CVaR Distance Between Univariate Probability Distributions and Approximation Problems[1]

Konstantin Pavlikov, Stan Uryasev

Research Report 2015-6

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

First draft: December 2012, This draft: March 23, 2016
Correspondence should be addressed to: Stan Uryasev

Abstract

We define new distances between univariate probability distributions, based on the concept of the CVaR norm. The problem of approximation of one discrete probability distribution by another one, potentially with a smaller number of outcomes, is considered. Such problems arise in the context of scenario reduction and the approximation is performed by minimizing the new distance. We find: (i) optimal locations of atoms of the approximating distribution with fixed probabilities and (ii) optimal probabilities with a priori fixed approximating positions of atoms. These two steps are further combined in an iterative procedure for finding both atom locations and their probabilities. Numerical experiments show high efficiency of the proposed approaches, solved with convex and linear programming.

Keywords: scenario reduction, distance minimization, Conditional Value-at-Risk, CVaR norm

---

1. Introduction

The problem of approximation of probability distributions is considered. In decision science literature, it is often reported a need to approximate one probability distribution by another simpler one. Typically, a continuous distribution is approximated by a discrete one with a small number of outcomes. For instance, three point approximations have been extensively studied, see Keefer and Bodily (1983) and Keefer (1994), and five point approximations are sometimes employed.

A number of approximation approaches has been discovered, including the mean (or the median) bracket or the moment matching methods. In the bracket approach, e.g., Miller III and Rice (1983) and Hammond and Bickel (2013), the support of the target

---

distribution is divided into several brackets (not necessary equal in probability) and the mean or the median of every bucket is chosen to be a discrete representation of that part of the target distribution in the approximation. Another approach is based on the idea that the approximation should match the moments of the original distribution. Matching the moments has been recognized to be especially important in computing value lotteries and their certain equivalents, Smith (1993). The idea is as follows: the value function frequently can be well approximated by a polynomial (with degree m) of a random variable. Thus, if that random variable is approximated by a simpler discrete variable having the same m first moments, the expected value function based on the approximation is no different from that one based on the original random variable. The key result here states that it is possible to match the first $2m - 1$ moments of the target distribution by a discrete one with only m outcomes, see Miller III and Rice (1983) and Smith (1993). Moreover, when the original distribution is not specified completely so fewer than $2m - 1$ moments of the original distribution are known, the resulting ambiguity of defining the approximation of the size m was suggested to be resolved using the entropy maximization, see Rosenblueth and Hong (1987).

While matching the moments is important in certain applications, we believe that accurately approximating the cumulative distribution function (cdf) of the target distribution is more important. In statistical literature, measuring the discrepancy between distributions is performed with help of distances based on the difference between cdfs. One of the most widely known distances is the Kolmogorov-Smirnov distance, which leads to the corresponding goodness of fit test, see for instance Gibbons and Chakraborti (2011) and Feller (1948). This distance is based on a single point where the absolute difference between two cdfs is maximized, and equals to the corresponding value of the absolute difference. Such measure is rather conservative. A number of alternative measures have been proposed, for instance the Cramer-von Mises distance, which is essentially based on the area under the weighted squared difference between two cumulative distribution functions, see Darling (1957). Another well-known example, the Anderson-Darling distance is also based on the squared difference between two cumulative distribution functions with a different weight function, see Boos (1981). The Kantorovich-Rubinstein distance is yet another popular example of distances. While it is originally defined as the cost of the optimal mass transportation plan to transport the probability mass from one distribution to another, see Villani (2009), in one dimensional case it is also connected to cdf functions and has been shown to be equal to the area between two cdf curves, Vallander (1973).

Recognizing the importance of accurate representation of the original cdf by the cdf of an approximation, we introduce the definition of the family of CVaR, (Rockafellar and Uryasev, 2000, 2002), distances between distributions, which extends the notion of the Kolmogorov-Smirnov distance. We consider the problem of approximation of the discrete distribution by another discrete distribution, possibly of a (much) smaller size. The objective of the approximation is to follow the cdf of the target distribution as closely as possible, which is achieved by minimizing new distances. Such an approximation problem generally consists of two subproblems: (i) how to find the outcomes of the approximating distribution with known probabilities and (ii) how to assign the probabilities when outcomes of the approximation are fixed. We show that each of these problems can be handled with linear or convex programming. Moreover, in a special case of the CVaR distance, when it corre-

sponds to the Kantorovich-Rubinstein distance, both approaches can be combined into an iterative procedure that simultaneously finds the outcomes of the approximation and their corresponding probabilities. Moreover, besides minimization of the distance, fitting the tails of the target distribution might be of a special importance in approximation problems dealing with risk management applications. The accuracy of tail approximation is measured by CVaRs (right and / or left tail) and it is shown that incorporating corresponding constraints can be done via addition of a set of linear constraints.

Finally, while in the current paper we consider approximations in one dimension, this might be an important subproblem of an approximation problem in higher dimensions, see Grigoriu (2009). There, approximation of distributions in higher dimensions is reduced to approximation of one-dimensional cdfs of marginal distributions, their moments and correlation matrices.

2. Risk Measure-based Distance between Maximal Monotone Relations

This section defines the notion of a risk measure-based distance between two probability distributions. Frequently, measuring the distance between probability distributions is based their cumulative distribution functions. Cumulative distribution functions received significant attention in statistical literature in studying problems like goodness of fit testing or determining whether two samples are withdrawn from the same distribution. A well-known example of cdf-based distance is the Kolmogorov-Smirnov distance, also called the uniform metric and employed heavily in statistical literature (Gibbons and Chakraborti, 2011) as well as in evaluation of distribution approximations, Smith (1993). It measures the distance solely based on the supremum of absolute difference between two cumulative distribution functions, which might be too insensitive and conservative for practical applications focused on tails of distributions. The main motivation of the definition of risk measure-based distances is to account for the differences between two cdfs beyond the maximum-distance point.

As it becomes clear later, it is convenient to define risk measure-based distances based on *maximal monotone relations*, Rockafellar and Royset (2014). A cdf can have points of discontinuity, however, if the corresponding jumps of that cdf function are filled with vertical segments, we obtain an example of the maximal monotone relation. In a similar fashion, the quantile function of a probability distribution can generate a maximal monotone relation. We will consider further distances based on both cumulative distribution functions and quantile functions, so it is convenient to define the notion of distance in a more general fashion. First, we define the notion of a monotone relation on a set $\mathcal{A} \subseteq \mathbb{R}$.

Definition 1. *(Rockafellar and Royset (2014)). Let $\mathcal{A} = [a, b] \subseteq \mathbb{R}$, possibly unbounded closed interval. A set $\Gamma = \{(x, p) \subset \mathcal{A} \times \mathbb{R}\}$ is called a monotone relation on $\mathcal{A}$ if $\forall (x_1, p_1), (x_2, p_2) \in \Gamma$*

$$(x_1 - x_2)(p_1 - p_2) \geq 0. \tag{1}$$

Set Γ is called a maximal monotone relation on $\mathcal{A}$ if there exists no monotone relation $\Gamma' \neq \Gamma$ on $\mathcal{A}$, such that $\Gamma \subset \Gamma'$.

Associated with a maximal monotone relation Γ on $\mathcal{A}$, the function $\Gamma(x)$, $x \in \mathcal{A}$, is defined. An arbitrary monotone relation can clearly contain a vertical segment, therefore,

3

let $\Gamma(x)$ be defined as

$$\Gamma(x) = \begin{cases} \inf\limits_{(x,p)\in\Gamma} p, & \text{if} \quad x = b < +\infty, \\ \sup\limits_{(x,p)\in\Gamma} p, & \text{otherwise}, \end{cases} \tag{2}$$

where b is the right point of the closed interval $\mathcal{A}$. Clearly, $\Gamma(x)$ is a nondecreasing function on $\mathcal{A}$.

Suppose F and G are two maximal monotone relations on $\mathcal{A}$ and we randomly pick a point $\xi \in \mathcal{A}$, so that the absolute difference between F and G becomes a random variable taking the value $|F(\xi) - G(\xi)|$. Specifically, we suppose there is an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and the random variable ξ is a $\mathcal{F}-$measurable function from Ω to $\mathcal{A}$, $\xi : \Omega \to A$. Let $\mathcal{A}$ be equipped with a Borel $\sigma-$algebra $\mathcal{B}$. Moreover, the auxiliary random variable ξ is supposed to have a probability distribution H, such that (i) it has a density function $h(x)$, $x \in \mathcal{A}$ and (ii) $h(x) > 0$ for any $x \in int(\mathcal{A})$. The distance (discrepancy metric) between F and G will be defined using a risk measure to the random variable $|F(\xi) - G(\xi)|$.

A risk measure $\mathcal{R}$ is a map from a space of random variables into $\mathbb{R}$. In our study $\mathcal{R}$ has to belong to a special class of risk measures, called coherent risk measures and defined in Artzner et al. (1999). To be coherent, a risk measure has to satisfy the following axioms (the axioms are in a slightly different from Artzner et al. (1999) form, due to Rockafellar and Uryasev (2013)):

- A1. $\mathcal{R}(\xi) = C$ for constant random variables $\xi = C$ a.s.,

- A2. $\mathcal{R}(\xi_1) \leq \mathcal{R}(\xi_2)$ for $\xi_1 \leq \xi_2$ a.s.,

- A3. $\mathcal{R}(\xi_1 + \xi_2) \leq \mathcal{R}(\xi_1) + \mathcal{R}(\xi_2)$,

- A4. $\mathcal{R}(\lambda\xi_1) = \lambda\mathcal{R}(\xi_1)$, for any $\lambda \in (0, +\infty)$.

