}^T[\mathbf{y} + \mathbf{e}_m]^+$. Finally,

$$\inf_{X\in\mathcal{X}} \bar{p}_x(X) = \inf_{\mathbf{y}\in C} \mathbf{p}^T[\mathbf{y} + \mathbf{e}_m]^+, \tag{14}$$

where $C = cl\ cone(S - x\mathbf{e}_m)$ is a closed convex cone, and S is defined by (13). Consider a particular example of $\mathcal{X}$.

Example 4. *For a finite probability space $\Omega = \{\omega_1, \dots, \omega_m\}$, and a finite set of random variables $\{Q^j\}_{j=1}^k$, suppose that the feasible set is $\mathcal{X} = \{X | EXQ^j \le b^j,\ for\ j = 1, \dots, k\}$. Since $EXQ^j = \sum_{i=1}^m p_i X(\omega_i) Q^j(\omega^i)$, where $p_i = P(\omega_i)$, then, defining matrix A with $A_{ji} = p_i Q^j(\omega^i)$ and $\mathbf{b} = (b_1, \dots, b_k)$, we get, according to (13), $S = \{\mathbf{x} | A\mathbf{x} \le \mathbf{b}\}$.*

In general, as (14) suggests, bPOE minimization in the case of finite probability space is reduced to convex programming. In some cases, like the one considered in the example above, bPOE minimization is reduced to LP, as showed in the proposition below.

Corollary 4.6. *Let $\mathcal{X}$ be a feasible set of random variables defined on a finite probability space with vector of probabilities $\mathbf{p} \in \Pi_+^m$, $x \in \mathbb{R}$, and let set S be defined by (13). If $S = \{\mathbf{x} | A\mathbf{x} \le \mathbf{b}\} \subseteq \mathbb{R}^m$, then bPOE minimization is reformulated as an LP:*

$$\begin{aligned} \inf_{X \in \mathcal{X}} \quad \bar{p}_x(X) \quad = \quad \inf \quad & \mathbf{p}^T \mathbf{z} \\ s.t. \quad & \mathbf{z} \ge \mathbf{y} + \mathbf{e}_m, \\ & A\mathbf{y} - a(\mathbf{b} - xA\mathbf{e}_m) \le 0, \\ & \mathbf{z} \ge 0, a \ge 0. \end{aligned}$$

Proof. By (14), $\inf_{X \in \mathcal{X}} \bar{p}_x(X) = \inf_{\mathbf{y} \in C} \mathbf{p}^T [\mathbf{y} + \mathbf{e}_m]^+$, with $C = cl\ cone(S - x\mathbf{e}_m)$. Note that

$$S - x\mathbf{e}_m = \{\mathbf{x} | A(\mathbf{x} + x\mathbf{e}_m) \le \mathbf{b}\} = \{\mathbf{x} | A\mathbf{x} \le \mathbf{b} - xA\mathbf{e}_m\}.$$

Therefore,

$$\begin{aligned} cone(S - x\mathbf{e}_m) &= \{a\mathbf{x} | A\mathbf{x} \le \mathbf{b} - xA\mathbf{e}_m, a > 0\} \cup \{0\} = \\ &= \{\mathbf{y} | A\mathbf{y} \le a(\mathbf{b} - xA\mathbf{e}_m), a > 0\} \cup \{0\}, \\ cl\ cone(S - x\mathbf{e}_m) &= \{\mathbf{y} | A\mathbf{y} \le a(\mathbf{b} - xA\mathbf{e}_m), a \ge 0\}. \end{aligned}$$

Finally, by introducing $\mathbf{z} = [\mathbf{y} + \mathbf{e}_m]^+$, we obtain the LP reformulation. □

Let us proceed with the second formulation, $\min_{\mathbf{w} \in W} \bar{p}_x(f(\mathbf{w}; \mathbf{X}))$, which assumes that variables $X_1, \dots, X_n$ can be observed, but can not be controlled, while the random variable $f(\mathbf{w}; \mathbf{X})$ can be controlled by the vector $\mathbf{w}$. We assume that $\mathbf{X} = (X_1, \dots, X_n)$ is a random vector of dimension n with a given distribution, $\mathbf{w} \in \mathbb{R}^k$ is a vector of control parameters, and $W \subseteq \mathbb{R}^k$ is a feasible convex set for the control variable. The function $f : \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}$ is assumed to be a *Caratheodory function*, i.e. $f(\mathbf{w}; \cdot)$ is measurable for every $\mathbf{w}$, and $f(\cdot; \mathbf{X})$ is continuous for almost every (a.e.) $\mathbf{X}$. We also assume f to be a convex function of $\mathbf{w}$ for a.e. $\mathbf{X}$ and L_1-finite: $E|f(\mathbf{w}; \mathbf{X})| < \infty$ for every $\mathbf{w} \in W$. These properties guarantee that function $\bar{p}_x(f(\mathbf{w}; \mathbf{X}))$ is properly defined and finite valued on W.

