Fréchet inequalities via convex optimization

Bart P.G. Van Parys[*1], Paul J. Goulart[†2], and Paul Embrechts[‡3]

[1]Massachusetts Institute of Technology
[2]Oxford University
[3]ETH Zürich

Abstract

Quantifying the risk carried by an aggregate position $S_d := \sum_{i=1}^{d} X_i$ comprising many risk factors X_i is fundamental to both insurance and financial risk management. Fréchet inequalities quantify the worst-case risk carried by the aggregate position given distributional information concerning its composing factors but without assuming independence. This marginal factor modeling of the aggregate position in terms of its risk factors X_i leaves, however, the distribution of S_d ambiguous. The resulting distributional ambiguity can be traced back directly to the lack of dependence structure imposed between the risk factors. Fréchet inequalities implicitly uphold already the work of George Boole on probabilistic logic but were explicitly derived by Maurice Fréchet only in 1935. These inequalities can be considered rules about how to bound calculations involving probabilities without assuming independence or, indeed, without making any dependence assumptions whatsoever. It is exactly this robustness to obscure risk dependencies which have stimulated their renewed interest in the post crisis economy. We will approach these Fréchet inequalities using a modern convex optimization lens. Our novel perspective provides data friendly computational tools applicable to practical Fréchet problems. Two important data friendly classes of Fréchet problems are presented, both of which are shown to admit exact tractable convex optimization reformulations. The efficacy of our approach is illustrated using a small insurance management problem.

1 Introduction

A fundamental problem in financial and insurance management is to quantify the risk of a financial position S_d. Most commonly, the position $S_d := \sum_{i=1}^{d} X_i$ is given as the aggregate of several risk factors $X := (X_1, \ldots, X_d)$ valued in $\mathbb{R}^d$ among which no dependence structure can be assumed a priori. For instance, the risk factors X_i might represent the uncertain returns of d different stock options in a portfolio S_d with known marginal return distributions $\mathbb{P}_i$. In practice the dependence structure between individual stock options is obscure at best or even completely unknown. This type of distributional information structure is justified in many practical situations as obtaining marginal information is often easier than obtaining exact dependence structures. Previous observation is especially pronounced in a high-dimensional setting $d \gg 1$. Information regarding the joint

[*]Sloan School of Management, USA, Email: vanparys@mit.edu. The first author was supported by the Early Post.Mobility fellowship No. 165226 of the Swiss NSF.
[†]Department of Engineering Science, Oxford University, UK, Email: paul.goulart@eng.ox.ac.uk
[‡]Department of Mathematics, ETH Zürich, Switzerland, Email: paul.embrechts@math.ethz.ch

distribution $\mathbb{P}$ of the aggregate S_d is thus limited to its marginal projections $\Pi_i \circ \mathbb{P}$ each coinciding with the distribution $\mathbb{P}_i$ of a factor X_i.

The risk associated with the position S_d must be quantified with the help of a user specified risk measure, c.f. McNeil et al. (2015). Ideally, the choice of measure reflects the risk preferences of the practitioner. The Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) are two of the most popular risk measures in financial and insurance management. The popularity of the VaR stems mainly from its simplicity and historical prevalence as reflected for instance by its promotion in the Basel II banking accords. See Danielsson et al. (2001) for an academic discussion on the (mis)use and of the VaR in financial regulation. Recently however, the widespread use of VaR has been put into question in favor of convex alternatives such as the CVaR measure. The results in this paper will not be specific to the VaR or CVaR measure but rather apply to other risk measures as well. We will do so by studying the expected loss

$$F(\ell, \mathbb{P}) := \int \ell\left(S_d := \textstyle\sum_{i=1}^d X_i\right) \, \mathrm{d}\mathbb{P} \tag{1}$$

of the aggregate position as measured by a positive loss function ℓ and assume for now that the loss function is given. Later we will show how both the VaR and CVaR are intimately related to our canonical form (1). Alternatively, the function ℓ can be chosen so as to reflect the risk preferences of the practitioner. Unfortunately, the risk $F(\ell, \mathbb{P})$ is as such beyond our grasp as the joint distribution $\mathbb{P}$ is not known.

The lack of dependence structure inherently results in ambiguity in terms of the joint probability distribution $\mathbb{P}$ between the risk factors. Indeed, in general a joint probability distribution is not uniquely determined by its marginals alone. Fréchet inequalities provide us the least upper bound on the risk carried by the aggregate position S_d given the marginal information $\mathbb{P}$ in $\mathcal{M}$, i.e.

$$\begin{aligned} F(\ell, \mathcal{M}) := \sup \quad & \int \ell\left(S_d := \textstyle\sum_{i=1}^d X_i\right) \, \mathrm{d}\mathbb{P} \\ \text{s.t.} \quad & \mathbb{P} \in \mathcal{M}. \end{aligned} \tag{P}$$

In that regard, the Fréchet bound (P) is recognized as the robust counterpart of the nominal risk in (1) with respect to the marginal model $\mathcal{M}$. The marginal model $\mathcal{M}$ hence should be taken as the smallest set of joint distributions consistent with the available information regarding the joint distribution between the risk factors.

Most results on Fréchet inequalities and risk management found for instance in the works Rüschendorf (1991); Puccetti and Rüschendorf (2013); Wang and Wang (2011) are concerned with complete marginal information. Complete marginal information corresponds to a marginal model $\mathcal{M}$ consisting of all joint distributions sharing given marginal distributions $\mathbb{P}_i$, i.e.

$$\mathcal{M}(\mathbb{P}_1, \ldots, \mathbb{P}_d) := \{\mathbb{P} \mid \Pi_i \circ \mathbb{P} = \mathbb{P}_i, \quad \forall i \in [d]\}. \tag{2}$$

This previous marginal model requires complete knowledge of the distributions $\mathbb{P}_i$ of the risk factors X_i and is not very data friendly. In practice indeed data rather than distributions are observed. As most historical data is both noisy and incomplete, some ambiguity concerning the marginal distributions $\mathbb{P}_i$ can never be avoided altogether. We will consider therefore incomplete marginal information as the more data friendly alternative. Partial marginal information can be represented through d marginal ambiguity sets $\mathcal{P}_i$. In this case the marginal model $\mathcal{M}$ corresponds to the set of joint distributions whose marginals are elements of the corresponding marginal ambiguity sets $\mathcal{P}_i$, i.e.

$$\mathcal{M}(\mathcal{P}_1, \ldots, \mathcal{P}_d) := \{\mathbb{P} \mid \Pi_i \circ \mathbb{P} \in \mathcal{P}_i, \quad \forall i \in [d]\}, \tag{3}$$

and makes for a more practical alternative to (2).

We will look at the Fréchet problem (P) from a convex optimization perspective. Indeed, when the marginal ambiguity sets $\mathcal{P}_i$ are convex then the marginal model (3) is a convex set as well. Remark however that the marginal model does not consists of finite dimensional vectors, but rather of potentially infinite dimensional distributions. This infinite optimization perspective will nevertheless allow us to leverage various powerful results from the theory of linear programming over sets of probability distributions. The practical power which this approach brings has been illustrated already by the results in Popescu (2005); Vandenberghe et al. (2007); Van Parys et al. (2016); Van Parys (2015). The main purpose of this paper is thus not only to offer new practical results, but also to argue that convex optimization can offer a novel valuable perspective on Fréchet problems. In doing so, we also will develop numerical tools applicable to the Fréchet problem (P) which are both exact and tractable in practice.

1.1 Background

Fréchet inequalities implicitly uphold already the work of Boole (1854) on probabilistic logic but were only explicitly derived by Fréchet (1935) almost a century later. Probabilistic logic concerns itself with extending classical binary logic to propositions which hold merely in probability. Suppose E_1 and E_2 are two logical propositions or events for which we know their respective probabilities $\mathbb{P}(E_1)$ and $\mathbb{P}(E_2)$. Nothing else is known. Despite not knowing the relationship between both events, we wish to characterize the probability of their logical disjunction $\mathbb{P}(E_1 \cap E_2)$ and conjunction $\mathbb{P}(E_1 \cup E_2)$. Evidently, these probabilities are ambiguous and hence can not be determined exactly. Nevertheless, characterizing exact upper (as well as lower) bounds can be cast as a Fréchet problem. The probability $\mathbb{P}(E_1 \cap E_2)$ for instance must be less than the Fréchet bound (P) for the risk factors $X_i = \frac{1}{2}\mathbb{1}_{E_i}$ with their respective marginal ambiguity sets $\mathcal{P}_i = \left\{\mathbb{Q} \;\middle|\; \int \mathbb{1}_{E_i} \,\mathrm{d}\mathbb{Q} = \mathbb{P}(E_i)\right\}$ and an indicator loss function $\ell = \mathbb{1}_{=1}$. Notice that the corresponding lower bound can be obtained in much the same manner by simply negating the loss function. These resulting four simple Fréchet bounds admit closed form solutions which we state here for the sake of completeness.

Fact 1 (Probabilistic logic)**.** *Given two events with probabilities $\mathbb{P}(E_1)$ and $\mathbb{P}(E_2)$, then their logical disjunction satisfies*

$$\max(0, \mathbb{P}(E_1) + \mathbb{P}(E_2) - 1) \le \mathbb{P}(E_1 \cap E_2) \le \min(\mathbb{P}(E_1), \mathbb{P}(E_2)),$$

while their logical conjunction satisfies

$$\max(\mathbb{P}(E_1), \mathbb{P}(E_2)) \le \mathbb{P}(E_1 \cup E_2) \le \min(1, \mathbb{P}(E_1) + \mathbb{P}(E_2)).$$

Please remark that if the probabilities of the events E_1 and E_2 are either zero or one, the previous fact reduces to the standard binary logic for AND and OR, respectively. These four inequalities can be considered rules about how to bound calculations involving probabilities without assuming independence or, indeed, without making any dependence assumptions whatsoever.