Definition 2. *The risk measure-based distance between maximal monotone relations on $\mathcal{A}$, F and G, is defined through the corresponding functions $F(\cdot)$ and $G(\cdot)$, as follows:*

$$d^H(F, G) = \mathcal{R}\left(|F(\xi) - G(\xi)|\right). \tag{3}$$

The function $d^H(F, G)$ satisfies the usual properties of a probability metric, discussed in, for instance, Rachev et al. (2008), Chapter 3.

Proposition 2.1. *Let F, Z, G be maximal monotone relations on $\mathcal{A} = [a, b]$. If H is a distribution with a density function $h(x) > 0$, $\forall x \in int(\mathcal{A})$, the following properties hold:*
1. $d^H(F, G) \geq 0$
2. $d^H(F, G) = 0 \iff \mu\left(\{x : F(x) \neq G(x)\}\right) = 0$ where μ denotes Lebesgue measure
3. $d^H(F, G) = d^H(G, F)$
4. $d^H(F, Z) \leq d^H(F, G) + d^H(G, Z)$

Proof. First, correctness of definition $|F(\xi)-G(\xi)|$ needs to be shown, i.e., that $|F(\xi)-G(\xi)|$ has to be a $\mathcal{F}-$ measurable function. It is sufficient to show that $F(\xi)$ is measurable, i.e., that the preimage of an open set in $\mathcal{A}$ is in $\mathcal{F}$. In order to see this sufficiency, note first that the sum or the difference of two measurable functions is measurable, see for instance McDonald and Weiss (1999), Chapter 3. Then, it is well-known that a preimage (with respect to a continuous function) of an open set is another open set, see McDonald and Weiss (1999), Chapter 2, therefore the absolute value function also preserves measurability. Thus, consider the measurability of the function $F(\cdot)$. The function $F(\cdot)$ is associated with a maximal monotone relation F, therefore it is nondecreasing and has at most countable number of points of discontinuity, cf. Rudin (1964) for example. Therefore, $F(\cdot)$ can be approximated by a sequence of continuous nondecreasing functions $F_n(\cdot)$ such that the sequence converges pointwise to F: $\forall x \in \mathcal{A}$, $\lim\limits_{n\to\infty} F_n(x) = F(x)$. By Theorem 4.5 in McDonald and Weiss (1999), the function $F(\cdot)$ is measurable.

Properties 1,3,4 are trivial and direct consequences of the axioms of coherent risk measures. The proof of Property 2 follows next.

1. We start from the $\implies$ implication. Suppose $d^H(F,G) = 0$, in other words, $0 = \mathcal{R}(|F(\xi) - G(\xi)|) \leq \mathcal{R}(0)$, which implies by the property A2 of coherent risk measures that $\mathbb{P}(\omega : |F(\xi(\omega)) - G(\xi(\omega))| \leq 0) = \mathbb{P}(\omega : F(\xi(\omega)) = G(\xi(\omega))) = 1$. Thus, we obtain the following

$$\mathbb{P}(\omega : F(\xi(\omega)) \neq G(\xi(\omega))) = 0. \tag{4}$$

Let $\mathcal{A}' \subset \mathcal{A}$ denote the image of ξ. Clearly, $\mu(\mathcal{A}') = \mu(\mathcal{A})$ because of the absolute continuity of the distribution of ξ. Then,

$$\mu\{x \in \mathcal{A} : F(x) \neq G(x)\} = \mu\{x \in \mathcal{A}' : F(x) \neq G(x)\}. \tag{5}$$

Let $E = \{x \in \mathcal{A}' : F(x) \neq G(x)\}$. Consider a sequence of $\{\epsilon_k > 0\}$, $\epsilon_k \to 0$, $k \to +\infty$ and let $E_k = \{x \in \mathcal{A}' : F(x) \neq G(x), h(x) \geq \epsilon_k\}$. Clearly, $\bigcup\limits_{k=1}^{\infty} E_k = E$. Also,

$$0 = \mathbb{P}(\omega : F(\xi(\omega)) \neq G(\xi(\omega))) \geq \int_{E_k} h d\mu \geq \epsilon_k \int_{E_k} d\mu = \epsilon_k \mu(E_k) \implies \tag{6}$$

$$\mu\{E_k\} = 0, \qquad\qquad\qquad k = 1, \ldots, +\infty. \tag{7}$$

Thus

$$\mu(E) = \mu\left(\bigcup\limits_{k=1}^{\infty} E_k\right) \leq \sum\limits_{k=1}^{\infty} \mu(E_k) = 0. \tag{8}$$

2. The implication $\impliedby$ is more obvious. Let $E = \{x \in \mathcal{A} : F(x) \neq G(x)\}$. Then, due to ξ having the density function h,

$$\mathbb{P}(\omega : F(\xi(\omega)) \neq G(\xi(\omega))) = \int_E h d\mu = 0, \tag{9}$$

as an integral of the nonnegative function over the set of measure 0. $\qquad\square$

While general properties of the $d^H(F, G)$ with respect to general H and various $\mathcal{R}$ remain to be studied, here we focus on the following special case. First, $F(\cdot)$ and $G(\cdot)$ are based on discrete distributions with finite supports and $\mathcal{R}$ is chosen to be the Conditional Value-at-Risk (CVaR) risk measure, Rockafellar and Uryasev (2000, 2002), which is known to be coherent, see, for instance, Pflug (2000). Moreover, the distribution H is selected to be uniform, $H = U(\mathcal{A})$ on a bounded set $\mathcal{A}$. When $F(\cdot)$ and $G(\cdot)$ represent cumulative distribution functions, the distance is called the CVaR distance between distributions. If $F(\cdot)$ and $G(\cdot)$ are quantile functions, denoted by $F^{-1}(\cdot)$ and $G^{-1}(\cdot)$, then the corresponding distance is called the CVaR distance between quantile functions of corresponding distributions. See a more detailed discussion on maximal monotone relations, cumulative distribution functions and their inverse (quantile) functions in Rockafellar and Royset (2014).

3. CVaR Distance for Univariate Discrete Distributions With Finite Domains

In this section, we explore and simplify the definition of CVaR distance for the case of discrete distributions. Let $\mathcal{A} = [a, b]$ be a bounded closed interval in $\mathbb{R}$. Let F and G be two distributions on $\mathcal{A}$. In other words, any random variable having the distribution F or G takes values in $\mathcal{A}$ with probability 1. Let probability distribution F be defined by a set of outcomes $\mathbf{x} = (x_1, \ldots, x_n)$, $x_i \in \mathcal{A}$ with probabilities $p = (p_1, \ldots, p_n)$ and probability distribution G be defined by discrete set of outcomes $\mathbf{y} = (y_1, \ldots, y_m)$, $y_j \in \mathcal{A}$ with probabilities $q = (q_1, \ldots, q_m)$. We assume without loss of generality that components of $\mathbf{x}$ and $\mathbf{y}$ are presented in ascending order, i.e., $x_i < x_j$ and $y_i < y_j$, $i < j$.

Cumulative distribution functions $F(\cdot)$ and $G(\cdot)$ of distributions F and G are defined as follows:

$$F(z) = \sum_{i=1}^{n} p_i \mathbb{1}_{z \geq x_i}, \tag{10}$$

$$G(z) = \sum_{i=1}^{m} q_i \mathbb{1}_{z \geq y_i}. \tag{11}$$

where $\mathbb{1}_{z \geq x_i} = 1$ if $z \geq x_i$ and 0, otherwise. Moreover, the auxiliary random variable ξ with the distribution H will be assumed uniformly distributed on $\mathcal{A}$ throughout the rest of the paper, $H = U(\mathcal{A})$. Since both $F(\cdot)$ and $G(\cdot)$ are nondecreasing functions on $\mathcal{A}$, Definition 3 of the distance can be applied and is simplified as follows when $\mathcal{R}$ is Conditional Value-at-Risk. CVaR with confidence level α of a continuous random variable X that describes a "loss", is defined as follows:

$$\mathrm{CVaR}_\alpha(X) = E\left(X \big| X > q_\alpha\right), \tag{12}$$

where q_α is the α−quantile of the random variable X, defined by

$$q_\alpha(X) = \inf\{l \in \mathbb{R} : \mathbb{P}(X > l) \leq 1 - \alpha\}. \tag{13}$$

For a discrete random variable X and $\alpha \in [0, 1)$, it is more convenient to use the generalized CVaR definition introduced in Rockafellar and Uryasev (2000), as follows:

$$\mathrm{CVaR}_\alpha(X) = \min_c \left(c + \frac{1}{1 - \alpha} E\left[X - c\right]^+ \right). \tag{14}$$

In short, CVaR of a random variable is the average of a specified percentage of its largest outcomes, see some basic examples with discrete distributions in Pavlikov and Uryasev (2014). With that, the distance between distributions F and G is defined as follows.