The following proposition shows that for the case of convex f, the function $\bar{p}_x(f(\mathbf{w}; \mathbf{X}))$ is quasi-convex, which is an attractive feature for optimization: see, for example, paper [9] on interior point methods or paper [10] on multiobjective optimization.

Proposition 4.7. *For some random vector $\mathbf{X}'$, let $f(\mathbf{w}; \mathbf{X}')$ be a convex function of $\mathbf{w}$ on a convex set $W \subseteq \mathbb{R}^k$, and let $f(\mathbf{w}; \mathbf{X}') \in \mathcal{L}^1(\Omega)$ for all $\mathbf{w} \in W$. Then, for $x \in \mathbb{R}$, $\bar{p}_x(f(\mathbf{w}; \mathbf{X}'))$ is a quasi-convex function of $\mathbf{w}$ on W.*

Proof. Convexity of f implies

$$f(\mathbf{w}^M;\mathbf{X}) \le \lambda f(\mathbf{w}^1;\mathbf{X}) + (1-\lambda) f(\mathbf{w}^2;\mathbf{X}),$$

for $\mathbf{w}^M = \lambda \mathbf{w}^1 + (1-\lambda)\mathbf{w}^2$. Then, using monotonicity of $\bar{p}_x(X)$, see proposition 3.6,

$$\bar{p}_x(f(\mathbf{w}^M;\mathbf{X})) \le \bar{p}_x(\lambda f(\mathbf{w}^1;\mathbf{X}) + (1-\lambda) f(\mathbf{w}^2;\mathbf{X})).$$

$\bar{p}_x(X)$ is a quasi-convex function of X, see proposition 3.4. Quasi-convexity of a function $\bar{p}_x(X)$ is equivalent to $\bar{p}_x(\lambda X^1 + (1-\lambda)X^2) \le \max\{\bar{p}_x(X^1), \bar{p}_x(X^2)\}$. Then,

$$\bar{p}_x(\lambda f(\mathbf{w}^1;\mathbf{X}) + (1-\lambda) f(\mathbf{w}^2;\mathbf{X})) \le \max\{\bar{p}_x(f(\mathbf{w}^1;\mathbf{X})), \bar{p}_x(f(\mathbf{w}^2;\mathbf{X}))\}.$$

Therefore,

$$\bar{p}_x(f(\mathbf{w}^M;\mathbf{X})) \le \max\{\bar{p}_x(f(\mathbf{w}^1;\mathbf{X})), \bar{p}_x(f(\mathbf{w}^2;\mathbf{X}))\},$$

i.e., $\bar{p}_x(f(\mathbf{w};\mathbf{X}))$ is a quasi-convex function of $\mathbf{w}$. □

Before we proceed to the convex reformulation of the considered problem, let us introduce the operation of *right scalar multiplication* from [17]:

$$(f\lambda)(x) = \begin{cases} \lambda f(x/\lambda), & \text{if } \lambda > 0; \\ 0, & \text{if } \lambda = 0, x = 0; \\ +\infty, & \text{if } \lambda = 0, x \neq 0, \end{cases} \tag{15}$$

and for any proper convex function f on $\mathbb{R}^k$, the function h on $\mathbb{R}^{k+1}$ defined by

$$h(x,\lambda) = \{(f\lambda)(x), \text{ if } \lambda \ge 0; +\infty, \text{ if } \lambda < 0\},$$