Bivariate $(d = 2)$ Fréchet problems are closely related to several classical problems in mathematics as noted by Rüschendorf (1991). Optimal transport problems discussed in Villani (2008) are indeed a superclass of bivariate Fréchet problems with complete marginal information. The bivariate joint distribution $\mathbb{P}$ is in this context usually referred to as the plan transporting its first marginal into its second. Analytical results for bivariate Fréchet problems are known dating back to the classical results by Kolmogorov and Makarov (1982). Furthermore, exact expressions are found in Puccetti and Rüschendorf (2013) for Fréchet problems in the context of VaR aggregation dealing with homogeneous marginals ($\mathbb{P}_1 = \mathbb{P}_2 = \cdots = \mathbb{P}_d$) enjoying stringent monotonicity properties. An impressive body of work by Wang and Wang (2011); Wang et al. (2013); Puccetti et al. (2013) is available concerning a deep connection of comonotonicity and the distributions for which Fréchet bounds hold with equality. Most of these results however apply only to very special cases and are consequently of limited practical value.

Due to their robustness against hard to come by dependence structures, Fréchet problems have witnessed a surge of interest from the financial management community in the context of the post crisis economy, c.f. Das et al. (2013) and references therein. Despite their practical relevance however, computational algorithms for the Fréchet problem (P) are scarce to come by. A notable exception in this regard is the `rearrangement method` presented in Puccetti and Rüschendorf (2012). The method is empirically found to scale well and often delivers high-quality solutions. Unfortunately, only approximate solutions can be hoped for. Furthermore, the method can not deal with ambiguity concerning the marginals despite the fact that in practice they invariable need to be estimated from noisy historical data.

1.2 Outline and contributions

We will approach in Section 2 Fréchet problems using a novel convex optimization lens yielding practical and data friendly computational tools. We will take the same angle to Fréchet problems as the one found in Bertsimas and Popescu (2005) and Owhadi et al. (2013). This angle looks at Fréchet problems from a strictly optimization point of view. When the marginal ambiguity sets $\mathcal{P}_i$ are convex, then the Fréchet problem (P) can indeed be recognized as a linear optimization problem albeit over distributions rather than vectors. Despite not (immediately) yielding any practical algorithm, this perspective is shown in Section 2.2 to reveal a powerful recursive structure underlying all Fréchet problem.

The popularity of Fréchet inequalities mainly derives from the observation that information concerning the marginals X_i is far easier to obtain from limited historical data than their dependence structure. Nevertheless, also the marginal distributions can not be expected to be uniquely determined from a finite amount of observations. Sections 3 and 4 discuss how our recursive structure underlying all Fréchet problems can be exploited to reduce the Fréchet problem (P) to a *tractable* convex optimization problem for two data friendly Fréchet classes:

(i) The marginal distributions $\mathbb{P}_i$ of the risk factors X_i have known means μ_i and variances σ_i^2. Both of these statistics can be determined with high accuracy using only a modest amount of historical data following Hoeffding (1963). The corresponding marginal ambiguity sets are

$$\mathcal{P}_{i,\mathcal{Q}} = \left\{ \mathbb{Q} \;\middle|\; \int t \,\mathrm{d}\mathbb{Q} = \mu_i \text{ and } \int t^2 \,\mathrm{d}\mathbb{Q} = s_i^2 := \mu_i^2 + \sigma_i^2 \right\}. \tag{4}$$

Theorem 2 establishes that the resulting Fréchet problem reduces to a tractable semidefinite optimization problem. This Fréchet problem with given marginal means and variances can be recognized as an instance of a generalized Chebyshev problem; c.f. Vandenberghe et al.

(2007); Van Parys et al. (2016). Our reformulation, however, presents a novel perspective on and yields a significant generalization of the results found in aforementioned works which all assume stringent conditions on the loss function ℓ.

(ii) The information concerning the marginal distributions $\mathbb{P}_i$ of the risk factors X_i comes in the form a histogram with b bins of fixed bin size Δ over an interval $[s_i, s_i+(b-1)\cdot\Delta]$ on the real line. These histograms give a rough sense of the marginal distributions and are well suited when working directly with empirically obtained data. Indeed, they were originally proposed by Pearson (1894) to do just that. The corresponding ambiguity sets are given as

$$\mathcal{P}_{i,\mathcal{H}} = \left\{ \mathbb{Q} \;\middle|\; \int \mathbb{1}_{\leq s_i+(j-1)\cdot\Delta} \, \mathrm{d}\mathbb{Q} = p_{i,j}, \quad \forall j \in [b] \right\}, \tag{5}$$

where $p_{i,j}$ represent the cumulative distribution functions of the marginals $\mathbb{P}_i$ in binned form. Theorem 3 establishes that the resulting Fréchet problem reduces to a tractable linear optimization problem. We remark that we have not been the first to consider Fréchet problems based on marginal histogram information. The `rearrangement method` discussed earlier goes about approximating Fréchet problems with complete information by first binning the known marginals $\mathbb{P}_i$ into histogram form. It then approximates the resulting Fréchet problem with histogram information using a shuffle heuristic. An exact linear programming approach taken by Prékopa (1988) could be extended to our setting as well, but would unfortunately yield a linear optimization characterization of exponential size in the dimension d. To the best of our knowledge, we are the first to obtain a polynomial sized exact tractable reformulation of Fréchet problems with histogram information.

Observe that both Fréchet classes are compatible with practical situations in which data rather than exact marginal distributions are given. Indeed, the means and variances as well as histograms are often the only information about the underlying distributions which can be determined with reasonable confidence based on limited historical data. We hope that the results presented in this paper will spur the research on this topic towards a more computational and quantitive direction prepared for a new age in which data rather than exact models take the center stage. The efficacy of our approach is illustrated using a small insurance management problem in Section 6.

Notation

We denote by $\mathbb{S}^n$ and $\mathbb{S}^n_+$ the sets of all symmetric and positive semidefinite symmetric matrices in $\mathbb{R}^{n\times n}$, respectively, and for any $X, Y \in \mathbb{S}^n$ the relation $X \succeq Y$ ($X \preceq Y$) indicates that $X - Y \in \mathbb{S}^n_+$ ($Y - X \in \mathbb{S}^n_+$). For functions $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$, their epi-addition is the function $f_1 \oplus f_2 : \mathbb{R}^n \to \mathbb{R}$ defined by $(f_1 \oplus f_2)(t) := \inf_s \{f_1(s) + f_2(t-s)\}$. The epi-addition is associative, commutative, and satisfies the well known identity

$$\inf \left\{ \textstyle\sum_{i=1}^n f_i(t_i) \;\mid\; \sum_{i=1}^n t_i = t \right\} = \left[\oplus_{i=1}^n f_i\right](t). \tag{6}$$

The epi-addition operation is referred to as inf-convolution as done for instance in Rockafellar (1970) and will come to play a preeminent role in this paper. Rockafellar shows that the epi-addition operator can be stated in terms of epigraphs as well. That is, the epigraph of $\oplus_i f_i$ coincides with the Minkowski sum of the epigraphs of the functions f_i. Further results on epi-addition can be found in (Rockafellar and Wets, 1998, §1.H). Finally, we denote with $\mathbb{1}_S$ the indicator function of a set S, i.e. $\mathbb{1}_S(x) = 1$ when $x \in S$; zero otherwise.

2 Fréchet problems

We state and briefly discuss here the properties particular to all Fréchet problems more in depth than was done in the introduction. The Fréchet problem (P) can be recognized as an optimization problem over a subset of the simplex of probability distributions. Taking this point of view has the benefit that we can bring to bear several powerful results from the theory of infinite dimensional programming discussed in for instance Shapiro and Kleywegt (2002); Bertsimas and Popescu (2005). Nevertheless, we will need to exploit the structure characteristic to Fréchet problems much further than done in those works if we are to develop computational tools. Indeed, optimizing over probability distributions is NP hard in general as shown by Bertsimas and Popescu (2005). We will focus primarily on the dual characterization possessed by all Fréchet problems. We will show how the lack of dependence structure in the primal formulation (P) translates to a powerful recursive formulation of its dual.

2.1 Dependence uncertainty

The key property common to all Fréchet problems is their complete lack of dependence information between the risk factors X_i. This dependence structure, or rather lack thereof, is justified from a practical point of view by the observation that obtaining marginal distributional information is often much more practical then justifying a dependence structure. Indeed, the problem of determining a dependence structure tends to suffer severely from the so-called *curse of dimensionality* discussed in Keogh and Mueen (2010). That is, the problem of estimating or justifying a dependence structure between the factors X_i tends to become problematic with increasing dimensionality d when only empirical data is given.

Despite the previous concerns, in the financial and insurance risk literature a dependence structure is nevertheless occasionally specified via copulas, c.f. Embrechts et al. (2003). Copula are often chosen based on expert knowledge of the particular problem at hand rather than chosen based on empirical considerations. The selection of these copulas is often prone to error and can result in grave underestimation of the risk of extreme events as pointed out articulately in Taleb (2010). From this point of view, the Fréchet problem (P) can be seen to try to find the worst-case copula maximizing the risk assigned to the aggregate position S_d. This also explains the current interest in Fréchet inequalities in the context of stress testing the financial system as discussed in Embrechts et al. (2014). Fréchet inequalities provide an exact and disciplined way to assess the worst-case risk of a particular financial position.

