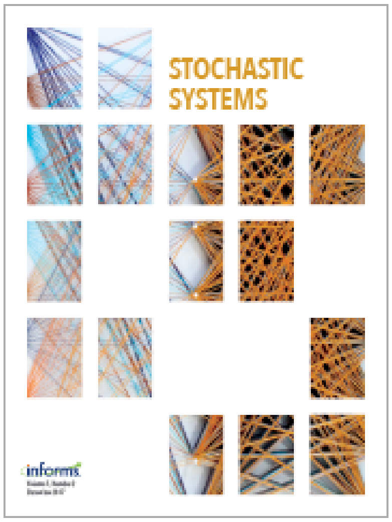


Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Efficient Scenario Generation for Heavy-Tailed Chance Constrained Optimization

Jose Blanchet, Fan Zhang, Bert Zwart

To cite this article:
Jose Blanchet, Fan Zhang, Bert Zwart (2024) Efficient Scenario Generation for Heavy-Tailed Chance Constrained Optimization. Stochastic Systems 14(1):22-46. https://doi.org/10.1287/stsy.2021.0021

informs
https://pubsonline.informs.org/journal/stsy

STOCHASTIC SYSTEMS
Vol. 14, No. 1, March 2024, pp. 22–46
ISSN 1946-5238 (online)

Efficient Scenario Generation for Heavy-Tailed Chance Constrained Optimization

Jose Blanchet,[a] **Fan Zhang,**[a,*] **Bert Zwart**[b,c]

[a] Department of Management Science and Engineering, Stanford University, Stanford, California 94305; [b] Department of Statistics, Centrum Wiskunde and Informatica, Amsterdam 1098 XG, Netherlands; [c] Department of Mathematics and Computer Science, Eindhoven University of Technology, 5612 AZ Eindhoven, Netherlands

*Corresponding author

Contact: jose.blanchet@stanford.edu (JB); fzh@stanford.edu, https://orcid.org/0000-0001-6735-4901 (FZ); bert.zwart@cwi.nl, https://orcid.org/0000-0001-9336-0096 (BZ)

Received: August 2, 2023
Revised: February 6, 2023
Accepted: August 2, 2023
Published Online in Articles in Advance: September 4, 2023

https://doi.org/10.1287/stsy.2021.0021

Abstract. We consider a generic class of chance-constrained optimization problems with heavy-tailed (i.e., power-law type) risk factors. As the most popular generic method for solving chance constrained optimization, the scenario approach generates sampled optimization problem as a precise approximation with provable reliability, but the computational complexity becomes intractable when the risk tolerance parameter is small. To reduce the complexity, we sample the risk factors from a conditional distribution given that the risk factors are in an analytically tractable event that encompasses all the plausible events of constraints violation. Our approximation is proven to have optimal value within a constant factor to the optimal value of the original chance constraint problem with high probability, uniformly in the risk tolerance parameter. To the best of our knowledge, our result is the first uniform performance guarantee of this type. We additionally demonstrate the efficiency of our algorithm in the context of solvency in portfolio optimization and insurance networks.

Funding: The research of B. Zwart is supported by the NWO (Dutch Research Council) [Grant 639.033.413]. The research of J. Blanchet is supported by the Air Force Office of Scientific Research [Award FA9550-20-1-0397], the National Science Foundation [Grants 1820942, 1838576, 1915967, and 2118199], Defense Advanced Research Projects Agency [Award N660011824028], and China Merchants Bank.

Keywords: stochastic optimization • Monte Carlo methods • regular variation • asymptotic optimality

1. Introduction

In this paper, we consider the following family of chance constrained optimization problems:

$$\begin{aligned} &\text{minimize} && c^\top x \\ &\text{subject to} && \mathrm{P}(\phi(x, L) > 0) \le \delta, \\ &&& x \in \mathbb{R}^{d_x}. \end{aligned} \tag{CCP$_\delta$}$$

where $x \in \mathbb{R}^{d_x}$ is a d_x-dimensional decision vector and L is a d_l-dimensional random vector in $\mathbb{R}^{d_l}$. The elements of L are often referred to as risk factors; the function $\phi : \mathbb{R}^{d_x} \times \mathbb{R}^{d_l} \longrightarrow \mathbb{R}$ is often assumed to be convex in x and often models a cost constraint; the parameter $\delta > 0$ is the risk level of the tolerance. Our framework encompasses the joint chance constraint of the form $\mathrm{P}(\phi_j(x, L) > 0,\ \exists j \in \{1, \ldots, n\}) \le \delta$, by setting $\phi(x, L) = \max_{j=1,\ldots,n} \phi_j(x, L)$.