is a positively homogeneous proper convex function. Note that $(f1)(x) = f(x)$, and $\lambda f(x) = (f\lambda)(\lambda x)$ for $\lambda \ge 0$. Suppose that $x \in S$, where the set S is convex, then $(\lambda x, \lambda) \in cone(S \times \{1\})$, i.e. lies in a convex cone. Therefore, *non-convex* problem of minimizing $\lambda f(x)$ subject to $x \in S$ and $\lambda \ge 0$ is equivalently reformulated as a *convex* problem of minimizing $(f\lambda)(y)$ subject to $(y,\lambda) \in cone(S \times \{1\})$, where $y = \lambda x$. The expression for the new feasible region $cone(S \times \{1\})$ can also be simplified in certain cases with (15). Suppose $S = \{x | \mathbf{g}(x) \le 0; x \in \mathcal{K}\}$, where $\mathbf{g}(x) = (g_1(x), \dots, g_d(x))$, $g_j(x)$ are convex functions for $j = 1, \dots, d$, and $\mathcal{K}$ is a convex cone. Then, $g_j(x) \le 0 \Leftrightarrow (g_j\lambda)(\lambda x) \le 0$, $\lambda > 0$, and $x \in \mathcal{K} \Leftrightarrow \lambda x \in \mathcal{K}$, $\lambda > 0$, hence, $cone(S \times \{1\}) = \{(\lambda x, \lambda) | \mathbf{h}(\lambda x, \lambda) \le 0; \lambda x \in \mathcal{K}; \lambda > 0\} \cup \{(0,0)\}$, where $\mathbf{h}(y,\lambda) = ((g_1\lambda)(y), \dots, (g_d\lambda)(y))$. Let $y = \lambda x$, then

$$cone(S \times \{1\}) = \{(y,\lambda) | \mathbf{h}(y,\lambda) \le 0; y \in \mathcal{K}; \lambda \ge 0\}. \tag{16}$$

Proposition 4.8. *Let $\mathbf{X}$ be a random vector, $x \in \mathbb{R}$, and let $f(\mathbf{w};\mathbf{X})$ be a convex function of $\mathbf{w}$, defined on a convex set $W \subseteq \mathbb{R}^k$. Let also $E|f(\mathbf{w};\mathbf{X})| < \infty$ for all $\mathbf{w} \in W$. Then bPOE minimization problem is reformulated equivalently as a following convex program:*

$$\inf_{\mathbf{w} \in W} \bar{p}_x(f(\mathbf{w};\mathbf{X})) = \inf_{(\mathbf{v},a) \in V} E\left[(fa)(\mathbf{v};\mathbf{X}) - ax + 1\right]^+,$$

where $(fa)(\mathbf{v};\mathbf{X})$ is defined by (15) for $f(\cdot;\mathbf{X})$, and $V = cone(W \times \{1\})$ is a convex cone. Furthermore, if $W = \{\mathbf{w} | \mathbf{g}(\mathbf{w}) \le 0; \mathbf{w} \in \mathcal{K}\}$, where $\mathbf{g}(\mathbf{w}) = (g_1(\mathbf{w}), \dots, g_d(\mathbf{w}))$, $g_j(\mathbf{w})$ are convex functions for $j = 1, \dots, d$, and $\mathcal{K}$ is a convex cone, then

$$V = \{(\mathbf{v},a) | \mathbf{h}(\mathbf{v},a) \le 0; \mathbf{v} \in \mathcal{K}; a \ge 0\},$$

where $\mathbf{h}(\mathbf{v},a) = ((g_1a)(\mathbf{v}), \dots, (g_da)(\mathbf{v}))$.

Proof. Note that

$$\inf_{\mathbf{w}\in W} \bar{p}_x(f(\mathbf{w};\mathbf{X})) = \inf_{a\geq 0, \mathbf{w}\in W} E[af(\mathbf{w};\mathbf{X}) - ax + 1]^+,$$

hence, we can apply (15) to rewrite $af(\mathbf{w};\mathbf{X}) = (fa)(a\mathbf{w};\mathbf{X})$. Since W is a convex set, then $\mathbf{w} \in W, a \geq 0 \Leftrightarrow (a\mathbf{w}, a) \in V := cone(W \times \{1\})$. Denoting $\mathbf{v} = a\mathbf{w}$, we finish the first part of the proof. The second part follows from (16). □

An optimization problem might not achieve the optimal inf value in any point from a given feasible region if this region is not a closed set, or if the objective function is not closed. Note that we were able to introduce such convex reformulation in proposition 4.5, that an objective function is closed and a feasible region closed. This is not necessarily the case with proposition 4.8 simply because the right scalar multiplication does not necessarily produce a closed convex function. It is easy to see that for definition (15) the only values of variable, which might violate closure condition, are $\lambda = 0$, $x \neq 0$. Instead, let us define

$$(f\circ\lambda)(x) = \lim_{\mu>0,\ \mu\to\lambda} \mu f\left(\frac{x}{\mu}\right) = \begin{cases} \lambda f(x/\lambda), & \text{if } \lambda > 0; \\ 0, & \text{if } \lambda = 0, x = 0; \\ \lim_{\lambda\to+0} \lambda f(x/\lambda), & \text{if } \lambda = 0, x \neq 0. \end{cases} \tag{17}$$

Since f is convex, the limit is well-defined, although not necessarily finite. For example, an affine function $f(x) = Ax + b$ will produce $(f\circ\lambda)(x) = Ax + \lambda b$, hence, $(f\circ 0)(x) = Ax$, while $(f0)(x) = +\infty$, except when $x = 0$.