Because we are interested in developing computational methods, the marginal distributional information represented through the ambiguity sets $\mathcal{P}_i$ needs to be suitably represented. We will therefore assume that the distributions $\mathbb{P}_i$ of the risk factors X_i in $\mathbb{R}$ are known only to belong to an ambiguity set $\mathcal{P}_i$ defined through a finite number of moment conditions

$$\mathbb{P}_i \in \mathcal{P}_i := \left\{ \mathbb{Q} \;\middle|\; \int g_i \, \mathrm{d}\mathbb{Q} = m_i \in \mathbb{R}^k \right\}, \tag{7}$$

where the functions $g_i : \mathbb{R} \to \mathbb{R}^k$ are called the moment functions and $m_i \in \mathbb{R}^k$ denoted as (generalized) moments following Van Parys (2015). To simplify exposition later on, we will assume that the last component of the moment functions is $g_{i,k} = \mathbb{1}$ with a corresponding moment $m_{i,k} = \int \mathrm{d}\mathbb{P}_i = 1$.

The previous moment condition guarantees that the marginals $\mathbb{P}_i$ are indeed probability distributions. We note that the marginal ambiguity sets $\mathcal{P}_i$ are convex by construction since they are defined through the linear moment constraints (7).

Although the canonical representation of the marginal ambiguity sets $\mathcal{P}_i$ in expression (7) is not universal, many types of distributional ambiguity can nonetheless be represented through a judicious choice of moment function g and moments m. Indeed, both classes of Fréchet problems discussed in Section 1.2 are easily recognized to be in the canonical form (7) for quadratic and piecewise constant moment functions, respectively.

2.2 Fréchet duality

The Fréchet problem (P) is recognized as a linear optimization problem over a convex marginal model $\mathcal{M}$. Unfortunately, the marginal model contains infinite dimensional distributions rather than finite dimensional vectors and hence offers no immediate computational approach. In this section we indicate how this issue can be resolved by considering a convex dual formulation. The convex dual of our Fréchet problem (P) reduces to a (semi-infinite) optimization problem over vectors rather than distributions. We will show here how the lack of dependence information implies a separable and recursive structure in the dual formulation of any Fréchet problem.

The classic dual characterization of Fréchet problem (P) with marginal ambiguity sets in the canonical form (7) is given as the semi-infinite optimization convex problem

$$\begin{aligned} F(\ell, \mathcal{M}) = \inf \quad & \textstyle\sum_{i=1}^{d} \lambda_i^\top m_i \\ \text{s.t.} \quad & \lambda_i \in \mathbb{R}^k, \quad \forall i \in [d], \\ & \textstyle\sum_{i=1}^{d} \lambda_i^\top g_i(x_i) \geq \ell\left(\textstyle\sum_{i=1}^{d} x_i\right), \quad \forall x \in \mathbb{R}^d \end{aligned} \tag{8}$$

as found for instance in Rüschendorf (1991). In contrast to the primal characterization where one looks to *maximize* $\int \ell(S_d)\,\mathrm{d}\mathbb{P}$ over a convex set of probability distributions $\mathcal{M}$, the dual characterization (8) seeks to *minimize* a linear combination of the moments m_i such that the same linear combination of moment functions g_i majorizes the loss function ℓ. Observe that while the dual optimization problem (8) is a convex semi-infinite optimization problem, the primal problem (P) is an infinite dimensional linear optimization problem. Strong duality can be shown to hold under rather mild technical conditions; see Shapiro (2001). For instance, strong duality holds already if the ambiguity set $\mathcal{M}$ has non-empty interior in the weak topology of distributions. Consequently, strong duality holds whenever the marginal ambiguity sets $\mathcal{P}_i$ have non-empty interiors. Unfortunately, constraints involving the positivity of a multivariate function such as those appearing in the dual characterization (8) result in very difficult (albeit convex) optimization problems in general. Hence, it is not clear yet whether the previous dual characterization is any more favorable than its primal counterpart.

Nevertheless, the particular structure of the Fréchet dual problem can be revealed to possess an interesting separable and recursive nature in terms of epi-additions of the moment functions g_i defining the marginal ambiguity sets $\mathcal{P}_i$. Indeed, the lack of dependence structure in the primal characterization of Fréchet problems can be exploited using epi-additions in their dual counterpart.

Lemma 1 (Fréchet dual characterization)**.** *The Fréchet dual* (8) *can be written equivalently using* *epi-additions as*

$$
\begin{aligned}
F(\ell, \mathcal{M}) = \inf \quad & \textstyle\sum_{i=1}^d \lambda_i^\top m_i \\
\text{s.t.} \quad & \lambda_i \in \mathbb{R}^k, \quad \forall i \in [d], \\
& \oplus_{i=1}^d (\lambda_i^\top g_i) \geq \ell.
\end{aligned} \tag{D}
$$

Proof. Observe that we can characterize the dual constraint in (8) using partial minimization equivalently as the condition $\inf_x \{\sum_{i=1}^d \lambda_i^\top g_i(x_i) : \sum_{i=1}^d x_i = t\} \geq \ell(t)$ for all t in $\mathbb{R}$. The last condition is recognized in terms of epi-additions as the requirement $\oplus_{i=1}^d (\lambda_i^\top g_i) \geq \ell$ and is to be understood point-wise. □

We will come to denote the characterization (D) as the Fréchet dual problem. By itself though, this alternative representation does not seem to offer any immediate advantages as the function $\oplus_{i=1}^d (\lambda_i^\top g_i)$ is in general not easy to work with. Nevertheless, working with the Fréchet dual (D) offers many benefits over the more general alternative (8). The following theorem will come to play the protagonist role in this paper.

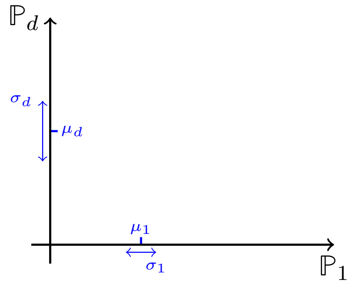
Theorem 1. *The Fréchet dual* (D) *can be recursively decomposed into*

$$
\begin{aligned}
F(\ell, \mathcal{M}) = \inf \quad & \textstyle\sum_{i=1}^d \lambda_i^\top m_i \\
\text{s.t.} \quad & \lambda_i \in \mathbb{R}^k,\ a_i \in \mathcal{F}, \quad \forall i \in [d], \\
& (\lambda_d^\top g_d, \lambda_{d-1}^\top g_{d-1}, a_{d-1}) \in E_{\mathcal{F}},\ \ (a_{i+1}, \lambda_i^\top g_i, a_i) \in E_{\mathcal{F}}, \\
& a_1(t) \geq \ell(t), \quad \forall t \in \mathbb{R}.
\end{aligned} \tag{9}
$$

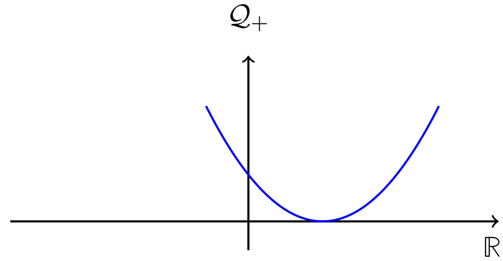
with $\mathcal{F} := \{f : \mathbb{R} \to \mathbb{R}\}$ *and* $E_{\mathcal{F}} := \{f_1, f_2, f_3 \in \mathcal{F} \mid f_1 \oplus f_2 \geq f_3\}$ *represents the epi-addition of two functions in hypograph form.*

Proof. Theorem 1 can easily be seen to hold by taking $a_i = \oplus_{j=i}^d (\lambda_j^\top g_j) \in \mathcal{F}$ for the auxiliary functions in characterization (9) in which case we recover representation (D). The opposite direction follows trivially from the associativity (6) and monotonicity of the epi-addition operator. □

The main benefit of the recursive characterization (9) is that only the epi-addition of two (instead of d) functions needs to be considered. The previous characterization is still troublesome as it involves univariate functions a_i as auxiliary variables. As the space of arbitrary positive univariate functions $\mathcal{F}$ is not finite dimensional, the characterization does indeed not give rise immediately to a computationally tractable representation. Nevertheless, Theorem 1 will come to take a central role in this paper. Using the recursive dual problem (9) as a blueprint, a tractable characterization for both data friendly Fréchet classes discussed in Section 1.2 can indeed be derived. For these particular Fréchet classes, the sets $\mathcal{F}$ and $E_{\mathcal{F}}$ can be sufficiently restricted to allow for efficient optimization over them, thereby rendering the recursive representation (9) computationally attractive. The presentation of the results in subsequent Sections 3 and 4 is deliberately chosen to mirror one another as to stress the fact that the stated results are derived using Theorem 1 as a common template.



(a) Mean/variance information



(b) Quadratic dual functions

Figure 1: The marginal means μ_i and variances σ_i^2 are the only information given concerning the marginal distributions $\mathbb{P}_i$ in the context of Fréchet problem $F(\ell, \mathcal{M}_\mathcal{Q})$. The corresponding dual functions in its dual characterization are positive convex quadratic functions. Theorem 2 exploits the fact that the set of positive quadratic functions $\mathcal{Q}$ is closed under the epi-addition.