Definition 3. *Let ξ be a uniform random variable on $\mathcal{A} = [a, b]$. Let $\alpha \in [0, 1)$. Then, the CVaR distance between F and G is defined as*

$$d_\alpha^U(F, G) = \langle\!\langle F(\xi) - G(\xi) \rangle\!\rangle_\alpha. \tag{15}$$

where $\langle\!\langle \cdot \rangle\!\rangle_\alpha$ denotes the CVaR norm (CVaR of the absolute value of a random variable) with confidence level α, introduced in Mafusalov and Uryasev (2016).

Let $\mathbf{t} = \{\mathbf{x} \cup \mathbf{y}\}$ be the union of sets of outcomes $\mathbf{x}$ and $\mathbf{y}$, with $t_1 = \min\{\mathbf{x} \cup \mathbf{y}\}$ and $t_s = \max\{\mathbf{x} \cup \mathbf{y}\}$. Then, the discrete random variable $F(\xi) - G(\xi)$ takes the following values

$$d_k = F(t_k) - G(t_k), \quad k = 1, \ldots, s - 1, \tag{16}$$

with probabilities

$$\mathbb{P}(d_k) = \frac{t_{k+1} - t_k}{|\mathcal{A}|}, \quad k = 1, \ldots, s - 1. \tag{17}$$

The definition of the CVaR distance is illustrated in Figure 1. The family of CVaR distances defined by (15) includes the Kolmogorov-Smirnov distance as a special case. First, recall the definition of the Kolmogorov-Smirnov distance.

Definition 4. *The Kolmogorov-Smirnov distance between two distributions with cumulative distribution functions $F(\cdot)$ and $G(\cdot)$ is defined as follows:*

$$d_{KS}(F, G) = \sup_z |F(z) - G(z)|. \tag{18}$$

The following remark establishes a connection between the Kolmogorov-Smirnov distance and the family of CVaR distances.

Remark 1. Definition 4 is a special case of the Definition 3 when $\alpha \to 1$, i.e.,

$$d_{KS}(F, G) = \lim_{\alpha \to 1_-} d_\alpha^U(F, G) =: d_1^U(F, G).$$

Another special case is the CVaR distance with confidence level $\alpha = 0$. It is also called the average distance and entails a simpler reformulation. The following definition explicitly simplifies the definition of the average distance.

Definition 5. *The average distance between two distributions F and G, denoted by d_{AV}^U, is defined as follows:*

$$d_{AV}^U(F, G) = \sum_{k=1}^{s-1} d_k \mathbb{P}(d_k), \tag{19}$$

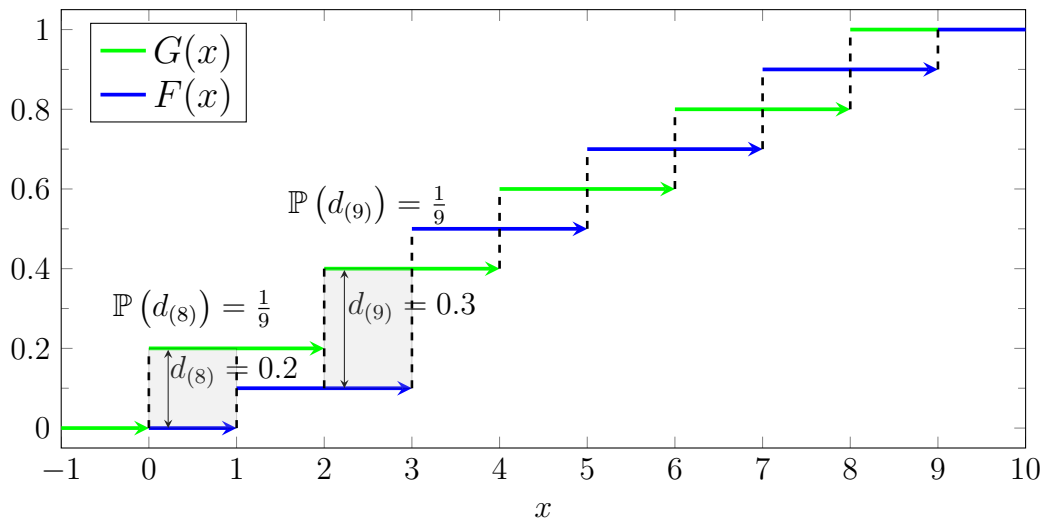*with d_k and $\mathbb{P}(d_k)$ defined according to (16), (17).*

Figure 1: Illustration of the CVaR distance with confidence level $\alpha = \frac{7}{9}$ between two discrete distributions, $\mathcal{A} = [0, 9]$. The largest absolute difference between two cdfs, $d_{KS}(F, G) = d_{(9)} = 0.3$, has the probability of occurrence $\frac{1}{9}$. The second largest absolute difference, $d_{(8)} = 0.2$ and also has the probability of occurrence $\frac{1}{9}$. Thus, $d^{U}_{7/9}(F, G) = \frac{0.3 + 0.2}{2} = 0.25$.
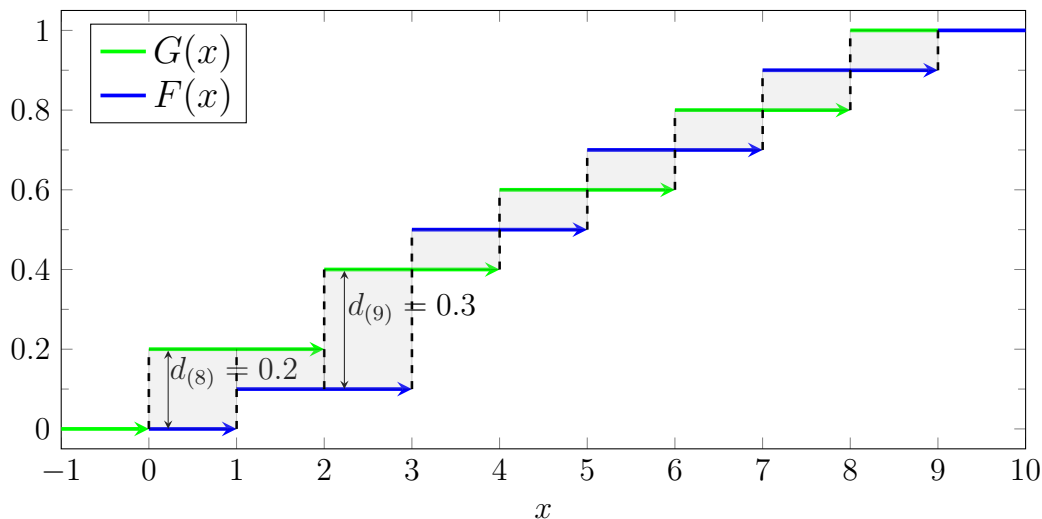
Figure 2: The area of the shaded region of the graph represents the Kantorovich-Rubinstein distance between F and G. The area scaled by the coefficient $\dfrac{1}{t_s - t_1} = \dfrac{1}{9}$ equals the average distance between F and G.

An illustration of the average distance is provided in Figure 2.

Now we discuss a connection between the average distance and another well-known distance, the Kantorovich-Rubinstein distance. Recall the definition of the Kantorovich-Rubinstein distance first, considered for the case of univariate distributions with finite domains.

Definition 6. *(Kantorovich-Rubinstein distance between two discrete distributions) Define a transportation plan of transporting the probability mass of F to the distribution G, as*

8

follows:

$w_{ij} =$ *probability transported from the outcome y_j of G to the outcome x_i of F,*

$c_{ij} =$ *transportation cost of the unit of probability mass from the y_j to the outcome x_i.*

Here we assume

$$c_{ij} = |x_i - y_j|, \quad i = 1, \ldots, n, \, j = 1, \ldots, m.$$

The Kantorovich-Rubinstein distance is defined as the optimal value of the following transportation problem:

$$d_K(F, G) = \min_{w_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} w_{ij} \tag{20}$$

subject to

$$\sum_{i=1}^{n} w_{ij} = q_j, \qquad\qquad\qquad j = 1, \ldots, m, \tag{21}$$

$$\sum_{j=1}^{m} w_{ij} = p_i, \qquad\qquad\qquad i = 1, \ldots, n, \tag{22}$$

$$w_{ij} \geq 0, \qquad\qquad\qquad i = 1, \ldots, n, \, j = 1, \ldots, m. \tag{23}$$

The next proposition establishes the following connection between d_{AV} and d_K.