Corollary 4.9. *In the formulation of proposition 4.8, assume that convergence of* $\{(a^i, a^i\mathbf{w}^i)\}_{i=1}^{\infty}$ *to* $(a, \mathbf{v})$ *implies convergence of variables* $\{a^i f(\mathbf{w}^i;\mathbf{X})\}_{i=1}^{\infty}$ *to variable* $(f\circ a)(\mathbf{v};\mathbf{X})$ *in* $\mathcal{L}^1$. *Then,*

$$\inf_{\mathbf{w}\in W} \bar{p}_x(f(\mathbf{w};\mathbf{X})) = \inf_{(\mathbf{v},a)\in V} E\left[(f\circ a)(\mathbf{v};\mathbf{X}) - ax + 1\right]^+,$$

where $V = cl\ cone(W \times \{1\})$*, and objective is closed. If* $W = \{\mathbf{w}|\mathbf{g}(\mathbf{w}) \leq 0; \mathbf{w} \in \mathcal{K}\}$*, then*

$$V = \{(\mathbf{v}, a)|\mathbf{h}(\mathbf{v}, a) \leq 0; \mathbf{v} \in \mathcal{K}; a \geq 0\},$$

where $\mathbf{h}(\mathbf{v}, a) = ((g_1\circ a)(\mathbf{v}), \ldots, (g_d\circ a)(\mathbf{v}))$.

Proof. Since the closed counterpart of $(fa)(\mathbf{v};\mathbf{X})$ is $(f\circ a)(\mathbf{v};\mathbf{X})$, then placing the latter instead of the former in an inf optimization problem under any feasible region will not change the optimal value and might add optimal solutions which were not attained previously. Note that the feasible region can be extended to $V = cl\ cone(W\times\{1\})$: let $\mathbf{v}^i = a^i\mathbf{w}^i$, then $a^i f(\mathbf{w}^i;\mathbf{X}) = (f\circ a)(\mathbf{v}^i;X)$; since $(a^i, \mathbf{v}^i) \to (a, \mathbf{v})$ implies $(f\circ a)(\mathbf{v}^i;\mathbf{X}) \xrightarrow{\mathcal{L}^1} (f\circ a)(\mathbf{v};\mathbf{X})$, then $E[(f\circ a)(\mathbf{v}^i;\mathbf{X}) - ax + 1]^+ \to E[(f\circ a)(\mathbf{v};\mathbf{X}) - ax + 1]^+$ is also implied. Change of right multiplication from (15) to (17) changes set $\{(\mathbf{v}, a)|(g_j a)(\mathbf{v}) \leq 0\}$ to its closure $\{(\mathbf{v}, a)|(g_j\circ a)(\mathbf{v}) \leq 0\}$, hence, using definition (17) is equivalent to extending the feasible region to $V = cl\ cone(W \times \{1\})$. □

Note that for the important case of linear constraints, i.e. $W = \{\mathbf{w}|A\mathbf{w} \leq b\}$ the transformed feasible set is $V = cl\ cone(W \times \{1\}) = \{(\mathbf{v}, a)|A\mathbf{v} - ab \leq 0\}$. That is, new constraints are also linear and easily represented.

The additional convergence property in corollary 4.9 may seem cumbersome, however, it holds in many cases. Below are some examples, for which the property holds.

Example 5. *Let the probability space for the vector* $\mathbf{X}$ *be finite. Suppose that* $\{(a^i, a^i\mathbf{w}^i)\}_{i=1}^{\infty} \to (a, \mathbf{v})$ *as* $i \to \infty$. *Then, due to the continuity of* f, *for every event* ω_j *from the probability space, values* $a^i f(a^i\mathbf{w}^i; \mathbf{X}(\omega_j)) \to (f \circ a)(\mathbf{v}; \mathbf{X}(\omega_j))$ *as* $i \to \infty$. *Hence, convergence holds for the finite sum:*

$$\sum_{j=1}^{m} P(\omega_j) \left[a^i f(a^i\mathbf{w}^i; \mathbf{X}(\omega_j))\right]^+ \to \sum_{j=1}^{m} P(\omega_j) \left[(f \circ a)(\mathbf{v}; \mathbf{X}(\omega_j))\right]^+ .$$