3 Fréchet inequalities with marginal means and variances

Estimating the marginal means μ_i and variances σ_i^2 of the risk factors X_i is often the only viable practical option when we only have access to empirical data. When confronted indeed with very limited historical data, marginal means and variances can often still be estimated with high confidence making use of for instance the Hoeffding (1963) inequality. In this section we therefore discuss the Fréchet class in which the marginal information concerning the distribution of the risk factors X_i is limited to second-order moment information. The marginal ambiguity sets for risk factors with known marginal means and variances are given in expression (4) and can represented in the canonical form (7) as

$$\mathcal{P}_{i,\mathcal{Q}} = \left\{ \mathbb{Q} \;\middle|\; \int q_i := t \mapsto (t^2, 2t, 1)\, \mathrm{d}\mathbb{Q} = (s_i^2, 2\mu_i, 1) \right\}$$

with the help of the quadratic moment functions. The marginal second-order moments are readily derived from the given marginal means and variances using the relationship $s_i^2 = \mu_i^2 + \sigma_i^2$. As long as the marginal variances σ_i^2 are strictly positive, then the marginal ambiguity sets $\mathcal{P}_{i,\mathcal{Q}}$ are non-empty and the strong duality holds. Indeed, the normal distributions with appropriate mean and variance are recognized immediately as elements of the marginal ambiguity sets $\mathcal{P}_i$. We will indicate that the Fréchet class corresponding to the marginal ambiguity sets (4) can be tractably solved via our convex optimization perspective.

The primal feasible set $\mathcal{M}_\mathcal{Q} := \mathcal{M}(\mathcal{P}_{1,\mathcal{Q}}, \dots, \mathcal{P}_{d,\mathcal{Q}})$ of the Fréchet problem of interest comprises here all joint probability distributions consistent with the given marginal means and variances; see Figure 1(a). The corresponding Fréchet bound can in this case be characterized using the epi-addition of quadratic moment functions

$$\begin{aligned} F(\ell, \mathcal{M}_\mathcal{Q}) := \inf \quad & \textstyle\sum_{i=1}^d \lambda_{i,1} s_i^2 + 2\lambda_{i,2}\mu_i + \lambda_{i,3} \\ \text{s.t.} \quad & \lambda_i \in \mathbb{R}^3, \quad \forall i \in [d] \\ & \oplus_{i=1}^d (\lambda_i^\top q_i) \,\geq\, \ell. \end{aligned} \tag{10}$$

We can now make the power of the recursive representation of Fréchet problems offered in Theorem 1 concrete by stating its specialization to Fréchet problems with given marginal means and variances. Theorem 2 will establish that the auxiliary functions appearing in the recursive dual Fréchet formulation (9) preventing a direct computational impact may in fact be restricted to the set $\mathcal{Q}_+$ of positive quadratic functions. The previous observation will render the Fréchet formulation (9) tractable as we can optimize over positive quadratics in $\mathcal{Q}_+$ efficiently.

Before we can state the main theorem of this section, we first offer a suitable representation for the set of positive quadratic functions and then derive a convex representation for the set $E_{\mathcal{Q}}$ representing the epi-addition of two positive quadratic functions.

Definition 1 (Positive quadratic functions)**.** *A univariate real function c is* positive quadratic *if we can write $c(t) = c_1 t^2 + 2c_2 t + c_3$ for a coefficient vector c in the convex set*

$$\mathcal{Q}_+ := \left\{ c \in \mathbb{R}^3 \;\middle|\; \begin{pmatrix} c_1 & c_2 \\ c_2 & c_3 \end{pmatrix} \in \mathbb{S}^2_+ \right\}. \tag{11}$$

As can be remarked from Definition 1, we do not make the distinction between a positive quadratic function and its coefficients. Although this implies a slight abuse of notation as both objects are of course distinct, it does not impede the readability of this paper. It can be remarked that the set of all positive quadratic functions $\mathcal{Q}_+$ is convex and semidefinite representable. It is well known that the positive quadratic nature of functions is preserved under epi-addition. Furthermore, in the next lemma we indicate that the epi-addition (its hypograph in fact) is semidefinite representable as well. The last two observations will constitute the backbone of our main Theorem 2.

Lemma 2 (Epi-addition of positive quadratic functions)**.** *For all positive quadratic functions a and b, we have that their epi-addition is positive quadratic as well. Furthermore, the set of all positive quadratic functions c less than $a \oplus b$ is representable as the set*

$$E_{\mathcal{Q}} = \left\{ (a, b, c) \in \mathcal{Q}^3_+ \;\middle|\; \begin{pmatrix} a_1 - c_1 & a_2 - c_2 & -a_1 \\ a_2 - c_2 & a_3 + b_3 - c_3 & b_2 - a_2 \\ -a_1 & b_2 - a_2 & a_1 + b_1 \end{pmatrix} \in \mathbb{S}^3_+ \right\}. \tag{12}$$

Proof. The previous lemma can be proven easily starting from the definition of epi-addition given in the introduction. The epi-addition of the two quadratic functions a and b can be written explicitly as

$$[a \oplus b](t) = -\frac{(b_2 - a_2 - a_1 t)^2}{a_1 + b_1} + a_1 t^2 + 2a_2 t + a_3 + b_3$$

where the previous expression can be recognized as a convex quadratic function in the variable $t \in \mathbb{R}$. The set $E_{\mathcal{Q}}$ can be obtained through a simple Schur complement transformation on the previous expression as discussed in Zhang (2006). $\square$

The subsequent theorem follows the blueprint of our general result in Theorem 1 and exploits the fact that the epi-addition of two quadratic moment functions is a quadratic function as well.

Theorem 2. *The Fréchet problem $F(\ell, \mathcal{M}_{\mathcal{Q}})$ with given marginal means μ_i and variances $0 < \sigma_i^2 = s_i^2 - \mu_i^2$ can be recursively represented as the convex optimization problem*

$$\begin{aligned} \inf \quad & \textstyle\sum_{i=1}^d \lambda_{i,1} s_i^2 + 2\lambda_{i,2}\mu_i + \lambda_{i,3} \\ \text{s.t.} \quad & \lambda_i \in \mathcal{Q}_+,\ a_i \in \mathcal{Q}_+,\ \forall i \in [d], \\ & (\lambda_d, \lambda_{d-1}, a_{d-1}) \in E_{\mathcal{Q}},\ (a_{i+1}, \lambda_i, a_i) \in E_{\mathcal{Q}}, \\ & a_{1,1} t^2 + 2a_{1,2} t + a_{1,3} \geq \ell(t), \quad \forall t \in \mathbb{R}. \end{aligned} \tag{13}$$

where the convex semidefinite sets $\mathcal{Q}_+$ *and* $E_{\mathcal{Q}}$ *are given in expressions* (11) *and* (12)*, respectively.*

Proof. Theorem 2 can easily be seen to hold by taking a_i as the functions $\oplus_{j=i}^d \lambda_j^\top q_j$ at which point we recover the dual Fréchet formulation (10). We can restrict ourselves to positive quadratic auxiliary functions as ℓ is positive and because $\oplus_{i=1}^d(\lambda_i^\top q_i) \geq 0$ implies $(\lambda_i^\top q_i) \geq 0$ for all i in $[d]$. Remark indeed that Lemma 2 guarantees that the resulting auxiliary functions $a_i \in \mathcal{Q}_+$ are positive quadratic functions. As before, the opposite direction follows trivially from the associativity and monotonicity of the epi-addition operator. $\square$

We do need to remark that the ultimate constraint of (13) can still pose a challenge. Fortunately, this semi-infinite convex constraint is not problematic as it is merely univariate and thus under quite mild conditions efficiently representable. When the cost function is merely piecewise polynomial for instance, the ultimate constraint reduces to an exact sum-of-squares condition as discussed at length in Lasserre (2009). We do not however go into the details here, but rather focus on an interesting particular case which will be illustrated with a small numerical experiment in Section 6.

If we are interested in determining the worst-case probability of the event the aggregate position S_d exceeds β, we consequently need to work with the indicator loss function $\ell = \mathbb{1}_{\geq \beta}$. In this very particular case the S-lemma discussed by Pólik and Terlaky (2007) guarantees that the ultimate constraint of (13) reduces to the convex semidefinite condition

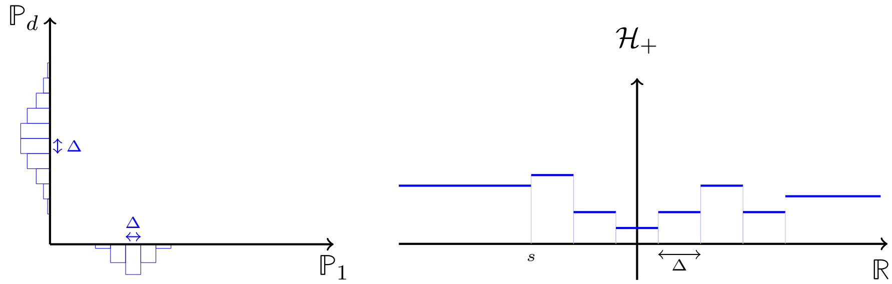
$$\exists \tau \in \mathbb{R}_+ : \quad \begin{pmatrix} a_{1,1} & a_{2,1} \\ a_{2,1} & a_{3,1} - 1 \end{pmatrix} + \tau \begin{pmatrix} 0 & -1 \\ -1 & 2\beta \end{pmatrix} \in \mathbb{S}^2_+.$$

Our Theorem 2 then consequently offers an exact reformulation of this worst-case probability Fréchet problem $F(\mathbb{1}_{\geq \beta}, \mathcal{M}_{\mathcal{Q}})$ as an semidefinite optimization problem counting $\mathcal{O}(d)$ variables for $\mathcal{O}(d)$ constraints all involving the semidefinite cone $\mathbb{S}^3_+$.