Chance constrained optimization problems have a rich history in operations research. Introduced by Charnes et al. (1958), chance constrained optimization formulations have proved to be versatile in modeling and decision making in a wide range of settings. For example, Prekopa (1970) used these types of formulations in the context of production planning. The work of Bonami and Lejeune (2009) illustrates how to take advantage of chance constrained optimization formulations in the context of portfolio selection. In the context of power and energy control the use of chance constrained optimization is illustrated in Andrieu et al. (2010). These are just examples of the wide range of applications that have benefited (and continue to benefit) from chance constrained optimization formulations and tools.

Consequently, there has been a significant amount of research effort devoted to the solution of chance constrained optimization problems. Unfortunately, however, these types of problems are provably NP-hard in the worst case (Luedtke et al. 2010). As a consequence, much of the methodological effort has been placed into developing (a) solutions in the case of specific models; (b) convex and, more generally, tractable relaxations; (c) combinatorial optimization tools; and (d) Monte Carlo sampling schemes. Of course, hybrid approaches are also developed. For example, as a combination of type b and type d approaches, Hong et al. (2011) show that the solution to a chance constraint optimization problem can be approximated by optimization problems with constraints represented as the difference of two convex functions. In turn, this is further approximated by solving a sequence of convex optimization problems, each of which can be solved by a gradient based Monte Carlo method. Another example is Peña-Ordieres et al. (2020), which combines relaxations of type b with sample-average approximation associated with type d methods. In addition to the aforementioned types, Hong et al. (2021) provides an upper bound for the chance constraint optimization problem using a robust optimization with a data-driven uncertainty set, achieving a dimension independent sample complexity.

Examples of type a approaches include the study of Gaussian or elliptical distributions when ϕ is affine both in L and x. In this case, the problem admits a conic programming formulation, which can be efficiently solved (Lagoa et al. 2005). Type b approaches include Hillier (1967); Seppälä (1971); Ben-Tal and Nemirovski (2000, 2002); Prékopa (2003); Bertsimas and Sim (2004); Nemirovski and Shapiro (2006a); Chen et al. (2010); and Tong et al. (2022). These approaches usually integrate probabilistic inequalities such as Chebyshev's bound, Bonferroni's bound, Bernstein's approximations, or large deviation principles to construct tractable analytical approximations. Type c methods are based on branch and bounding algorithms, which connect squarely with the class of tools studied in areas such as integer programming (Ahmed and Shapiro 2008, Luedtke et al. 2010, Küçükyavuz 2012, Luedtke 2014, Zhang et al. 2014, Lejeune and Margot 2016). Type d methods include the sample gradient method, the sample average approximation, and the scenario approach. The sample gradient method is usually combined with a smooth approximation (see Hong et al. (2011) for example). The sample average approximations studied by Luedtke and Ahmed (2008) and Barrera et al. (2016), although simplifying the constraint's probabilistic structure via replacing the population distribution by sampled empirical distribution, are nevertheless hard to solve due to nonconvex feasible regions. The method we consider in this paper is the scenario approach. The scenario approach is introduced and studied in Calafiore and Campi (2005) and is further developed in a series of papers, including Calafiore and Campi (2006); Nemirovski and Shapiro (2006b).

The scenario approach is the most popular generic method for (approximately) solving chance constrained optimization. The idea is to sample a number N of scenarios (each scenario consists of a sample of L) and enforce the constraint in all of these scenarios. The intuition is that if for any scenario, say $L^{(i)}$, the constraint $\phi(L^{(i)},x)<0$ is convex in x, and $\delta>0$ is small, we expect that by suitably choosing N the constrained regions can be relaxed by enforcing $\phi(L^{(i)},x)<0$ for all $i=1,\dots,N$, leading to a good and, in some sense, tractable (if N is of moderate size) approximation of the chance constrained region. Of course, this intuition is correct only when $\delta>0$ is small and we expect the choice of N to be largely influenced by this asymptotic regime.