Denoting $p_j := P(\omega_j)$ *and* $\mathbf{X}_j := \mathbf{X}(\omega_j)$, *we can rewrite the corresponding bPOE minimization problem:*

$$\begin{array}{ll} \inf\limits_{(\mathbf{v},a)\in V} & \sum\limits_{j=1}^{m} p_j \left[(f \circ a)(\mathbf{v}; \mathbf{X}_j) - ax + 1\right]^+ \\ s.t. & (\mathbf{v}, a) \in V = cl\ cone(W \times \{1\}). \end{array}$$

Papers [14, 12, 25] considered several bPOE minimization problems with a finite observed data sample $(\mathbf{X}_1, \ldots, \mathbf{X}_m)$.

Example 6. *Let* $E\|\mathbf{X}\| < \infty$ *and* $f(\mathbf{w}; \mathbf{X}) = \max_{t=1,\ldots,T}\{\mathbf{Y}^t\mathbf{w} + Z^t\}$, *where* $\mathbf{X} = (\mathbf{Y}^1, Z^1, \ldots, \mathbf{Y}^T, Z^T)$. *Then* $(fa^i)(\mathbf{v}^i; \mathbf{X}) - (fa)(\mathbf{v}; \mathbf{X}) \le (\|\mathbf{v}^i - \mathbf{v}\| + |a^i - a|)\|\mathbf{X}\|$*:*

$$\begin{aligned} (fa^i)(\mathbf{v}^i; \mathbf{X}) - (fa)(\mathbf{v}; \mathbf{X}) &= \max_{t=1,\ldots,T}\{\mathbf{Y}^t\mathbf{v}^i + a^i Z^t\} - \max_{t=1,\ldots,T}\{\mathbf{Y}^t\mathbf{v} + aZ^t\} \le \\ &\le \max_{t=1,\ldots,T}\{\mathbf{Y}^t(\mathbf{v}^i - \mathbf{v}) + (a^i - a)Z^t\} + \max_{t=1,\ldots,T}\{\mathbf{Y}^t\mathbf{v} + aZ^t\} - \max_{t=1,\ldots,T}\{\mathbf{Y}^t\mathbf{v} + aZ^t\} \le \\ &\le \max_{t=1,\ldots,T}\{\|\mathbf{Y}^t\|\|\mathbf{v}^i - \mathbf{v}\| + |a^i - a||Z^t|\} \le (\|\mathbf{v}^i - \mathbf{v}\| + |a^i - a|)\|\mathbf{X}\|, \end{aligned}$$

where the second transition holds since the max *function is subadditive, and the third transition uses the Cauchy-Bunyakovsky-Schwarz inequality. Similarly,* $(fa)(\mathbf{v}; \mathbf{X}) - (fa^i)(\mathbf{v}^i; \mathbf{X}) \le (\|\mathbf{v} - \mathbf{v}^i\| + |a - a^i|)\|\mathbf{X}\|$, *therefore,*

$$E|(fa^i)(\mathbf{v}^i; \mathbf{X}) - (fa)(\mathbf{v}; \mathbf{X})| \le (\|\mathbf{v} - \mathbf{v}^i\| + |a - a^i|)E\|\mathbf{X}\| \to 0,$$

as $i \to \infty$, *i.e.* $\{(fa^i)(\mathbf{v}^i; \mathbf{X})\}_{i=1}^{\infty}$ *converges to* $(fa)(\mathbf{v}; \mathbf{X})$ *in* $\mathcal{L}^1$. *We can rewrite the corresponding bPOE minimization problem as follows:*

$$\begin{array}{ll} \inf\limits_{(\mathbf{v},a)\in V} & E\left[\max\limits_{t=1,\ldots,T}\{\mathbf{Y}^t\mathbf{v} + aZ^t\} - ax + 1\right]^+ \\ s.t. & (\mathbf{v}, a) \in V = cl\ cone(W \times \{1\}). \end{array}$$

Note here that if also W *is polyhedral, then* V *is polyhedral, and the problem can be reformulated as an LP*[4] *for a discretely distributed* $\mathbf{X}$. *This is exactly the problem solved in [25]. Problems with* $T = 1$, *i.e. without maximization over several terms, were formulated in [14, 12].*

[4] The resulting linear programming problem might be finite- or infinite- dimensional, depending on the number of atoms in $\mathbf{X}$. For infinite dimensional linear programming see [23].