Finally, we remark that the particular Fréchet problem $F(\mathbb{1}_{\geq \beta}, \mathcal{M}_{\mathcal{Q}})$ can be recognized as a generalized Chebyshev problem. Generalized Chebyshev problems are well studied and are known to admit exact semidefinite optimization formulations as stated in for instance the works of Vandenberghe et al. (2007); Van Parys et al. (2016). They are commonly employed to guard against overfitting data as pointed out in Delage and Ye (2010); Stellato et al. (2016). Our particular reformulation (13) offers nevertheless a new valuable perspective which is at the same time applicable to a much wider range of loss functions. In the subsequent section, we will carry over the results presented here for Fréchet problems with given marginal means and variances, to a Fréchet class based on histogram data. The presentation of the results in the next section is deliberately chosen to parallel the presentation here to highlight the fact that both sections follow the common template suggested in Theorem 1.

4 Fréchet inequalities with marginal histograms

In practice it is often the case that information beyond merely the mean and variance of the marginal distributions can be recovered either from data or expert knowledge concerning the problem at hand. In those situations, the bound provided by the Fréchet problem $F(\ell, \mathcal{M}_{\mathcal{Q}})$ can be rather loose despite its practical appeal of needing only the marginal means and variances. It is indeed exactly



(a) Histogram information

(b) Piecewise constant dual functions

Figure 2: The distributional ambiguity represented by the ambiguity sets $\mathcal{P}_{i,\mathcal{H}}$ corresponds to the ambiguity left when we are given histograms with fixed bin size Δ of the marginal distributions $\mathbb{P}_i$. The dual functions in the dual characterization of the Fréchet problem $F(\ell, \mathcal{M}_\mathcal{H})$ are piecewise constant functions. Theorem 3 exploits the fact that the set of positive piecewise constant functions $\mathcal{H}$ is closed under the epi-addition.

because only two moments of each marginal distribution are considered that the corresponding bound can be overly pessimistic.

Histograms give a more precise sense of the marginal distributions. Hence instead, we will consider here Fréchet problems in which the marginal information concerning the risk factors X_i are given in terms of histograms with b bins of equal width Δ and minimum bin locations s. The corresponding marginal ambiguity sets $\mathcal{P}_{i,\mathcal{H}}$ are given in equation (5). We immediate recognize that these marginal ambiguity sets

$$\mathcal{P}_{i,\mathcal{H}} := \left\{ \mathbb{Q} \;\middle|\; \int h_i := (\mathbb{1}_{\leq s_i}, \ldots, \mathbb{1}_{\leq s_i+(b-1)\cdot\Delta}, 1)\, \mathrm{d}\mathbb{Q} = p_{i,j} \right\} \tag{14}$$

fit our canonical form (7) for positive piecewise constant moment functions on $b+1$ pieces. The moments $p_{i,j}$ in the marginal ambiguity set (14) represent a binned version of the cumulative distribution function of each of the distribution of the risk factors X_i. If the binned values $p_{i,j}$ are increasing in j then the marginal ambiguity set $\mathcal{P}_{i,\mathcal{H}}$ is non-empty. We refer to Figure 2 for an illustration of a Fréchet problem with histogram information.

Observe that this type of marginal information is particularly well suited to describe distributional ambiguity when working directly with data. Distributions obtained using empirical observations are often presented in the form of a histogram. They were indeed originally proposed by Pearson (1894) to do just that. The number of bins b and bin size Δ must in practice be properly adapted to the data it must present. The more and better historical data is available the more refined histograms can be used. Thanks to the flexibility of histograms, Fréchet problems with complete marginal information discussed in the introduction can also be approximated to an arbitrary degree by using sufficiently refined histograms.

The primal feasible set $\mathcal{M}_\mathcal{H} := \mathcal{M}(\mathcal{P}_{1,\mathcal{H}}, \ldots, \mathcal{P}_{d,\mathcal{H}})$ of the Fréchet problem of interest consists of all joint probability distributions having marginal histograms with binned values $p_{i,j}$ at evenly spaced locations. The corresponding Fréchet bound can in this case be characterized using the epi-addition of piecewise constant moment functions instead

$$
\begin{aligned}
F(\ell, \mathcal{M}_{\mathcal{H}}) := \inf \quad & \textstyle\sum_{i=1}^{d} \sum_{j=1}^{b+1} \lambda_{i,j} \cdot p_{i,j} \\
\text{s.t.} \quad & \lambda_i \in \mathbb{R}^{b+1}, \quad \forall i \in [d] \\
& \oplus_{i=1}^{d} (\lambda_i^\top h_i) \geq \ell.
\end{aligned} \tag{15}
$$

We now repeat the exercise done in the previous section in the context of the alternative Fréchet problem $F(\ell, \mathcal{M}_{\mathcal{H}})$. Once more we will make the common recursive dual template (9) concrete by stating its specialization to Fréchet problems with histogram information. Theorem 3 will establish that the auxiliary functions a_i appearing in the recursive dual (9) preventing a direct computational impact may in fact be restricted to the set $\mathcal{H}_+$ of positive piecewise constant functions. The previous observation will render the template (9) tractable as we can optimize over positive piecewise constant function in $\mathcal{H}_+$ efficiently.

Before we can state the main theorem of this section, we first offer a suitable representation for the set of positive piecewise constant functions and then derive a convex representation for the set $E_{\mathcal{H}}$ representing the epi-addition of two of those functions.

Definition 2 (Positive piecewise constant functions)**.** *A univariate real function c is a positive piecewise constant function of degree n if we can write $c(t) = \sum_{j=1}^{n} c_j \, \mathbb{1}_{\leq s+(j-1)\cdot\Delta}(t) + c_{n+1}$ for a coefficient vector c in the convex set*

$$
\mathcal{H}(s, n)_+ := \left\{ c \in \mathbb{R}^{n+1} \;\middle|\; \forall k \in [n] : \quad \textstyle\sum_{j=k}^{n} c_j \geq 0 \right\}. \tag{16}
$$

Again we will not make the distinction between a positive piecewise constant function and its coefficients. The set of all positive piecewise constant functions $\mathcal{H}(s, n)_+$ is remarked to be a polyhedral convex set. It can be shown that epi-addition preserves the piecewise constant property of functions in the same way as it did for quadratic functions. The situation is slightly more complicated with regards to the degree of the piecewise constant functions. Indeed, the degree of the epi-addition of two piecewise constant functions is one short of the sum of their respective degrees. In the next lemma we show that the epi-addition (rather its hypograph) of two of piecewise constant functions is nevertheless representable as a convex polyhedron. The previous two observations will form the backbone of our main Theorem 3.

Lemma 3 (Epi-addition of piecewise constant functions)**.** *Given two positive piecewise constant functions $a \in \mathcal{H}_+(s_1, n_1)$ and $b \in \mathcal{H}_+(s_2, n_2)$, then their epi-addition $a \oplus b \in \mathcal{H}_+(s_1+s_2, n_1+n_2-1)$ is positive piecewise constant as well. Furthermore, the set of all piecewise constant minorants c in $\mathcal{H}_+(s_1 + s_2, n_1 + n_2 - 1)$ to the epi-addition $a \oplus b$ is representable as the polytopic set*

$$
E_{\mathcal{H}} := \left\{ (a, b, c) \;\middle|\; \begin{array}{c} \forall k \in [n_1 + n_2], \ \forall j \in [\min(1, k - n_2 + 1), \ldots, \max(n_1 + 1, k + 1)] : \\ \sum_{i=k}^{n_1+n_2} c_i \leq \sum_{i=j}^{n_1+1} a_i + \sum_{i=k+1-j}^{n_2+1} b_i, \\ \sum_{i=k}^{n_1+n_2} c_i \leq \sum_{i=j}^{n_1+1} a_i + \sum_{i=k+2-j}^{n_2+1} b_i. \end{array} \right\}, \tag{17}
$$

where we use the summation symbol $\sum_{i=a}^{b}$ as shorthand for $\sum_{i=\min(\max(1,a),b)}^{b}$ in order to preserve readability.

Proof. The fact that two positive piecewise constant functions $a \in \mathcal{H}(s_1, n_1)$ and $b \in \mathcal{H}(s_2, n_2)$ yield a positive piecewise constant epi-addition $a \oplus b \in \mathcal{H}(s_1 + s_2, n_1 + n_2 - 1)$ can be seen from its definition

$$\begin{aligned} a \oplus b(t) := & \ \inf_s & a(s) + b(t-s), \\ = & \ \min_{i \in \mathbb{N}} & a(i \cdot \Delta) + b(t - i \cdot \Delta), \end{aligned}$$

where the equality follows from the fact that both the functions a and b are constant and positive on pieces of size Δ. As each one of the terms $a(i \cdot \Delta) + b(t - i \cdot \Delta)$ is piecewise constant in t on the same pieces, it follows that their minimum is as well. The remainder of the theorem is an exercise in (tedious) bookkeeping between the coefficients of the functions a, b and c. $\square$

One could remark that the previous theorem is quite remarkable in that the degree of the epi-addition of two piecewise constant functions on nonuniform pieces is the product of the their degrees rather than the sum. The even spacing of the histogram bins is hence the critical property that will ultimately render our approach tractable. We are not the first to make an observation along these lines. The most famous observation concerns the discrete Fourier transformation. The discrete Fourier transformation of a sequence of length d requires generally $\mathcal{O}(d^2)$ work. In case the sequence is uniformly spaced, however, the fast Fourier transformation reduces the needed work to merely $\mathcal{O}(d \log d)$ effort. A similar phenomena seems to occur here.