Proposition 3.1. *Let distribution F be characterized by outcomes $\mathbf{x}$ and their probabilities $\mathbf{p}$. The distribution G is characterized by outcomes $\mathbf{y}$ and their probabilities $\mathbf{q}$. Let $\mathcal{A} = [t_1, t_s]$, where $t_1 = \min\{\mathbf{x} \cup \mathbf{y}\}$ and $t_s = \max\{\mathbf{x} \cup \mathbf{y}\}$. Then, the cost of an optimal probability mass transportation plan, (20), equals the scaled average distance,*

$$d_K(F, G) = (t_s - t_1) d_{AV}^U(F, G). \tag{24}$$

Proof. As shown in Vallander (1973),

$$d_K(F, G) = \int_{\mathbb{R}} |F(z) - G(z)| dz = \int_{t_1}^{t_s} |F(z) - G(z)| dz.$$

Thus,

$$\int_{t_1}^{t_s} |F(z) - G(z)| dz = \sum_{i=1}^{s-1} d_k(t_{k+1} - t_k) = (t_s - t_1) d_{AV}(F, G).$$

$\square$

4. Approximation of Discrete Distributions: Distance Minimization Problem

In this section we define the approximation problem. We assume there exists a known reference distribution G with m outcomes, characterized by $(\mathbf{y}, \mathbf{q})$, and the goal is to define an approximation F, possibly of (much) smaller size. In this section we assume that the outcomes of the approximation, i.e., vector $\mathbf{x}$, is known, while their probabilities $\mathbf{p}$ are unknown. The vector $\mathbf{p}$ will be defined as any optimal solution of the following distance minimization problem:

$$\min_{\mathbf{p}} \; d_\alpha^U (F, G) \,, \tag{25}$$

$$\sum_{i=1}^n p_i = 1 \,, \tag{26}$$

$$p_i \geq 0 \,, \qquad\qquad i = 1, \ldots, n \,. \tag{27}$$

The following proposition is important in obtaining the convexity property of the problem $(25) - (27)$.

Proposition 4.1. *Let $\mathcal{R}$ be a coherent risk measure. Then, $d^H(F,G)$ is a convex functional, i.e., with $\lambda \in (0,1)$, for any maximal monotone relations F, G, $\widetilde{F}$, $\widetilde{G}$ on $\mathcal{A}$, the following holds:*

$$d^H(\lambda F + (1-\lambda)\widetilde{F}, \lambda G + (1-\lambda)\widetilde{G}) \leq \lambda d^H(F,G) + (1-\lambda)d^H(\widetilde{F}, \widetilde{G})\,. \tag{28}$$

Proof. Note that since $F(\cdot)$ associated with any maximal monotone relation F is a nondecreasing function on $\mathcal{A}$, then the convex combination $\lambda F(\cdot) + (1-\lambda)\widetilde{F}(\cdot)$ is also a nondecreasing function on $\mathcal{A}$, which is why $d^H(\lambda F + (1-\lambda)\widetilde{F}, \lambda G + (1-\lambda)\widetilde{G})$ is correctly defined. Using A2, A3, A4 axioms of coherent risk measures, we obtain:

$$d^H(\lambda F + (1-\lambda)\widetilde{F}, \lambda G + (1-\lambda)\widetilde{G}) = \tag{29}$$

$$\mathcal{R}(|\lambda F(\xi) + (1-\lambda)\widetilde{F}(\xi) - \lambda G(\xi) - (1-\lambda)\widetilde{G}(\xi)|) = \tag{30}$$

$$\mathcal{R}(|\lambda(F(\xi) - G(\xi)) + (1-\lambda)(\widetilde{F}(\xi) - (1-\lambda)\widetilde{G}(\xi))|) \overset{\text{A2}}{\leq} \tag{31}$$

$$\mathcal{R}(|\lambda(F(\xi) - G(\xi))| + |(1-\lambda)(\widetilde{F}(\xi) - (1-\lambda)\widetilde{G}(\xi))|) \overset{\text{A3}}{\leq} \tag{32}$$

$$\mathcal{R}(|\lambda(F(\xi) - G(\xi))|) + \mathcal{R}(|(1-\lambda)(\widetilde{F}(\xi) - (1-\lambda)\widetilde{G}(\xi))|) \overset{\text{A4}}{=} \tag{33}$$

$$\lambda\mathcal{R}(|F(\xi) - G(\xi)|) + (1-\lambda)\mathcal{R}(|\widetilde{F}(\xi) - \widetilde{G}(\xi)|) = \tag{34}$$

$$\lambda d^H(F,G) + (1-\lambda)d^H(\widetilde{F}, \widetilde{G})\,. \tag{35}$$

$$\square$$

Proposition 4.1 has the following important corollary.

Corollary 4.1. *Let F be a discrete distribution on $\mathbf{x}$ with probabilities $\mathbf{p}$; G is a distribution on $\mathbf{y}$ with probabilities $\mathbf{q}$. Then $d_\alpha^U(F,G)$ is a convex function of variables $(\mathbf{p}, \mathbf{q})$.*

Proof. Let $\mathbf{p} = (p_1, \ldots, p_n)$ and $\widetilde{\mathbf{p}} = (\widetilde{p}_1, \ldots, \widetilde{p}_n)$ be two arbitrary distributions on $\mathbf{x} = (x_1, \ldots, x_n)$. Let $\mathbf{q} = (q_1, \ldots, q_m)$ and $\widetilde{\mathbf{q}} = (\widetilde{q}_1, \ldots, \widetilde{q}_m)$ be two arbitrary distributions on $\mathbf{y} = (y_1, \ldots, y_m)$. With $\lambda \in (0, 1)$

$$F_\lambda(t) = \sum_{i=1}^{n} (\lambda p_i + (1 - \lambda)\widetilde{p}_i)\mathbb{1}_{t \geq x_i} = \lambda F(t) + (1 - \lambda)\widetilde{F}(t), \tag{36}$$

$$G_\lambda(t) = \sum_{j=1}^{m} (\lambda q_j + (1 - \lambda)\widetilde{q}_j)\mathbb{1}_{t \geq y_j} = \lambda G(t) + (1 - \lambda)\widetilde{G}(t). \tag{37}$$

Then,

$$d_\alpha^U(\lambda F + (1 - \lambda)\widetilde{F}, \, \lambda G + (1 - \lambda)\widetilde{G}) \leq \lambda d_\alpha^U(F, \, G) + (1 - \lambda)d_\alpha^U(\widetilde{F}, \, \widetilde{G}). \tag{38}$$

$\square$

As the result of Corollary 4.1, problem $(25) - (27)$ is a convex problem in variables $\mathbf{p} = (p_1, \ldots, p_n)$. Since linearization of the problem $(25) - (27)$ for various levels α can be done using standard approaches, we place them in Appendix.

5. CVaR Distance Minimization With Cardinality Constraint

Defining the problem $(25) - (27)$, we assumed that the outcomes of the approximation $\mathbf{x}$ were known, which rises the question how to set $\mathbf{x}$. One way to deal with that is to assume the set of outcomes $\mathbf{x}$ to be a subset of the outcomes of the target distribution $\mathbf{y}$ with a specified cardinality. Consider now the same $(25) - (27)$ problem with the restriction that only at most r out of m atoms of approximating distribution G will be used. In other words, the number of outcomes with positive probabilities is at most r. This problem can be formulated as the following combinatorial optimization problem:

$$\min_{\mathbf{p}} \; d_\alpha^U(F, G) \tag{39}$$

subject to

$$\sum_{i=1}^{m} r_i \leq r, \tag{40}$$

$$\sum_{i=1}^{m} p_i = 1, \tag{41}$$

$$p_i \leq r_i, \qquad\qquad i = 1, \ldots, m, \tag{42}$$

$$p_i \geq 0, \qquad\qquad i = 1, \ldots, m, \tag{43}$$

$$r_i \in \{0, 1\}, \qquad\qquad i = 1, \ldots, m. \tag{44}$$

The complete linear mixed integer formulation of this problem can be found in Appendix.

6. CVaR Function

Earlier we studied the definition of the CVaR distance applied to the special case of discrete distributions with finite domains. The goal is to set up the optimization problem that approximates one distribution by another distribution with known outcomes and unknown probabilities. The unknown probabilities are defined as an optimal solution to a distance minimization problem.

As noted in Mason and Schuenemeyer (1983), the Kolmogorov-Smirnov distance is often insensitive to differences between distributions in tails. For higher values of α, this can be also true for the CVaR distance as well. However, dealing with approximations of distributions, it is desirable to ensure that tails of an approximation distribution are as heavy as the tails of the original distribution. For instance, this can be ensured through imposing a set of constraints that CVaR of approximating distribution is at least as large as the CVaR of the original distribution for several confidence levels.

Consider a discrete distribution G on the set of outcomes $\mathbf{y} = (y_1, \ldots, y_m)$ with probabilities $\mathbf{q} = (q_1, \ldots, q_m)$. Let an approximating distribution F be located at outcomes $\mathbf{x} = (x_1, \ldots, x_n)$ with unknown probabilities $\mathbf{p} = (p_1, \ldots, p_n)$. Let α be the confidence level of interest. We would like to impose constraints on the CVaR of a random variable X having the distribution F to be at least as great as the CVaR of a random variable Y having the distribution G:

$$\mathrm{CVaR}_\alpha(X) \geq \mathrm{CVaR}_\alpha(Y) \,. \tag{45}$$

The above expression constraints the right tail of the approximating distribution. However, the left tail of the distribution can be of the same interest, thus leading us to the following constraint:

$$- \mathrm{CVaR}_\alpha(-X) \leq -\mathrm{CVaR}_\alpha(-Y) \,. \tag{46}$$

Since distribution G is known, then $\mathrm{CVaR}_\alpha(Y)$ is just a constant for every α, then it is clear that both (46) and (45) are the constraints of the same type,

$$\mathrm{CVaR}_\alpha(X) \geq a \,. \tag{47}$$

Next, we study the behaviour of these constraints and make sure they are well-defined.