5. Conclusions

This paper defined bPOE and studied its mathematical properties. It was proved that bPOE is the minimal upper bound for POE in the class of Schur-convex functions. The presented procedure for finding such bounds is universal, and could be applied to other practically important functionals. The two problem families, bPOE and superquantile minimization, share the same, with minor exceptions, set of optimal solutions. Also, frontiers of parameter-objective pairs have a significant intersection. Thus, optimization of bPOE can be performed via multiple calls of superquantile minimization solver. In the case of a convex feasible region and a convex objective function, bPOE minimization is reduced to the convex programming.

A potential question for future research is how to define a buffered counterpart to a multidimensional POE, calculated for random vectors rather than random variables. This is a very important question because POE and distribution function for random vectors are very popular and evidently bPOE and superdistribution function also can be defined in multidimensional case.

Appendix A. Properties of Superquantile (CVaR)

Proposition Appendix A.1. *Suppose $X \in \mathcal{L}^1(\Omega)$. Let us define $\beta = P(X = \sup X)$ for $\sup X < \infty$ and $\beta = 0$, otherwise. The $\bar{q}_\alpha(X)$ is a continuous strictly increasing function of α on $[0, 1-\beta]$, and $\bar{q}_\alpha(X) = \sup X$ on $(1-\beta, 1)$.*

Proof. Note that if for $\alpha_1 < \alpha_2$ we have $\bar{q}_{\alpha_1}(X) = \bar{q}_{\alpha_2}(X)$, then, by property of superquantile,

$$\min_c \left\{ c + \frac{1}{1-\alpha_1} E[X-c]^+ \right\} = \min_c \left\{ c + \frac{1}{1-\alpha_2} E[X-c]^+ \right\}.$$

For each value of c, if $c < \sup X$, then $E[X-c]^+ > 0$ and $c + \frac{1}{1-\alpha_1}E[X-c]^+ < c + \frac{1}{1-\alpha_2}E[X-c]^+$. Therefore, $\arg\min_c \left\{ c + \frac{1}{1-\alpha_1}E[X-c]^+ \right\} = \sup X$. It proves that $\bar{q}_\alpha(X)$ as a function of α can have only one interval of constancy, which is for $\alpha \in [1 - P(X = \sup X), 1]$. For the interval $\alpha \in [0, 1 - P(X = \sup X)]$ the function $\bar{q}_\alpha(X)$ is strictly increasing in α. □

It is easy to show that superquantile is also concave w.r.t. mixture operation. The next proposition shows slightly more general statement.

Proposition Appendix A.2. $(1-\alpha)\bar{q}_\alpha(X)$ *is a concave function of* (X, α) *w.r.t. mixture operation and addition operation, correspondingly, i.e.,*

$$(1-\alpha_{\mathrm{m}})\bar{q}_{\alpha_{\mathrm{m}}}(\lambda X_1 \oplus (1-\lambda)X_2) \geq \lambda \left[(1-\alpha_1)\bar{q}_{\alpha_1}(X_1)\right] + (1-\lambda)\left[(1-\alpha_2)\bar{q}_{\alpha_2}(X_2)\right],$$

where $\alpha_{\mathrm{m}} = \lambda\alpha_1 + (1-\lambda)\alpha_2$, $\lambda, \alpha_1, \alpha_2 \in [0,1]$, $X_1, X_2 \in \mathcal{L}^1(\Omega)$.

Proof. Note that $c(1-\alpha) + E[X-c]^+$ is jointly linear w.r.t. mixture operation on X and addition operation on α. Since $(1-\alpha)\bar{q}_\alpha(X) = \min_c c(1-\alpha) + E[X-c]^+$, then, as a minimum over a collection of linear functions, $(1-\alpha)\bar{q}_\alpha(X)$ is concave jointly on (X, α) w.r.t. mixture operation and addition operation, correspondingly. □

The statement below rephrases the Proposition 6.1 in [16], it has motivated the proposition Appendix A.2 in the first place, and it follows immediately from proposition Appendix A.2.