By choosing N sufficiently large, the scenario approach allows obtaining both upper and lower bounds which become asymptotically tighter as $\delta\to 0$. In a celebrated paper, Calafiore and Campi (2006) provide rigorous support for this claim. In particular, given a confidence level $\beta\in(0,1)$, if $N\geq(2/\delta)\times\log(1/\beta)+2d+(2d/\delta)\times\log(2/\delta)$, with probability at least $1-\beta$, the optimal solution of the scenario approach relaxation is feasible for the original chance constrained problem and, therefore, an upper bound to the problem is obtained. Unfortunately, the required sample size of N grows with $(1/\delta)\times\log(1/\delta)$ as δ becomes small, limiting the scope of the scenario methods in applications.

Many applications of chance constraint optimization require a very small δ. For example, in the 5G ultra-reliable communication system design, the failure probability δ is no larger than 10^{-5} (Alsenwi et al. 2019); for fixed income portfolio optimization, an investment grade portfolio has a historical default rate of 10^{-4}, reported by Frank (2008).

Motivated by this, Nemirovski and Shapiro (2006b) developed a method that lowers the required sample size to the order of $\log(1/\delta)$, making additional assumptions on the function ϕ (which is taken to be biaffine), and the risk factors L, which are to be assumed light-tailed. Specifically, the moment generating function $E[\exp(sL)]$ is assumed to be finite in a neighborhood of the origin. No guarantee is given in terms of how far the upper bound is from the optimal value function of the problem as $\delta\to 0$.

In the present paper, we focus on improving the scalability of N in terms of $1/\delta$ for the practically important case of heavy-tailed risk factors. Heavy-tailed distributions appear in a wide range of applications in science, engineering, and business (Wierman and Zwart 2012, Embrechts et al. 2013) but, in some aspects, are not as well understood as light-tails. One reason is that techniques from convex duality cannot be applied as the moment generating function of L does not exist in a neighborhood of zero. In addition, probabilistic inequalities, exploited in Nemirovski and Shapiro (2006b), do not hold in this setting. Only very recently, a versatile algorithm for heavy-tailed rare event simulation has been developed in Chen et al. (2019).

The main contribution of our paper is to develop an algorithm that has a sample complexity N uniformly bounded in the risk tolerance parameter, assuming a versatile class of heavy-tailed distributions for L. Specifically, we shall assume that L follows a semiparametric class of models known as multivariate regular variation, which is quite standard in multivariate heavy-tail modeling (Embrechts et al. 2013, Resnick 2013). Moreover, our estimator is shown to be within a constant factor to the solution to (CCP$_\delta$) with high probability, uniformly as $\delta \to 0$. We are not aware of other approaches that provide a uniform performance guarantee of this type.

The main idea of our algorithm is to construct an analytically tractable event C_δ that uniformly contains the violation events $\{l \in \mathbb{R}^{d_l} \mid \phi(x,l) > 0\}$ for all x plausible to be feasible. In view of the reformulation of the probabilistic constraint in (CCP$_\delta$) as $\mathrm{P}(\phi(x,L) > 0 \mid L \in C_\delta) \le (\delta/\mathrm{P}(L \in C_\delta))$, the problem (CCP$_\delta$) can be solved by the scenario approach where L is sampled from the conditional distribution given $L \in C_\delta$. The risk tolerance parameter is adjusted to $\delta/\mathrm{P}(L \in C_\delta)$. The primary challenge is to construct C_δ as tight as possible so that the new risk tolerance parameter $\delta/\mathrm{P}(L \in C_\delta)$ is bounded. (This is at the heart of Property 1 defined later. This property is facilitated in the heavy-tailed setting if we assume that $\phi(x,L)$ has appropriate scaling properties, similar to the distribution of L uniformly over a suitable compact set of decisions.)

We illustrate our assumptions and our framework with a risk problem of independent interest. This problem consists in computing a collective salvage fund in a network of financial entities whose liabilities and payments are settled in an optimal way using the Eisenberg-Noe model (Eisenberg and Noe 2001). The salvage fund is computed to minimize its size to guarantee a probability of collective default after settlements of less than a small prescribed margin. For the sake of demonstrating the broad applicability of our method, we also present a portfolio optimization problem with value-at-risk constraints as an additional running example.