Definition 7. *Consider the following function $f_\alpha(\cdot)$ of variables $\mathbf{p}$*

$$f_\alpha(\mathbf{p}) = \min_c \left(c + \frac{1}{1-\alpha} \sum_{i=1}^n p_i \big[x_i - c\big]^+ \right), \tag{48}$$

where

$$\big[x_i - c\big]^+ = \begin{cases} x_i - c \,, & \text{if } x_i \geq c \,, \\ 0 \,, & \text{otherwise} \,. \end{cases}$$

In other words, the function $f_\alpha(\mathbf{p})$ defines the CVaR with confidence level α of a random variable with the distribution F and so the constraint (47) will be expressed using notation $f_\alpha(\cdot)$ as

$$f_\alpha(\mathbf{p}) \geq a \,. \tag{49}$$

The value of $f_\alpha(\mathbf{p})$ can be obtained using one of the following representations:

1. As a solution of the primal problem (48):

$$\min_{c,\, z_i} \left(c + \frac{1}{1-\alpha} \sum_{i=1}^{n} p_i z_i \right) \tag{50}$$

subject to

$$z_i \geq x_i - c \,, \qquad\qquad\qquad i = 1, \ldots, n\,, \tag{51}$$
$$z_i \geq 0 \,, \qquad\qquad\qquad i = 1, \ldots, n\,. \tag{52}$$

2. As a solution of the problem dual to (48):

$$\max_{w_i} \ \sum_{i=1}^{n} w_i x_i \tag{53}$$

subject to

$$w_i \leq \frac{p_i}{1-\alpha} \,, \qquad\qquad\qquad i = 1, \ldots, n\,, \tag{54}$$

$$\sum_{i=1}^{n} w_i = 1 \,, \tag{55}$$

$$w_i \geq 0 \,, \qquad\qquad\qquad i = 1, \ldots, n\,. \tag{56}$$

In order to be used further in optimization problems, we need to know the properties of $f_\alpha(\mathbf{p})$. The following proposition establishes the concavity property of $f_\alpha(\mathbf{p})$. This property also follows from the general result about convexity of the pointwise supremum of an arbitrary collection of convex functions, e.g., Theorem 5.5 in Rockafellar (1970), however, here we can show the statement directly.

Proposition 6.1. $f_\alpha(\mathbf{p})$ *is a concave function.*

Proof. Let $\mathbf{p} = (p_1, \ldots, p_n)$ and $\widetilde{\mathbf{p}} = (\widetilde{p}_1, \ldots, \widetilde{p}_n)$. With (48) and $\lambda \in (0,1)$ we obtain

$$f_\alpha\left(\lambda \mathbf{p} + (1-\lambda)\widetilde{\mathbf{p}}\right) = \min_c \left(c + \frac{1}{1-\alpha} \sum_{i=1}^{n} (\lambda p_i + (1-\lambda)\widetilde{p}_i)\, [x_i - c]^+ \right) =$$

$$\min_c \left(\lambda c + \frac{\lambda}{1-\alpha} \sum_{i=1}^{n} p_i [x_i - c]^+ + (1-\lambda)c + \frac{1-\lambda}{1-\alpha} \sum_{i=1}^{n} \widetilde{p}_i [x_i - c]^+ \right) \geq$$

$$\min_c \left(\lambda c + \frac{\lambda}{1-\alpha} \sum_{i=1}^{n} p_i [x_i - c]^+ \right) + \min_c \left((1-\lambda)c + \frac{1-\lambda}{1-\alpha} \sum_{i=1}^{n} \widetilde{p}_i [x_i - c]^+ \right) =$$

$$\lambda f_\alpha(\mathbf{p}) + (1-\lambda) f_\alpha(\widetilde{\mathbf{p}}) \,.$$

$\square$

Corollary 6.1. *The set* $\{\mathbf{p} : f_\alpha(\mathbf{p}) \geq a\}$ *is convex.*

With representation (53) - (56) and Corollary 6.1 we obtain that

- Constraint (49) is a convex constraint

- It can be can be formulated using the following linear representation:

$$\sum_{i=1}^{n} w_i x_i \geq a \tag{57}$$

$$w_i \leq \frac{p_i}{1-\alpha} \, , \qquad\qquad i = 1, \ldots, n \, , \tag{58}$$

$$\sum_{i=1}^{n} w_i = 1 \, , \tag{59}$$

$$w_i \geq 0 \, , \qquad\qquad i = 1, \ldots, n \, . \tag{60}$$

Finally, we note that multiple constraints for various confidence levels α can be added to the distance minimization optimization problems described above, both to constrain the right tail and the left tail as well through the expression (46).

7. CVaR Distance Between Quantile Functions

Previously, the definition of the CVaR distance between distributions was introduced, based on their cumulative distribution funtions. As noted earlier, c.d. function is a classical source for measuring differences between distributions in many ways, and the maximum absolute difference value (the Kolmogorov-Smirnov distance) is widely used. It is reported, e.g., Mason and Schuenemeyer (1983), that the maximum value of the absolute difference between two cdfs is poorly sensitive to differences in tails of the distributions and frequently is attained at levels around the mean of the distribution. In the context of financial risk management where distributions describe losses of financial portfolios, tails represent extreme events, which are crucially important for measurement and management.

This section defines the CVaR distance between quantile functions. The quantile function of a probability distribution with cdf $F(\cdot)$ is defined as follows:

$$F^{-1}(z) = \inf \{l \in \mathbb{R} : F(l) \geq z\} \, . \tag{61}$$

In financial risk management, the quantity $F^{-1}(z)$ is known as the Value-at-Risk (VaR) with confidence level z of the probability distribution of losses. Therefore, $F^{-1}(z) - G^{-1}(z)$ represents the difference between the VaRs of two portfolios. The maximum absolute difference $\sup_z |F^{-1}(z) - G^{-1}(z)|$ can be introduced as another way of measuring the distance between distributions, however, Rachev et al. (2008) report that this maximum is usually attained at values of z close to 0 or 1. Thus, the CVaR distance based on quantile funtions will be based on extreme events, tails of distributions, and hence will suplement corresponding distances based on cumulative distribution functions.

The notation F^{-1} is used to point out the fact that a quantile function is, in some sense, the inverse function to a cumulative distribution function $F(\cdot)$. Thus, $F^{-1}(\cdot)$ is also a nondecreasing function on $\mathcal{A} = [0, 1]$. A quantile function, possibly after filling the vertical gaps, is another example of maximum monotone relations, see Rockafellar and Royset (2014), therefore previously developed theory is directly applicable in this case. The following defines the CVaR distance between quantile functions F^{-1} and G^{-1}.

Definition 8. *Let ξ be a uniform random variable on $[0,1]$. Then*

$$d_\alpha^U \left(F^{-1}, G^{-1} \right) = \langle\!\langle F^{-1}(\xi) - G^{-1}(\xi) \rangle\!\rangle_\alpha \,. \tag{62}$$

where $\langle\!\langle \cdot \rangle\!\rangle_\alpha$ denotes the CVaR norm, Mafusalov and Uryasev (2016).

Now we will simplify Definition 8 for the case of discrete distributions. If $F(\cdot)$ is the cdf of a discrete distribution, the corresponding quantile function will be expressed as

$$F^{-1}(z) = \inf \left\{ l \in \mathbb{R} : \sum_{i=1}^n p_i \mathbb{1}_{l \geq x_i} \geq z \right\}. \tag{63}$$

The above expression for the quantile function is somewhat complicated since it involves an optimization problem and the variable l in the argument of indicator function. In order to simplify it, let us use some features of the problem we are considering. Recall that we are given the probability distribution $\mathbf{p} = (p_1, \ldots, p_n)$ on the set of outcomes $\mathbf{x}$, where one can assume without loss of generality that the set of outcomes is ordered:

$$x_1 \leq \ldots \leq x_n \,.$$

Let us also define the cumulative distribution vector of the distribution F:

$$\mathbf{f} = (f_1, \ldots, f_n) = \left(p_1, p_1 + p_2, \ldots, \sum_{i=1}^n p_i \right). \tag{64}$$

With these notations, the quantile F^{-1} function can be equivalently defined in the following way:

$$F^{-1}(z) = x_{i_z} \,, \tag{65}$$
$$i_z = \min_i i \, : \, f_i \geq z \,, \ 0 \leq z \leq 1 \,. \tag{66}$$

Introducing the same notations as (64) and (65) – (66) for distribution G and letting $\{\gamma_1, \ldots, \gamma_s\} = \left\{ p_1, p_1 + p_2, \ldots, \sum_{i=1}^n p_i \right\} \cup \left\{ q_1, q_1 + q_2, \ldots, \sum_{j=1}^m q_j \right\}$, the random variable $F^{-1}(\xi) - G^{-1}(\xi)$ takes values

$$f_k^{-1} - g_k^{-1} = x_{i_{\gamma_k}} - y_{j_{\gamma_k}} \,, \qquad\qquad k = 1, \ldots, s-1 \,, \tag{67}$$
$$i_{\gamma_k} = \min_i i \, : \, f_i \geq \gamma_k \,, \qquad\qquad k = 1, \ldots, s-1 \,, \tag{68}$$
$$j_{\gamma_k} = \min_j j \, : \, g_j \geq \gamma_k \,, \qquad\qquad k = 1, \ldots, s-1 \,, \tag{69}$$

with probabilities

$$\mathbb{P} \left(f_k^{-1} - g_k^{-1} \right) = \gamma_{k+1} - \gamma_k \,, \qquad\qquad k = 1, \ldots, s-1 \,.$$