The subsequent theorem follows the blueprint of our general result in Theorem 1 and exploits the fact that the epi-addition of piecewise constant functions is a piecewise constant function itself.

Theorem 3. *The Fréchet problem $F(\ell, \mathcal{M}_{\mathcal{H}})$ with given marginal histograms bin values $p_{i,j}$ can be recursively represented as the linear optimization problem*

$$\begin{aligned} \inf \ & \textstyle\sum_{i=1}^{d} \sum_{j=1}^{b+1} \lambda_{i,j} \cdot p_{i,j} \\ \text{s.t.} \ & \lambda_i \in \mathcal{H}_+(s_i, b),\ a_i \in \mathcal{H}_+(\tilde{s}_i, \tilde{n}_i),\ \ \forall i \in [d], \\ & (\lambda_d, \lambda_{d-1}, a_{d-1}) \in E_{\mathcal{H}},\ \ (a_{i+1}, \lambda_i, a_i) \in E_{\mathcal{H}}, \\ & \textstyle\sum_{i=1}^{\tilde{n}_1} a_{1,i} \cdot \mathbb{1}_{\leq \tilde{s}_1 + \Delta \cdot (i-1)}(t) + a_{1,\tilde{n}_1+1} \geq \ell(t), \quad \forall t \in \mathbb{R}. \end{aligned} \tag{18}$$

were the convex polyhedra $\mathcal{H}_+$ and $E_{\mathcal{H}}$ are given in (16) *and* (17)*, respectively. Additionally, $\tilde{s}_i$ is taken to be the running sum $\sum_{j=i}^{d} s_j$ and $\tilde{n}_i := (b-1)(d-i) + b$.*

Proof. Theorem 3 can be seen to hold by taking a_i as piecewise constant functions $\oplus_{j=i}^{d} (\lambda_j^\top h_j)$ at which point we recover the dual Fréchet formulation (15). We can again restrict ourselves to positive piecewise constant functions as the loss function ℓ is positive and because $\oplus_{i=1}^{d} (\lambda_i^\top h_i) \geq 0$ implies $(\lambda_i^\top h_i) \geq 0$ for all i in $[d]$. Lemma 3 then indeed guarantees that the auxiliary functions a_i constructed that way are piecewise constant functions in $\mathcal{H}_+(\tilde{s}_i, \tilde{n}_i)$. Once more, the opposite direction holds trivially. $\square$

The ultimate semi-infinite constraint in the dual formulation (18) again justifies a few words of commentary. The constraint enforces a semi-infinite condition but is nevertheless quite easy to deal with as it is merely univariate. A moments reflection learns that it can always be represented using only $\tilde{n}_1 + 1$ linear constraints. Indeed, one constraint for every piece of the auxiliary function a_1. The previous observation does assume that we can obtain the maxima of the loss function ℓ restricted to each single piece efficiently. Rather than dwelling on the details here, we again focus attention to the interesting particular case of characterizing the worst-case probability of the event $S_d \geq \beta$ building on a similar discussion in the previous section.

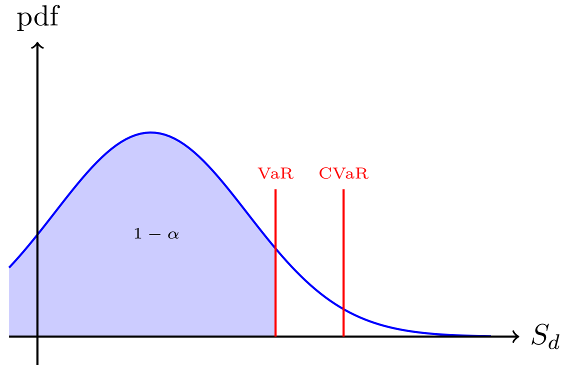


Figure 3: The VaR of the position S_d at risk level $\alpha \in (0,1)$ quantifies the largest value of S_d which occurs with frequency at least α. The CVaR at level α coincides for continuous distributions with the conditional expectation of the distribution of S_d above its corresponding VaR.

The worst-case probability of the event $S_d \geq \beta$ is characterized by the corresponding Fréchet problem for the indicator loss function $\ell = \mathbb{1}_{\geq \beta}$. In this particular instance, the ultimate constraint of dual formulation (18) reduces to set the linear constraints

$$\textstyle\sum_{k=j}^{\tilde{n}_1+1} a_{1,k} \geq 1, \quad \forall \left\{ j \in [\tilde{n}_1 + 1] \;\middle|\; j \geq \lfloor (\beta - \tilde{s}_1)/\Delta \rfloor \right\}.$$

Theorem 3 consequently offers an exact reformulation of the worst-case probability Fréchet problem $F(\mathbb{1}_{\geq \beta}, \mathcal{M}_{\mathcal{H}})$ as a linear optimization problem counting $\mathcal{O}(d^2 \cdot b)$ variables for $\mathcal{O}(d^3 \cdot b^2)$ linear inequality constraints.

We remark that that we are not the first to consider Fréchet classes based on marginal histogram information. The `rearrangement method` presented in Puccetti and Rüschendorf (2012) for instance goes about approximating Fréchet problems with complete information by first binning the known marginals $\mathbb{P}_i$ into histogram form. It then approximates the resulting Fréchet problem with histogram information using a shuffle heuristic with no guarantees of optimality or even finite execution time. We remark that although the exact linear programming approach taken by Prékopa (1988) could be extended to this setting as well, it would unfortunately yield a linear optimization characterization of size exponential in the dimension d. To the best of our knowledge, we are the first to obtain a relatively small polynomial sized exact convex reformulation for Fréchet problems with histogram information.

5 Extensions

5.1 The VaR and CVaR Fréchet problems

The most popular risk measures in finance and insurance applications are the VaR and, to a lesser extent, the CVaR of the aggregate position S_d. To be self contained, we briefly discuss how both measures are related to our canonical form (1) and refer the reader to McNeil et al. (2015) for further information.

The VaR of a position S_d quantifies the largest economic loss occurring with odds at least $\alpha \in (0, 1)$; see Figure 3.

Definition 3 (Value-at-Risk)**.** *The Value-at-Risk of a univariate random variable S_d distributed as $\mathbb{P}$ is defined as*

$$\mathbb{P}\text{-VaR}_\alpha(S_d) := \inf\left\{\beta \in \mathbb{R} \mid \mathbb{P}(S_d \geq \beta) \leq \alpha\right\}. \tag{19}$$

The predominance of the VaR measure in financial and insurance risk quantification largely derives from its simplicity — it is easily recognized as a quantile function of the distribution of S_d — and its encouragement in the Basel II banking accords as discussed in Danielsson et al. (2001). While the VaR is likely the most commonly used measure to quantify the risk carried by the aggregate position S_d in the literature, it comes with a few shortcomings. Indeed, the VaR measure is blind to the severity of the risk taken and hence encourages large but remote risks to be taken can lead to potentially devastating results. Some even go so far as to blame the 2007 financial crisis on the misuse of this risk measure. Taleb (2010) provides a compelling nontechnical account for this conviction. A risk measure which has been proposed as an alternative to the VaR measure is its convex counterpart, the CVaR measure.

Definition 4 (Conditional Value-at-Risk)**.** *The Conditional Value-at-Risk of a univariate random variable S_d distributed as $\mathbb{P}$ is defined as*

$$\mathbb{P}\text{-CVaR}_\alpha(S_d) := \inf_{\beta\in\mathbb{R}}\left\{L(\beta,\mathbb{P}) := \beta + \frac{1}{\alpha}\int (S_d-\beta)^+ \mathrm{d}\mathbb{P}\right\}. \tag{20}$$

where $(S_d-\beta)^+$ is taken to mean $\max(0, S_d-\beta)$.

Rockafellar and Uryasev (2000) have shown that the set of optimal solutions for β in (20) is a closed interval whose left endpoint is given by $\mathbb{P}\text{-VaR}_\alpha(S_d)$. Moreover, it can be shown that if the random variable S_d follows a continuous distribution, then its CVaR at risk level α coincides with the conditional expectation above the $\mathbb{P}\text{-VaR}_\alpha(S_d)$-quantile as illustrated in Figure 3. This observation originally motivated the term *conditional value-at-risk*.

In the view of their popularity, much of the literature focusses on either the VaR Fréchet problem or its CVaR counterpart. Slightly abusing notation, we refer here to either problem respectively as

$$F(\text{VaR},\mathcal{M}) := \sup\nolimits_{\mathbb{P}\in\mathcal{M}} \mathbb{P}\text{-VaR}_\alpha(S_d) \quad \text{and} \quad F(\text{CVaR},\mathcal{M}) := \sup\nolimits_{\mathbb{P}\in\mathcal{M}} \mathbb{P}\text{-CVaR}_\alpha(S_d).$$

In the introduction we stated without much discussion that our canonical Fréchet problem (P) offers a solution to both the VaR and CVaR Fréchet problem. The worst-case VaR and CVaR measures are not immediately recognized to be of the canonical form (1), however, and thus require some massaging. In what remains we will indicate that both the VaR and CVaR Fréchet inequalities can be stated using our canonical Fréchet problem (P) as a computational primitive.