The rest of the paper is organized as follows. In Section 2, we introduce the portfolio optimization problem and the minimal salvage fund problem as particular applications of chance constraint optimization. We use both problems as running examples to provide a concrete and intuitive explanation for the concepts we introduce throughout the paper. In Section 3, we provide a brief review of the scenario approach in Calafiore and Campi (2006). The ideas behind our main algorithmic contributions are given in Section 4, where we introduce its intuition, rooted in ideas originating from rare event simulation. Our algorithm requires the construction of several auxiliary functions and sets, and we summarized the explicit expressions of the sets for the running examples in Table 1. How to do this for a more general setting is detailed in Section 5, in which we also present several additional technical assumptions required by our constructions. In Section 5, we also explain that our procedure results in an estimate that is within a constant factor of the optimal solution of the underlying chance constrained problem with high probability as $\delta \to 0$. In Section 6, we show that the assumptions imposed are valid in our motivating example (as well as a second example with quadratic cost structure inside the probabilistic constraint). Numerical results for the examples are provided in Section 7. Throughout our discussion, in each section we present a series of results that summarize the main ideas of our constructions.

To keep the discussion fluid, we present the corresponding proofs in Appendix A unless otherwise indicated. In Appendix B, we introduce an importance sampling algorithm to sample from a parametric family of regularly varying distribution. We present additional numerical experiments in Appendix C.

1.1. Notations

In the rest of this paper, $\mathbb{R}_+ = [0, +\infty)$ is the set of nonnegative real numbers, $\mathbb{R}_{++} = (0, +\infty)$ is the set of positive real numbers, and $\overline{\mathbb{R}} = [-\infty, +\infty]$ is the extended real line. A column vector with zeros is denoted by $\mathbf{0}$, and a column vector with ones is denoted by $\mathbf{1}$. For any matrix Q, the transpose of Q is denoted by $Q^\top$; the Frobenius norm of Q is denoted by $\|Q\|_F$. The identity matrix is denoted by I. For $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^d$, we use $\alpha \cdot x$ to denote the scalar multiplication of x with α. For two column vectors $x, y \in \mathbb{R}^d$, we say $x \preceq y$ if and only if $y - x \in \mathbb{R}^d_+$. For a

Table 1. Examples of O_δ and C_δ That Satisfy Property 1 When L Is Multivariate Regularly Varying

Examples	Outer approximation set O_δ	Uniform conditional event C_δ
Portfolio optimization (1)	$\{x \in \mathbb{R}^d_{++} \mid \eta \cdot x \ge \overline{F}^{-1}_{\mathbf{1}^\top L}(\delta)\}$	$\{l \in \mathbb{R}^d_{++} \mid 2 \cdot \mathbf{1}^\top l \ge \overline{F}^{-1}_{\mathbf{1}^\top L}(\delta)\}$
Minimal salvage fund (3)	$\bigcap_{j=1}^d \{x \in \mathbb{R}^d_{++} \mid \overline{F}^{-1}_{L_j}(\delta) \le \mathbf{e}_j^\top (I - Q^\top)^{-1} x + m_j\}$	$\bigcup_{j=1}^d \{l \in \mathbb{R}^d_{++} \mid L_j > \overline{F}^{-1}_{L_j}(\delta)\}$

column vector $x \in \mathbb{R}^d$ and a scalar $\alpha \in \mathbb{R}$, we say that $x \preceq \alpha$ if and only if $x \preceq \alpha \cdot \mathbf{1}$. For $\alpha \in \mathbb{R}$ and $E \subseteq \mathbb{R}^d$, we define $\alpha \cdot E = \{\alpha \cdot x | x \in E\}$. The optimal value of an optimization problem (Prob) is denoted by Val(Prob). For any real-valued random variable X with probability measure P, define the inverse tail distribution function $\overline{F}_X^{-1} : [0,1] \to \overline{\mathbb{R}}$ as $\overline{F}_X^{-1}(\delta) = \inf\{x \in \mathbb{R} | \mathrm{P}(x > X) \le \delta\}$. We also use Landau's notation. In particular, if $f(\cdot)$ and $g(\cdot)$ are nonnegative real valued functions, we write $f(t) = O(g(t))$ if $f(t) \le c_0 \times g(t))$ for some $c_0 \in (0,\infty)$ and $f(t) = \Omega(g(t))$ if $f(t) \ge g(t))/c_0$ for some $c_0 \in (0,\infty)$.