The definition of CVaR distance between quantile functions is illustrated in Figure 3. Previous sections defined the CVaR distance between two cumulative distribution functions and concerned the problem of obtaining optimal probabilities $\mathbf{p} = (p_1, \ldots, p_n)$ of outcomes of the approximation F, given that its outcomes $\mathbf{x} = (x_1, \ldots, x_n)$ were known. This definition is used to address the following problem. Suppose the probability distribution with m outcomes G is known (as well as function G^{-1}), and we would like to approximate it by a distribution F with n outcomes, for which we assume probabilities to be known. For instance, the approximating distribution might be set to belong to the class of uniform probability distributions, i.e., p_i can be set to $\dfrac{1}{n}$, $i = 1, \ldots, n$. The problem of finding the optimal positions $(x_1, \ldots, x_n)$ can be formulated as follows:

$$\min_{x_i} \quad d_\alpha^U \left(F^{-1}, G^{-1} \right) \tag{70}$$

subject to

$$x_1 \leq \ldots \leq x_n \, . \tag{71}$$

Convexity of the problem $(70) - (71)$ follows as another corollary of Proposition 4.1.

Corollary 7.1. *Let F be a discrete distribution on $\mathbf{x}$ with probabilities $\mathbf{p}$; G is a distribution on $\mathbf{y}$ with probabilities $\mathbf{q}$. Then $d_\alpha^U \left(F^{-1}, G^{-1} \right)$ is a convex function of variables $(\mathbf{x}, \mathbf{y})$.*

Proof. Let F be a discrete probability distribution with ordered outcomes $\mathbf{x} = (x_1, \ldots, x_n)$ and corresponding probabilities $\mathbf{p} = (p_1, \ldots, p_n)$. Let $\widetilde{F}$ be a discrete probability with ordered outcomes $\widetilde{\mathbf{x}} = (\widetilde{x}_1, \ldots, \widetilde{x}_n)$ and the same corresponding probabilities $\mathbf{p} = (p_1, \ldots, p_n)$. Similarly, G and $\widetilde{G}$ are two distributions on ordered $\mathbf{y} = (y_1, \ldots, y_m)$ and $\widetilde{\mathbf{y}} = (\widetilde{y}_1, \ldots, \widetilde{y}_m)$ having the same probabilities $\mathbf{q} = (q_1, \ldots, q_m)$. With $\lambda \in (0, 1)$,

$$F_\lambda^{-1}(z) = \inf \left\{ l \in \mathbb{R} : \sum_{i=1}^{n} p_i \mathbb{1}_{l \geq \lambda x_i + (1-\lambda)\widetilde{x}_i} \geq z \right\} = \lambda x_{i_z} + (1-\lambda)\widetilde{x}_{i_z} =$$

$$\lambda F^{-1}(z) + (1-\lambda)\widetilde{F}^{-1}(z) \, , \tag{72}$$

$$G_\lambda^{-1}(z) = \inf \left\{ l \in \mathbb{R} : \sum_{j=1}^{m} q_j \mathbb{1}_{l \geq \lambda y_j + (1-\lambda)\widetilde{y}_j} \geq z \right\} = \lambda y_{j_z} + (1-\lambda)\widetilde{y}_{j_z} =$$

$$\lambda G^{-1}(z) + (1-\lambda)\widetilde{G}^{-1}(z) \, . \tag{73}$$

Then, by Proposition 4.1

$$d_\alpha^U \left(\lambda F^{-1} + (1-\lambda)\widetilde{F}^{-1}, \lambda G^{-1} + (1-\lambda)\widetilde{G}^{-1} \right) \leq \lambda d_\alpha^U \left(F^{-1}, G^{-1} \right) + (1-\lambda) d_\alpha^U \left(\widetilde{F}^{-1}, \widetilde{G}^{-1} \right) .$$

$$\square$$

With representation $(67) - (69)$ of $F^{-1}(\xi) - G^{-1}(\xi)$, the linearization of the problem $(70) - (71)$ is straight forward and is moved to Appendix.
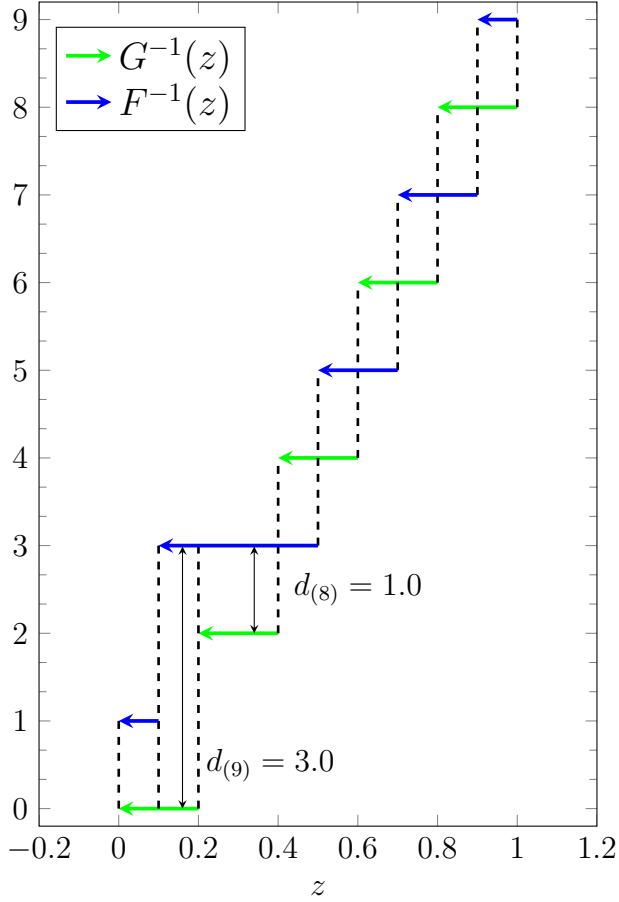
Figure 3: Illustration of the CVaR distance between quantile functions with confidence level $\alpha = \frac{7}{10}$. The largest absolute difference between two quantile functions, $d_{(9)} = 3.0$, has the probability of occurrence $\frac{1}{10}$. The second largest absolute difference, $d_{(8)} = 1.0$ and also has the probability of occurrence $\frac{2}{10}$. Thus, $d_{7/10}\left(F^{-1}, G^{-1}\right) = \frac{3.0 \cdot 0.1 + 1.0 \cdot 0.2}{0.3} \approx 1.67$.

8. Minimization of the Kantorovich-Rubinstein Distance

It is easy to notice that $d_\alpha^U(F, G)$ and $d_\alpha^U\left(F^{-1}, G^{-1}\right)$ distances are equal (up to a pre-defined scaling coefficient $(t_s - t_1)$) to each other when $\alpha = 0$. At the same time, the values of these distances correspond to the area between two cdf curves, and thus to the Kantorovich-Rubinstein distance. Hence, the above described two approaches of adjusting probabilities and assigning positions of outcomes can be combined in an iterative procedure that will adjust probabilities on one step, then use the obtained probabilities to adjust outcomes and so on.

Let G be a known distribution characterized by $(\mathbf{y}, \mathbf{q})$. The approximation F is required to have n outcomes $\mathbf{x}$ with probabilities $\mathbf{p}$. Formally, the iterative optimization procedure to find both $\mathbf{x}$ and $\mathbf{p}$ can be outlined as follows:

- Step 1. Let $\mathbf{p}^0 = \left(\dfrac{1}{n}, \dfrac{1}{n}, \ldots, \dfrac{1}{n}\right)$ be the uniform initial distribution. Find an optimal

17

$\mathbf{x}^1 = (x_1^1, \ldots, x_n^1)$ by solving the following quantile distance minimization problem:

$$\min_{x_i} \; d_0^U(F^{-1}(\mathbf{p}^0, x_1, \ldots, x_n), G^{-1}) \tag{74}$$

- Step 2. Find an optimal $\mathbf{p}^1 = (p_1^1, \ldots, p_n^1)$ by solving the following cdf distance minimization problem:

$$\min_{p_i} \; d_0^U(F(p_1, \ldots, p_n, \mathbf{x}^1), G) \tag{75}$$

- Repeat Steps 1-2 until the iterative procedure converges, i.e., for a small predefined $\epsilon > 0$, the following holds:

$$\left| d_0^U\left(F^{-1}\left(\mathbf{p}^k, \mathbf{x}^{k+1}\right), G^{-1}\right) - |\mathcal{A}| \, d_0^U\left(F\left(\mathbf{p}^{k+1}, \mathbf{x}^{k+1}\right), G\right)\right| < \epsilon. \tag{76}$$

This iterative procedure provides a series of approximations with nonincreasing values of the Kantorovich-Rubinstein distance. The procedure generates a sequence of nonincreasing numbers (distances) bounded from below (for instance, by 0), therefore the process necessary converges. We would like to note that the procedure does not necessary provide the global optimum for the Kantorovich-Rubinstein distance minimization problem of approximation of one discrete distribution by another one. Note that at the same time, there exists a simple algorithm that finds a minimum (Kantorovich-Rubinstein) distance approximation to a continuous distribution by a discrete one, see Kennan (2006).