Corollary Appendix A.3. *Let $X(\mathbf{x}, \mathbf{p})$ be a discretely distributed random variable, taking values $\mathbf{x} = (x_1, \dots, x_m)$ with probabilities $\mathbf{p} = (p_1, \dots, p_m)$, $p_i \ge 0$, $\sum_{i=1}^m p_i = 1$. Then function $\bar{q}_\alpha(X(\mathbf{x}, \mathbf{p}))$ is a concave function of $\mathbf{p}$.*

Appendix B. Properties of Schur-Convex Functions

Proposition Appendix B.1. *For $f : \mathcal{L}^1(\Omega) \to \mathbb{R}$, in the class of Schur-convex functions, its minimal upper bound is unique and is given by $\bar{f}(X) \equiv \sup_{Y \le_{cx} X} f(Y)$.*

Proof. For $Z \le_{cx} X$, by transitivity of convex order, $\{Y | Y \le_{cx} Z\} \subseteq \{Y | Y \le_{cx} X\}$, hence, $\bar{f}(Z) \le \bar{f}(X)$, that is, $\bar{f}$ is indeed Schur-convex. If there is a Schur-convex $g \ge f$, then for any $X \in \mathcal{L}^1(\Omega)$ holds $g(X) \ge g(Y)$ when $Y \le_{cx} X$, therefore, $g(X) \ge \sup_{Y \le_{cx} X} g(Y) \ge \sup_{Y \le_{cx} X} f(Y) = \bar{f}(X)$, which finishes the proof. □

Proposition Appendix B.2. *Let the probability space be non-atomic, then a closed law-invariant quasi-convex function on $\mathcal{L}^1(\Omega)$ is Schur-convex.*

Proof. Let function f be closed law-invariant quasi-convex on $\mathcal{L}^1(\Omega)$. For $X \in \mathcal{L}^1(\Omega)$ denote $K_X = \{Y \in \mathcal{L}^1(\Omega) | Y =_{st} X\}$, where $=_{st}$ means that random variables are identically distributed. By [1, Proposition 4.1][5], closed convex hull of K_X is $\{Y \in \mathcal{L}^1(\Omega) | Y \le_{cx} X\}$. Note that $L_{f(X)} = \{Y \in \mathcal{L}^1(\Omega) | f(Y) \le f(X)\}$ contains K_X since f is law-invariant. Since f is closed quasi-convex, $L_{f(X)}$ is closed convex, hence, contains *cl conv*(K_X). Therefore, $Y \le_{cx} X$ implies $f(Y) \le f(X)$, and f is Schur-convex. □

Proposition Appendix B.3. *For a non-atomic probability space, lower quantile defined on $\mathcal{L}^1(\Omega)$ has superquantile as its unique minimal Schur-convex upper bound.*

Proof. Since superquantile is closed convex and law-invariant, it is Schur-convex. To finalize the proof, let us provide for any X and α a random variable $Y_\beta \le_{cx} X$ such that $q_\beta(Y_\beta) = \bar{q}_\beta(X)$. Indeed, consider Y_β, a random variable such that $\bar{q}_\alpha(Y_\beta) = \bar{q}_\alpha(X)$ for $\alpha \in [0, \beta]$ and $\bar{q}_\alpha(Y_\beta) = \bar{q}_\beta(X)$ for $\alpha \in [\beta, 1]$. Note that $EY_\beta = EX$, hence, $Y_\beta \le_{cx} X$, and also that $P(Y_\beta = \bar{q}_\beta(X)) = 1 - \beta$, hence, $q_\beta(Y_\beta) = \bar{q}_\beta(X)$. □

Acknowledgements

Authors would like to thank Prof. R. Tyrrell Rockafellar for the valuable insights and the motivation for this research; Dr. Matthew Norton, who has found the first variant of the optimization formula for bPOE (the corner stone for bPOE minimization); Dr. Viktor Kuzmenko, for the productive discussions; Prof. Alexander Shapiro, for his helpful comments and suggestions.

[5]Equivalent result follows from Theorem 2 and Remark 2 of [6], which show that *cl conv*$(\{Y \in \mathcal{L}^1(\Omega) | Y \le_1 X\}) = \{Y \in \mathcal{L}^1(\Omega) | Y \le_2 X\}$. Since $\{Y \in \mathcal{L}^1(\Omega) | Y \le_1 X\} = K_X - \mathcal{L}^1_+$, then its closed convex hull is *cl conv*$(K_X) - \mathcal{L}^1_+$. Intersection with the hyperplane $EY = EX$ results in *cl conv*(K_X). On the other hand, $\{Y \in \mathcal{L}^1(\Omega) | Y \le_2 X, EY = EX\} = \{Y \in \mathcal{L}^1(\Omega) | Y \le_{cx} X\}$. Finally, *cl conv*$(K_X) = \{Y \in \mathcal{L}^1(\Omega) | Y \le_{cx} X\}$.

This research 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; by the AFRL Mathematical Modeling and Optimization Institute; and by the DARPA EQUiPS program, grant SNL 014150709, Risk-Averse Optimization of Large-Scale Multiphysics Systems.