From relation (19) it is clear that the VaR measure can be computed efficiently using

$$F(\text{VaR},\mathcal{M}) = \inf\nolimits_\beta\left\{\beta \mid F(\mathbb{1}_{\geq\beta},\mathcal{M}) \leq \alpha\right\},$$

where the Fréchet problem $F(\mathbb{1}_{\geq\beta},\mathcal{M})$ discussed throughout the paper is indeed used as a computational primitive. The outer optimization over the variable β can be done using the golden section search described in Kiefer (1953). Similarly, from a computational point of view the CVaR Fréchet problem can be reduced to the canonical Fréchet problem (P) in much the same manner. The CVaR Fréchet problem can indeed be restated as

$$\begin{aligned}F(\text{CVaR},\mathcal{M}) &= \sup_{\mathbb{P}\in\mathcal{M}}\inf_\beta L(\beta,\mathbb{P}) = \inf_\beta\sup_{\mathbb{P}\in\mathcal{M}} L(\beta,\mathbb{P})\\ &= \inf_\beta\left\{\beta + \sup_{\mathbb{P}\in\mathcal{M}}\frac{1}{\alpha}F\left((S_d-\beta)^+,\mathcal{M}\right)\right\}\end{aligned}$$

Since $L(\beta, \mathbb{P})$ is convex in β and linear in $\mathbb{P}$, the interchange of the supremum and infimum operations is justified when the ambiguity set $\mathcal{M}$ is weakly closed by virtue of a stochastic saddle point theorem due to Shapiro and Kleywegt (2002). The Fréchet problem $F(t \mapsto (t-\beta)^+, \mathcal{M})$ can now be seen to constitute an inner problem in the CVaR Fréchet problem.

5.2 Worst-case joint distributions $\mathbb{P}^\star$

Although Fréchet problems may be approached from their convex dual perspective, as in fact done up to this point in this paper, the dual optimal solution $\lambda^\star$ does not have an immediately physical interpretation in stark contrast to its primal counterpart $\mathbb{P}^\star$. It may indeed be of interest to identify a worst-case distribution $\mathbb{P}^\star$ in $\mathcal{M}$ which achieves the Fréchet bound $F(\ell, \mathcal{M}) = \int \ell(\sum_{i=1}^d X_i) \, \mathrm{d}\mathbb{P}^\star$ should one exist. Discovering what type of uncertainty causes the gravest harm is always worth pursuing. In this section we will therefore consider the relationship between the optimal primal probability distributions $\mathbb{P}^\star$ and the optimal solutions $\lambda^\star$ in its dual characterization. We will show that once we have the dual optimal solution, the optimal probability distribution can be extracted as well at little extra computational cost.

Because of strong duality it follows that a primal maximizer $\mathbb{P}^\star$ and dual minimizer $\lambda^\star$ of the Fréchet problem (P) and its dual (8), respectively, are related as $\int \ell(\sum_{i=1}^d X_i) \, \mathrm{d}\mathbb{P}^\star = \sum_{i=1}^d \lambda_i^{\star\top} m_i$. As the primal maximizer $\mathbb{P}^\star$ is feasible we have by definition that $m_i = \int g_i \, \mathrm{d}\mathbb{P}^\star$ according to (7). The primal and dual extrema are thus related as

$$\int \ell\left(\textstyle\sum_{i=1}^d X_i\right) \mathrm{d}\mathbb{P}^\star = \int \textstyle\sum_{i=1}^d \lambda_i^{\star\top} g_i(X_i) \, \mathrm{d}\mathbb{P}^\star \tag{21}$$

making use of the linearity of integration. Condition (21) is often denoted in the optimization community as the complementarity condition between the primal and dual extrema, $\mathbb{P}^\star$ and $\lambda^\star$, respectively. Because of dual feasibility we must have that $\sum_{i=1}^d \lambda_i^{\star\top} g_i(x_i) - \ell(\sum_{i=1}^d x_i) \geq 0$ point-wise for x in $\mathbb{R}^d$. A direct consequence of this inequality, in combination with the complementarity condition (21), is that $\mathbb{P}^\star$ must be supported on the points at which the dual function $\sum_{i=1}^d \lambda_i^{\star\top} g_i(x_i)$ coincides with the function $\ell(\sum_{i=1}^d x_i)$, i.e.

$$\operatorname{supp} \mathbb{P}^\star \subseteq S^\star := \left\{ x \in \mathbb{R}^d \;\middle|\; \textstyle\sum_{i=1}^d \lambda_i^{\star\top} g_i(x_i) = \ell(\sum_{i=1}^d x_i) \right\}. \tag{22}$$

It is clear that the worst-case distribution $\mathbb{P}^\star$ in $\mathcal{M}$ is not necessarily uniquely characterized by its support. Any joint distribution in the marginal model $\mathcal{M}$ supported on $S^\star$ is in fact optimal in the Fréchet problem (P). If the support $S^\star$ has finite cardinality, then constructing a worst-case joint distribution supported on $S^\star$ matching the required moment conditions requires only the solution of a system of linear equations. In those cases, condition (22) allows for the efficient extraction of a worst-case joint distribution $\mathbb{P}^\star$ from a dual optimal solution $\lambda^\star$.

6 Numerical example

We consider in this final section a risk aggregation problem commonly faced by an insurer holding a portfolio of several distinct types of insurance options. Suppose that our insurer sells three distinct types of insurance policies; life, car and fire insurances. The size of the claims made by the customers holding its sold life insurance policies X_1, car insurance policies X_2 and fire insurance

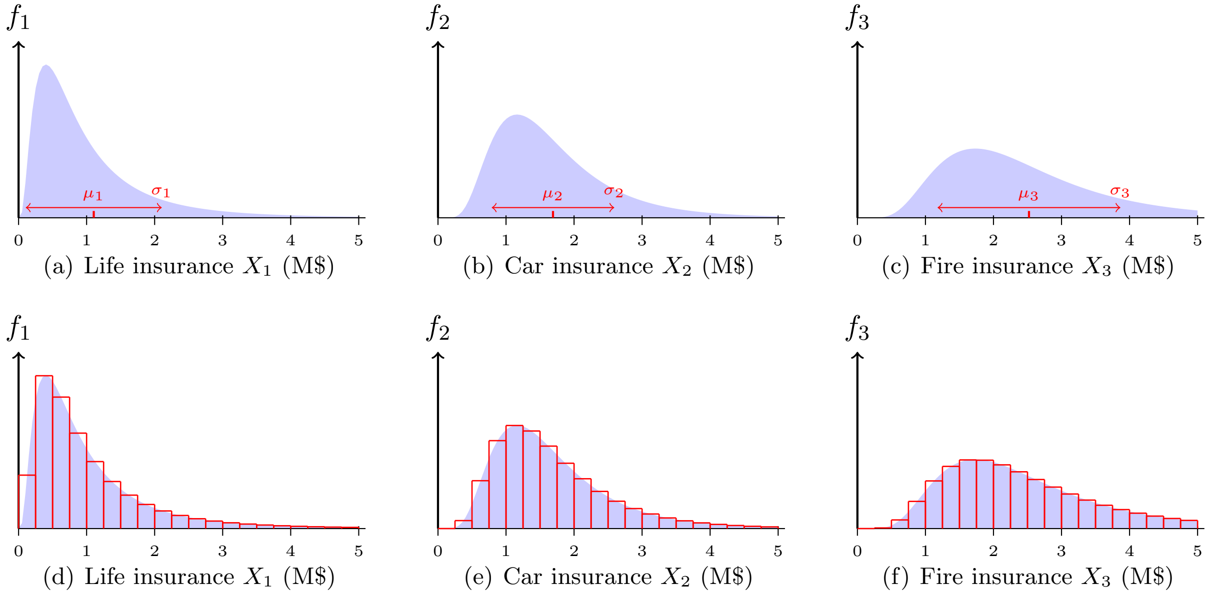


Figure 4: The risk aggregation problem for an insurance portfolio comprising three unrelated insurance policies. In this setting we are interested in determining the risk $\mathbb{P}(S_3 \geq \beta)$ for the monetary threshold $\beta = 15$ M$ given only partial information on the distributions of the insurance claims X_i. The first row visualizes the moment information used by the Fréchet bound $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{Q}})$. The second row attempts the same for the histogram information used by the alternative Fréchet bound $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{H}})$. It is evident that as more information is contained in the histograms than the first two moments of the distributions, the Fréchet bound based on histogram information will be more informative than its counterpart based only on two moments.

policies X_3 is unfortunately not known in advance. This uncertainty is exactly why its costumers buy their policies from the insurer at a premium. Nevertheless, the insurer is not completely blind and has a vague idea of how much of each of the separate insurance policies sold is going to cost the company over the course of a year. Practically such information can be obtained based on a combination of historical costs and probabilistic modeling. An excellent discussion of this topic is provided in McNeil et al. (2015). However, justifying a particular dependence relation between insurance claims incurred through several distinct types of insurance policies is much more challenging and often practically infeasible. The insurance company is nonetheless still interested in obtaining estimates of the total claim it needs to cover. Insurance companies are particularly interested in the probability

$$\mathbb{P}(S_3 \geq \beta) = \int \mathbb{1}\{X_1 + X_2 + X_3 \geq \beta\}\, \mathbb{P}(\mathrm{d}x) \tag{23}$$

of the event in which the total claim S_3 exceeds a certain threshold β. The threshold β might be the maximum amount the insurer can cover before becoming insolvent. Alternatively, it may also simply be the threshold after which the total claim S_3 exceeds the premiums payed for the options by its customers. Due to the fact that the joint distribution $\mathbb{P}$ between the claims is left ambiguous only an upper bound on the previous quantity of interest can be obtained. We will use the Fréchet bounds developed in this document to compute a provably tight bounds on the probability (23).