9. Computational Experiments

The goal of computational experiments is to illustrate the computational efficiency of the distance minimization problems described in the previous sections. The target probability distribution G in our experiments is based on a real-life data set from the aerospace industry, describing the errors in the locations of fastener holes. The original dataset contains $8,165$ observations, with only 448 unique values thus having different probabilities. The dataset is further referred to as "Holes" dataset. Distance minimization problems were solved using two different solvers, Xpress (2014) software and AORDA (2014). For fair comparison, both solvers were set up to work using 1 thread. Computational results were obtained on a machine equipped with Windows 8.1x64 operating system, Intel Core(TM) i5-4200M CPU 2.5GHz, 6GB RAM.

Table 1 presents computational results of the minimization problem (25) – (27) for a wide range of values of parameter α solved with linear (Xpress (2014)) and convex (AORDA (2014)) solvers. The problem (25) – (27) requires setting the outcomes $\mathbf{x}$ of the approximating probability distribution, which were selected uniformly across the range of the original distribution G. Similarly, Table 2 presents the computational results of approximating the original probability distribution by a uniform discrete distribution via the quantile distance minimization.

Table 3 presents computational results for the cardinality constrained version of the (25) – (27) problem. It is a challenging optimization problem, even for moderately sized instances. Mainly, this is due to the poor quality of the linear relaxation of the corresponding

18

Dataset	m	n	α	Xpress		PSG	
				Objective	CPU Time (sec)	Objective	CPU Time (sec)
			0.0	0.00253	0.14	0.00253	0.32
			0.1	0.00281	0.17	0.00281	0.37
			0.2	0.00316	0.17	0.00316	0.35
			0.3	0.00361	0.16	0.00361	0.38
			0.4	0.00419	0.19	0.00419	0.29
Holes	448	100	0.5	0.00501	0.16	0.00501	0.29
			0.6	0.00619	0.18	0.00619	0.21
			0.7	0.00802	0.17	0.00802	0.14
			0.8	0.01056	0.15	0.01056	0.14
			0.9	0.01416	0.16	0.01416	0.11
			1.0	0.02995	0.06	0.02995	0.01

Table 1: Computational results of solving (25) – (27) problem.

Dataset	m	n	α	Xpress	
				Objective	CPU Time (sec)
			0.0	0.00058	0.03
			0.1	0.00065	0.04
			0.2	0.00073	0.04
			0.3	0.00083	0.03
			0.4	0.00097	0.03
Holes	448	100	0.5	0.00116	0.04
			0.6	0.00137	0.03
			0.7	0.00166	0.04
			0.8	0.00224	0.03
			0.9	0.00399	0.04
			1.0	0.03300	0.02

Table 2: Computational results of solving (70) – (71) problem.

$0-1$ problem formulation. Because of that, we compare the best objective values obtained by either of the solvers within a specified time limit. The optimality gaps for the considered instances and time limits are large, up to 95%, which suggests that both solvers essentially run as heuristics with little or no performance guarantee.

Finally, computational experiments of the iterative procedure for the Kantorovich-Rubinstein distance minimization are presented in Table 4. Experiments show that only few iterations are needed to achieve convergence of the procedure with a small, e.g., $\epsilon = 10^{-5}$, precision. Figure 4 provides an illustration of the original dataset by 10 discrete points using the iterative procedure. It is worth to note that contrary to the result of Kennan (2006), which states that the minimum Kantorovich-Rubinstein distance approximation of the continuous distribution is necessary uniform, the output of the procedure is not a uniform discrete distribution. The illustration also demontrates that the right tail of the original distruibution is heavier than one of the obtained approximation. Therefore, corrections in form of CVaR constraints discussed in Section 6 can be particularly helpful in situations, where matching the tail is of primal importance. Results of a more detailed case

Dataset	m	n	r	α	Xpress		PSG	
					Objective	CPU Time (sec)	Objective	CPU Time (sec)
				0.0	0.0130	954.15	0.0140	955.58
				0.1	0.0156	739.80	0.0155	740.31
				0.2	0.0176	911.73	0.0183	912.12
				0.3	0.0197	936.78	0.0211	937.31
				0.4	0.0213	680.77	0.0237	681.04
Holes	448	448	10	0.5	0.0264	1,504.77	0.0248	1,505.64
				0.6	0.0304	1,805.75	0.0278	1,806.03
				0.7	0.0346	1,244.74	0.0330	1,245.04
				0.8	0.0414	1,059.78	0.0483	1,060.46
				0.9	0.0494	1,997.78	0.0525	1,998.47
				1.0	0.0672	631.76	0.0763	632.42
				0.0	0.0071	817.59	0.0078	818.21
				0.1	0.0081	580.76	0.0087	581.18
				0.2	0.0092	1,144.8	0.0106	1,145.61
				0.3	0.0103	1,698.75	0.0117	1,699.10
				0.4	0.0118	2,194.75	0.0111	2,195.01
Holes	448	448	20	0.5	0.0138	1,776.74	0.0150	1,777.31
				0.6	0.0160	2,668.73	0.0150	2,669.80
				0.7	0.0187	2,498.7	0.0170	2,499.81
				0.8	0.0230	2,100.74	0.0242	2,101.60
				0.9	0.0247	1,524.74	0.0305	1,525.67
				1.0	0.0350	590.77	0.0360	591.79

Table 3: Computational experiments solving the cardinality constrained approximation problem (39) – (43). The optimization problems were first run with AORDA software package. Solution time obtained by AORDA solver was used as an upper time limit for the FICO Xpress solver and the corresponding potentially suboptimal objective values are presented in the table.

Dataset	m	n	Xpress		
			Objective	CPU Time (sec)	# of iterations
		5	0.00551	0.91	5
		10	0.00321	1.71	7
		20	0.00186	3.06	9
		30	0.00125	4.00	10
		40	0.00106	1.37	5
Holes	448	50	0.00079	3.92	9
		60	0.00066	2.72	7
		70	0.00049	6.55	11
		80	0.00045	4.87	9
		90	0.00038	9.98	13
		100	0.00034	10.53	13

Table 4: Approximation of the dataset via minimization of the Kantorovich-Rubinstein distance. Approximation is done both with respect to outcome positions and their probabilities.

study are posted online[2].

---

[2]http://www.ise.ufl.edu/uryasev/research/testproblems/advanced-statistics/
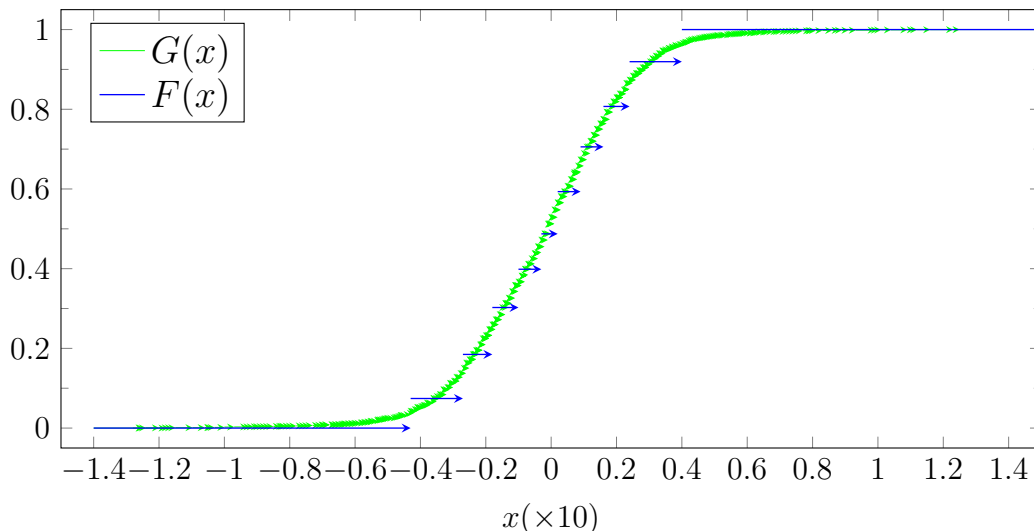approximation-discrete-distribution-using-cvar-distance/

Figure 4: Illusrtation of a result of the iterative Kantorovich-Rubinstein distance minimization procedure described in section 8. Approximation is done by a distribution with 10 outcomes.

10. Conclusion

This paper described several approaches to the problem of approximation of one discrete distribution on the line with finite support by another one, potentially with a smaller number of outcomes. The approximation problem clearly consists of two subproblems: where to locate the atoms of an approximation and which probabilities should be assigned to them. These problems are considered separately and it is shown that each of these problems can be approached using special distance measures between cumulative distribution functions or quantile functions. An important special case of these distances is the Kantorovich-Rubinstein distance, which is equal to the area between functions (cdf or quantile) of two distributions. This fact allows to combine subproblems into an iterative procedure to find an approximation via minimization of the Kantorovich-Rubinestein distance with respect to both atom locations and corresponding probabilities.