2. Running Examples

2.1. Portfolio Optimization with Value-at-Risk Constraint

We first introduce a portfolio optimization problem. Suppose that there are d assets to invest. If we invest a dollar in the jth asset, the investment has mean return μ_j and a nonnegative random loss L_j. Let $x = (x_1, \dots, x_d)$ represent the amount of dollars invested in different assets, and let $\mu = (\mu_1, \dots, \mu_d)$ and $L = (L_1, \dots, L_d)$. We assume that L follows a multivariate heavy-tailed distribution.

A precise definition of this concept is rather involved and will be given in Section 5. Intuitively, $\mathrm{P}(\|L\|_2 > x)$ follows a power law, and the direction $L/\|L\|_2$ is assumed to converge in a suitable sense on the unit sphere, conditioned on the event that $\|L\|_2$ is large.

The portfolio manager's goal is to maximize the mean return of the portfolio, which is equal to $\mu^\top x$, with a portfolio risk constraint prescribed by a risk measure called value-at-risk (VaR). The VaR at level $1-\delta \in (0,1)$ for a random variable X is defined as

$$\mathrm{VaR}_{1-\delta}(X) = \min\{z \in \mathbb{R} : F_X(z) \ge 1-\delta\}.$$

For a given number $\eta > 0$, we formulate the following portfolio optimization problem.

$$\begin{aligned} &\text{maximize} && \mu^\top x \\ &\text{subject to} && \mathrm{VaR}_{1-\delta}(x^\top L) \le \eta, \\ &&& x \in \mathbb{R}^d_{++}. \end{aligned}$$

Using the definition of VaR and the fact that the cumulative distribution function is right continuous, we conclude that $\mathrm{VaR}_{1-\delta}(x^\top L) \le \eta$ is equivalent to $\mathrm{P}(x^\top L - \eta > 0) \le \delta$. To facilitate the technical exposition, we apply the change of variable $x_j \longmapsto 1/x_j$ to homogenize the constraint function, yielding the following equivalent chance constrained optimization problem in standard form:

$$\begin{aligned} &\text{maximize} && \sum_{j=1}^d (\mu_j/x_j) \\ &\text{subject to} && \mathrm{P}(\phi(x,L) > 0) \le \delta, \\ &&& x \in \mathbb{R}^d_{++}, \end{aligned} \tag{1}$$

where $\phi(x,l) = \sum_{j=1}^d (l_j/x_j) - \eta$. Despite the nonlinear objective, Calafiore and Campi (2005, section 4.3) show that it admits an epigraphic reformulation with a linear objective so that the standard scenario approach is applicable.

2.2. Minimal Salvage Fund

In this section, we use chance-constrained optimization to determine the minimal total salvage fund required for a reinsurance company to control its default probability, where the policy holders have complex liability structures.

A reinsurance policy is a contract sold to insurance companies for transferring the financial risk exposure and smoothing the cash flow. In a certain type of reinsurance contract, the reinsurance company is responsible to pay a fixed percentage of the net liability for its clients (in this paper we assume the percentage is 100% for simplicity). Therefore, the total amount of net liability is the minimal amount of salvage fund required for the reinsurance company to avoid default. However, calculating the distribution of the minimal salvage fund is nontrivial because the clients may also have insurance contracts with each other.

Suppose that the reinsurance company has d clients, each is an entity or an insurance firm. Let $L = (L_1, \dots, L_d) \in \mathbb{R}^d_+$ denote the vector of incurred losses by each firm, where L_j denotes the total incurred loss that entity j is responsible to pay. We assume that L follows a multivariate heavy-tailed distribution as in the previous example. Let $Q = (Q_{i,j} : i,j \in \{1,\dots,d\})$ be a deterministic matrix where $Q_{i,j}$ denotes the amount of money received by entity j when entity i pays one dollar. We assume that $Q_{i,j} \ge 0$ and $\sum_{j=1}^d Q_{i,j} < 1$.