	Claim (M$)	Location m	Scale v
Life insurance	X_1	-0.3	0.8
Car insurance	X_2	0.4	0.5
Fire insurance	X_3	0.8	0.5

Table 1: The numerical parameter values specifying the log-normal distributions followed by the insurance claims X_i. The corresponding probability density functions are visualized in Figure 4.

For the sake of this example we assume that each of the claims X_i follows a log-normal distribution with probability density function $f_i(x) := (xv_i\sqrt{2\pi})^{-1} \exp\left(-(\log x - m_i)^2/(2v_i^2)\right)$. The particular numerical values which the location m_i and scale v_i parameters take on for each of the insurance types sold are presented in Table 1. Evidently, the fact that the claims X_i are log-normally distributed is not known to the insurer who will have to do with estimates based on historical data. We will determine the probability of the event for which the total claim S_3 exceeds $\beta = 15$M$ using partial information concerning the marginal distributions. For the Fréchet bounds $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{Q}})$ based on marginal means and variances only

$$\mu_i = \exp\left(m_i + v_i^2/2\right) \text{ and } s_i^2 = \exp\left(2m_i + 2v_i^2\right) \tag{24}$$

needs to be determined from data. Fréchet bounds $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{H}})$ based on the histogram information provide a flexible way to incorporate more information regarding the distribution of the insurance claims X_i through their sampled cumulative distributions

$$p_{i,j} = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{\log\left((j-1)\cdot\Delta - m_i\right)}{\sqrt{2}v_i}\right), \quad \forall j \in [b]. \tag{25}$$

For this example we consider histograms with $b = 41$ bins of size $\Delta = 0.25$ M$ starting at the origin. It is clear that the histogram captures more information regarding the insurance claim distributions than two moments do and hence the Fréchet bound $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{H}})$ is expected to be more informative than its rival $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{Q}})$. The risk aggregation problem faced by the insurer based on these distinct distributional information sets are visually illustrated in Figure 4.

The Fréchet bounds $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{Q}})$ and $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{H}})$ were posed as, respectively, a semidefinite and linear optimization problem and can also be numerically solved as such. All algorithms in this document are implemented in `MATLAB` and executed on a standard `Intel Core i5 @ 2.60GHz` running `OS X release 10.10.5`. All optimization was done using `SeDuMi` by Sturm (1999) for the semidefinite optimization problems and likewise the commercial mathematical optimization distribution `GUROBI version 6.5` for the linear optimization problems. Both these solvers were interfaced through the problem parser `YALMIP` developed by Löfberg (2004).

Using the marginal means and variances in (24), the probability $\mathbb{P}(S_3 \geq \beta)$ can be upper bounded by $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{Q}}) = 12.7\%$. The corresponding semidefinite optimization reformulation problem was solved using `SeDuMi` in well under a second. From Figure 4 it is visually quite clear that this Fréchet bound might not be so informative as the underlying distributions are not captured particularly well by their first two moments alone. Based on the more informative histogram information (25), a much better bound $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{H}}) = 5.8\%$ can be constructed. The corresponding linear optimization problem was fed to `GUROBI` and solved in approximately five seconds. Although both bounds are distinct, either is tight for the information they require concerning the underlying insurance claim distributions. In practice it will be the amount of available data which will decide whether to use the crude bound $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{Q}})$ or its more sophisticated alternative $F(\mathbb{1}_{\geq\beta}, \mathcal{M}_{\mathcal{H}})$.

To put these numbers in perspective, the probability $\mathbb{P}(S_3 \geq \beta)$ when the claims X_i are independent can be numerically determined to be approximately 0.2% following Fenton (1960). However, risk assessments obtained through independence assumptions are often hopelessly optimistic as eloquently pointed by Taleb (2010).

7 Conclusion

Risk aggregation in the absence of dependence information is a problem fundamental to both insurance and financial management. Fréchet inequalities provide an exact and disciplined perspective on risk aggregation without assuming independence, indeed, without making any dependence assumptions whatsoever. Due to their inherent robustness against hard to come by dependence structures, Fréchet problems have witnessed recently a surge of interest. Despite their practical relevance however, computational algorithms have been scarce to come by. In this paper we provided a novel computational and data friendly perspective on Fréchet inequalities via a modern convex optimization lens. We considered ambiguous marginal distributions based on mean/variance or histograms as data friendly alternatives to exact marginal distributions. In reality indeed data rather than distributions are given and hence some ambiguity can never be avoided completely. We unveiled an interesting separable and recursive nature common to all Fréchet problems. This special nature of Fréchet problems was exploited to derive exact and tractable numerical algorithms. To the best of our knowledge we are the first to provide data friendly and exact computational tools applicable to practical Fréchet problems.

References

Bertsimas, D. and Popescu, I. (2005). Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization*, 15(3):780–804.

Boole, G. (1854). *An Investigation of The Laws of Thought: On Which Are Founded the Mathematical Theories of Logic and Probabilities.* Dover Publications.

Danielsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O., and Shin, H. (2001). An academic response to Basel II. Technical report, LSE Financial Markets Group.

Das, B., Embrechts, P., and Fasen, V. (2013). Four theorems and a financial crisis. *International Journal of Approximate Reasoning*, 54(6):701–716.

Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612.

Embrechts, P., Lindskog, F., and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. *Handbook of Heavy Tailed Distributions in Finance*, 8(1):329–384.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, 2(1):25–48.

Fenton, L. (1960). The sum of log-normal probability distributions in scatter transmission systems. *IRE Transactions on Communications Systems*, 8(1):57–67.

Fréchet, M. (1935). Généralisation du théoreme des probabilités totales. *Fundamenta Mathematicae*, 1(25):379–387.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.

Keogh, E. and Mueen, A. (2010). Curse of dimensionality. In *Encyclopedia of Machine Learning*, pages 257–258. Springer.

Kiefer, J. (1953). Sequential minimax search for a maximum. *Proceedings of the American Mathematical Society*, 4(3):502–506.

Lasserre, J. (2009). *Moments, Positive Polynomials and their Applications*, volume 1. World Scientific.

Löfberg, J. (2004). YALMIP: A toolbox for modeling and optimization in MATLAB. In *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, pages 284–289. IEEE.

Makarov, G. (1982). Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability & its Applications*, 26(4):803–806.

McNeil, A., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Owhadi, H., Scovel, C., Sullivan, T., McKerns, M., and Ortiz, M. (2013). Optimal uncertainty quantification. *SIAM Review*, 55(2):271–345.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London*, 185:71–110.

Pólik, I. and Terlaky, T. (2007). A survey of the S-lemma. *SIAM review*, 49(3):371–418.

Popescu, I. (2005). A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Mathematics of Operations Research*, pages 632–657.

Prékopa, A. (1988). Boole-bonferroni inequalities and linear programming. *Operations Research*, 36(1):145–162.

Puccetti, G. and Rüschendorf, L. (2012). Computation of sharp bounds on the distribution of a function of dependent risks. *Journal of Computational and Applied Mathematics*, 236(7):1833–1840.

Puccetti, G. and Rüschendorf, L. (2013). Sharp bounds for sums of dependent risks. *Journal of Applied Probability*, 50(1):42–53.

Puccetti, G., Wang, B., and Wang, R. (2013). Complete mixability and asymptotic equivalence of worst-possible var and es estimates. *Insurance: Mathematics and Economics*, 53(3):821–828.

Rockafellar, R. (1970). *Convex Analysis*. Number 28 in Princeton Mathematics Series. Princeton University Press.

Rockafellar, R. and Uryasev, S. (2000). Optimization of conditional Value-at-Risk. *Journal of Risk*, 2:21–42.

Rockafellar, R. T. and Wets, R. J.-B. (1998). *Variational Analysis*. Springer-Verlag.

Rüschendorf, L. (1991). *Fréchet Bounds and their Applications*. Springer.

Shapiro, A. (2001). On duality theory of conic linear problems. *Nonconvex Optimization and its Applications*, 57:135–155.

Shapiro, A. and Kleywegt, A. (2002). Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542.

Stellato, B., Van Parys, B., and Goulart, P. (2016). Multivariate Chebyshev inequality with estimated mean and variance. *The American Statistician*, (just-accepted):1–13.

Sturm, J. (1999). Using SeDuMi `1.02`, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11(1-4):625–653.

Taleb, N. (2010). *The Black Swan: The Impact of the Highly Improbable*. Random House.

Van Parys, B. (2015). *Distributionally robust control and optimization*. PhD thesis, ETH Zürich.

Van Parys, B., Goulart, P., and Kuhn, D. (2016). Generalized Gauss inequalities via semidefinite programming. *Mathematical Programming*, 156(1-2):271–302.

Vandenberghe, L., Boyd, S., and Comanor, K. (2007). Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52–64.

Villani, C. (2008). *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media.

Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, 102(10):1344–1360.

Wang, R., Peng, L., and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, 17(2):395–417.

Zhang, F. (2006). *The Schur Complement and Its Applications*, volume 4. Springer Science & Business Media.