Some possible directions for future research may include the following. First, the cardinality constrained version of the approximation problem, where atom positions of the approximation are selected from the set of original atom positions, may require a stronger problem formulation. Extension of the ideas behind the iterative Kantorovich-Rubinstein distance minimization procedure in higher dimensions can also be of interest. Finally, introduced concept of the CVaR distance as an extension of the Kolmogorov-Smirnov distance leads to the question whether it can be applied for hypothesis testing of equality of distributions, similar to the Kolmogorov-Simrnov test.

11. References

AORDA, 2014. Portfolio Safeguard Version 2.1. `www.aorda.com/aod/welcome`.

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent Measures of Risk. Mathematical Finance 9 (3), 203–228.

Boos, D. D., 1981. Minimum Distance Estimators for Location and Goodness of Fit. Journal of the American Statistical Association 76 (375), 663–670.

Darling, D. A., 1957. The Kolmogorov-Smirnov, Cramer-von Mises Tests. The Annals of Mathematical Statistics, 823–838.

Feller, W., 1948. On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions. The Annals of Mathematical Statistics 19 (2), 177–189.

Gibbons, J. D., Chakraborti, S., 2011. Nonparametric Statistical Inference. Springer.

Grigoriu, M., 2009. Reduced Order Models for Random Functions. Application to Stochastic Problems. Applied Mathematical Modelling 33 (1), 161–175.

Hammond, R. K., Bickel, J. E., 2013. Reexamining Discrete Approximations to Continuous Distributions. Decision Analysis 10 (1), 6–25.

Keefer, D. L., 1994. Certainty Equivalents for Three-Point Discrete-Distribution Approximations. Management Science 40 (6), 760–773.

Keefer, D. L., Bodily, S. E., 1983. Three-Point Approximations for Continuous Random Variables. Management Science 29 (5), 595–609.

Kennan, J., 2006. A Note on Discrete Approximations of Continuous Distributions, University of Wisconsin-Madison.

Mafusalov, A., Uryasev, S., 2016. CVaR (Superquantile) Norm: Stochastic Case. European Journal of Operational Research 249 (1), 200–208.

Mason, D. M., Schuenemeyer, J. H., 1983. A Modified Kolmogorov-Smirnov Test Sensitive to Tail Alternatives. The Annals of Statistics, 933–946.

McDonald, J. N., Weiss, N. A., 1999. A Course in Real Analysis. Academic Press.

Miller III, A. C., Rice, T. R., 1983. Discrete Approximations of Probability Distributions. Management Science 29 (3), 352–362.

Pavlikov, K., Uryasev, S., 2014. CVaR Norm and Applications in Optimization. Optimization Letters 8 (7), 1999–2020.

Pflug, G. C., 2000. Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk. In: Probabilistic constrained optimization. Springer, pp. 272–281.

Rachev, S. T., Stoyanov, S. V., Fabozzi, F. J., 2008. Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measures. Vol. 149. John Wiley & Sons.

Rockafellar, R. T., 1970. Convex Analysis. No. 28. Princeton University Press.

Rockafellar, R. T., Royset, J. O., 2014. Random Variables, Monotone Relations, and Convex Analysis. Mathematical Programming 148 (1-2), 297–331.

Rockafellar, R. T., Uryasev, S., 2000. Optimization of Conditional Value-at-Risk. Journal of Risk 2 (3), 21–41.

Rockafellar, R. T., Uryasev, S., 2002. Conditional Value-at-Risk for General Loss Distributions. Journal of Banking & Finance 26 (7), 1443–1471.

Rockafellar, R. T., Uryasev, S., 2013. The Fundamental Risk Quadrangle in Risk Management, Optimization and Statistical Estimation. Surveys in Operations Research and Management Science 18 (1), 33–53.

Rosenblueth, E., Hong, H. P., 1987. Maximum Entropy and Discretization of Probability Distributions. Probabilistic Engineering Mechanics 2 (2), 58–63.

Rudin, W., 1964. Principles of Mathematical Analysis. Vol. 3. McGraw-Hill New York.

Smith, J. E., 1993. Moment Methods for Decision Analysis. Management Science 39 (3), 340–358.

Vallander, S. S., 1973. Calculation of the Wasserstein Distance between Probability Distributions on the Line. Teoriya Veroyatnostei i ee Primeneniya 18 (4), 784–786.

Villani, C., 2009. Optimal Transport: Old and New. Vol. 338. Springer.

Xpress, 2014. FICO$^{\text{TM}}$ Xpress Optimization Suite 7.8. http://www.fico.com.

Appendix A.

Formulation 1. *The problem (25) – (27) for $0 < \alpha < 1$ can be reformulated as the following linear problem:*

$$\min_{c,\, p_i,\, z_k} \left(c + \frac{1}{1-\alpha} \sum_{k=1}^{s-1} \mathbb{P}(d_k) z_k \right) \tag{A.1}$$

subject to

$$z_k \geq \quad F(t_k) - G(t_k) - c\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.2}$$

$$z_k \geq -F(t_k) + G(t_k) - c\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.3}$$

$$F(t_k) = \sum_{i=1}^{n} p_i \mathbb{1}_{t_k \geq x_i}\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.4}$$

$$\sum_{i=1}^{n} p_i = 1\,, \tag{A.5}$$

$$z_k \geq 0\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.6}$$

$$p_i \geq 0\,, \qquad\qquad i = 1, \ldots, n\,. \tag{A.7}$$

Formulation 2. *The problem (25) – (27) with $\alpha = 1$ can be reformulated as the following linear problem:*

$$\min_{a,\, p_i} \quad a \tag{A.8}$$

subject to

$$a \geq \quad F(t_k) - G(t_k)\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.9}$$

$$a \geq -F(t_k) + G(t_k)\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.10}$$

$$F(t_k) = \sum_{i=1}^{n} p_i \mathbb{1}_{t_k \geq x_i}\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.11}$$

$$(A.5)\,, (A.7)\,.$$

Formulation 3. *The problem (25) – (27) with $\alpha = 0$ can be reformulated as the following linear problem:*

$$\min_{p_i,\, d_k} \quad \sum_{k=1}^{s-1} d_k \mathbb{P}(d_k) \tag{A.12}$$

subject to

$$d_k \geq \quad F(t_k) - G(t_k)\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.13}$$

$$d_k \geq -F(t_k) + G(t_k)\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.14}$$

$$F(t_k) = \sum_{i=1}^{n} p_i \mathbb{1}_{t_k \geq x_i}\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.15}$$

$$(A.5)\,, (A.7)\,.$$

Formulation 4. *The problem (39) – (43) with $\alpha \in (0,1)$ can be reformulated as the following linear problem:*

$$\min_{c,\, p_i,\, r_i,\, z_k} \quad \left(c + \frac{1}{1-\alpha} \sum_{k=1}^{m-1} \mathbb{P}(d_k) z_k \right) \tag{A.16}$$

subject to

$$z_k \geq \quad F(y_k) - G(y_k) - c\,, \qquad\qquad k = 1, \ldots, m-1\,, \tag{A.17}$$

$$z_k \geq -F(y_k) + G(y_k) - c\,, \qquad\qquad k = 1, \ldots, m-1\,, \tag{A.18}$$

$$F(y_k) = \sum_{i=1}^{m} p_i \mathbb{1}_{y_k \geq y_i}\,, \qquad\qquad k = 1, \ldots, m-1\,, \tag{A.19}$$

$$p_i \leq r_i\,, \qquad\qquad i = 1, \ldots, m\,, \tag{A.20}$$

$$\sum_{i=1}^{m} r_i \leq r\,, \tag{A.21}$$

$$z_k \geq 0\,, \qquad\qquad k = 1, \ldots, m-1\,, \tag{A.22}$$

$$r_i \in \{0,1\}\,, \qquad\qquad i = 1, \ldots, m\,, \tag{A.23}$$

$$(A.5)\,, (A.7)\,.$$

Corresponding reformulations for problem (39) – (43) with $\alpha = 1$ and $\alpha = 0$ can be obtained similar to Formulations 2 and 3.

Formulation 5. *The problem (70) – (71) with $\alpha \in (0, 1)$ for the minimization of CVaR distance between quantile functions can be reformulated as the following linear problem:*

$$\min_{c,\, x_i,\, z_k} \left(c + \frac{1}{1 - \alpha} \sum_{k=1}^{s-1} \mathbb{P}(d_k) z_k \right) \tag{A.24}$$

subject to

$$z_k \geq \quad x_{i_{\gamma_k}} - y_{j_{\gamma_k}} - c\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.25}$$

$$z_k \geq -x_{i_{\gamma_k}} + y_{j_{\gamma_k}} - c\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.26}$$

$$z_k \geq 0\,, \qquad\qquad k = 1, \ldots, s-1\,, \tag{A.27}$$

$$x_1 \leq \ldots \leq x_n\,. \tag{A.28}$$

Corresponding reformulations for problem (70) – (71) with $\alpha = 1$ and $\alpha = 0$ can be obtained similar to Formulations 2 and 3.