References

[1] R.-A. Dana, *A representation result for concave schur concave functions*, Mathematical finance, 15 (2005), pp. 613–634.

[2] J. R. Davis and S. Uryasev, *Analysis of tropical storm damage using buffered probability of exceedance*, Natural Hazards, (2015), pp. 1–19.

[3] E. Delage, D. Kuhn, and W. Wiesemann, *"Dice"-sion Making under Uncertainty: When Can a Random Decision Reduce Risk?*, -, (2016). Available from Optimization Online.

[4] D. Dentcheva and G. Martinez, *Two-stage stochastic optimization problems with stochastic ordering constraints on the recourse*, European Journal of Operational Research, 219 (2012), pp. 1–8.

[5] D. Dentcheva and A. Ruszczynski, *Optimization with stochastic dominance constraints*, SIAM Journal on Optimization, 14 (2003), pp. 548–566.

[6] D. Dentcheva and A. Ruszczynski, *Convexification of stochastic ordering*, Comptes Rendus de l'Academie Bulgare des Sciences, 57 (2004), pp. 4–11.

[7] H. Föllmer and A. Schied, *Convex measures of risk and trading constraints*, Finance and stochastics, 6 (2002), pp. 429–447.

[8] H. Föllmer and A. Schied, *Stochastic finance: an introduction in discrete time*, Walter de Gruyter, 2011.

[9] N. Langenberg and R. Tichatschke, *Interior proximal methods for quasiconvex optimization*, Journal of Global Optimization, 52 (2012), pp. 641–661.

[10] K. Miettinen, *Nonlinear multiobjective optimization*, vol. 12, Springer Science & Business Media, 2012.

[11] A. Nemirovski and A. Shapiro, *Convex approximations of chance constrained programs*, SIAM Journal on Optimization, 17 (2006), pp. 969–996.

[12] M. Norton, A. Mafusalov, and S. Uryasev, *Cardinality of Upper Average and Application to Network Optimization*, tech. report, Research Report 2015-1, ISE Dept., University of Florida, 2015.

[13] M. Norton, A. Mafusalov, and S. Uryasev, *Soft Margin Support Vector Classification as Buffered Probability Minimization*, Journal of Machine Learning Research, (2017). accepted.

[14] M. Norton and S. Uryasev, *Maximization of AUC and Buffered AUC in Classification*, tech. report, Research Report 2014-2, ISE Dept., University of Florida, 2014.

[15] I. Olkin and A. W. Marshall, *Inequalities: theory of majorization and its applications*, vol. 143, Academic press, 2016.

[16] K. Pavlikov and S. Uryasev, *CVaR Distance between Univariate Probability Distributions and Approximation Problems*, tech. report, Research Report 2015-6, ISE Dept., University of Florida, 2015.

[17] R. T. Rockafellar, *Convex analysis. princeton landmarks in mathematics*, 1997.

[18] R. T. Rockafellar, *Safeguarding Strategies in Risky Optimization*, in Presentation at the International Workshop on Engineering Risk Control and Optimization, Gainesville, FL, 2009.

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

[20] R. T. Rockafellar and J. O. Royset, *Random variables, monotone relations, and convex analysis*, Mathematical Programming, 148 (2014), pp. 297–331.

[21] R. T. Rockafellar and S. Uryasev, *Conditional Value-at-Risk for general loss distributions*, Journal of Banking and Finance, (2002), pp. 1443–1471.

[22] R. T. Rockafellar and S. Uryasev, *The Fundamental Risk Quadrangle in Risk Management, Optimization and Statistical Estimation*, Surveys in Operations Research and Management Science 18, (2013).

[23] H. E. Romeijn, R. L. Smith, and J. C. Bean, *Duality in infinite dimensional linear programming*, Mathematical programming, 53 (1992), pp. 79–97.

[24] M. Shaked and J. G. Shanthikumar, *Stochastic orders*, Springer Science & Business Media, 2007.

[25] D. Shang, V. Kuzmenko, and S. Uryasev, *Cash flow matching with risks controlled by buffered probability of exceedance and conditional value-at-risk*, Annals of Operations Research, (2016).

[26] S. Uryasev, *Buffered Probability of Exceedance and Buffered Service Level: Definitions and Properties*, tech. report, Research Report 2014-3, ISE Dept., University of Florida, 2014.

[27] M. Zabarankin and S. Uryasev, *Statistical decision problems*, Springer, 2014.