Let $x = (x_1, \ldots, x_d)$ denote the total amount that the salvage fund allocated to each entity, and $y^* = (y_1^*, \ldots, y_d^*)$ denote the amount of the final settlement. The amount of final settlement is determined by the following optimization problem:

$$y^* = y^*(x, L) = \arg\max\{\mathbf{1}^\top y \,|\, 0 \preceq y \preceq L, \quad (I - Q^\top)y \preceq x\}.$$

In words, the system maximizes the payments subject to the constraint that nobody pays more than what they have (in the final settlement), and nobody pays more than what they owe. Notice that $y^* = y^*(x, L)$ is also a random variable (the randomness comes from L) satisfying $\mathbf{0} \preceq y^* \preceq L$.

Suppose that entity j bankrupts if the deficit $L_j - y_j^* \geq m_j$, where $m \in \mathbb{R}_+^d$ is a given vector. We are interested in finding the minimal amount of salvage fund that ensures no bankruptcy happens with probability at least $1 - \delta$. The problem can be formulated as a chance constraint programming problem as follows:

$$\begin{aligned} &\text{minimize} && \mathbf{1}^\top x \\ &\text{subject to} && \mathrm{P}(L - y^*(x, L) \preceq m) \geq 1 - \delta, \\ &&& x \in \mathbb{R}_{++}^d. \end{aligned} \tag{2}$$

Now we write Problem (2) into standard form. Notice that $L - y^*(x, L) \preceq m$ if and only if $\phi(x, L) \leq 0$, where $\phi(x, L)$ is defined as follows:

$$\phi(x, L) := \min_{b, y}\{b \,|\, (L - y - m) \preceq b \cdot \mathbf{1}, \, (I - Q^\top)y \preceq x, \, y \geq \mathbf{0}\}.$$

Therefore, Problem (2) is equivalent to

$$\begin{aligned} &\text{minimize} && \mathbf{1}^\top x \\ &\text{subject to} && \mathrm{P}(\phi(x, L) > 0) \leq \delta, \\ &&& x \in \mathbb{R}_{++}^d. \end{aligned} \tag{3}$$

3. Review of Scenario Approach

As mentioned in the Introduction, a popular approach to solve the chance constraint problem proceeds by using the scenario approach developed by Calafiore and Campi (2006). They suggest to approximate the probabilistic constraint $\mathrm{P}(\phi(x, L) > 0) \leq \delta$ by N sampled constraints $\phi(x, L^{(i)}) \leq 0$ for $i = 1, \ldots, N$, where $\{L^{(1)}, \ldots, L^{(N)}\}$ are independent samples. Instead of solving the original chance constraint problem (CCP_δ), which is usually intractable, we turn to solve the following optimization problem:

$$\begin{aligned} &\text{minimize} && c^\top x \\ &\text{subject to} && \phi(x, L^{(i)}) \leq 0, \quad i = 1, \ldots, N, \\ &&& x \in \mathbb{R}^{d_x}. \end{aligned} \tag{SP_N}$$

The total sample size N should be large enough to ensure the feasible solution to the sampled problem (SP_N) is also a feasible solution to the original problem (CCP_δ) with a high confidence level. According to Calafiore and Campi (2006), for any given confidence level parameter $\beta \in (0, 1)$, if

$$N \geq \frac{2}{\delta}\log\frac{1}{\beta} + 2d + \frac{2d}{\delta}\log\frac{2}{\delta},$$

then any feasible solution to the sampled optimization problem (SP_N) is also a feasible solution to (CCP_δ) with probability at least $1 - \beta$. However, when δ is small, the total number of sampled constraints is of order $\Omega((1/\delta)\log(1/\delta))$, which could be a problem for implementation. For example, as we shall see in Section 7, when $\beta = 10^{-5}$, $d = 15$ and $\delta = 10^{-3}$, the number of sampled constraints N is required to be larger than 2×10^5. In contrast, our method only requires sampling 2×10^3 constraints.

4. General Algorithmic Idea

To facilitate the development of our algorithm, we introduce some additional notation and a desired technical property (Property 1).

In our setting, a *property* is an *intermediate assumption* that facilitates the construction of an efficient scenario approach algorithm. We shall impose the technical property for now, and in Section 5, we will provide assumptions based on more direct model primitives, providing easy-to-verify sufficient conditions for the properties to hold.

We exploit key intuition borrowed from rare event simulation. A common technique exploited, for example, in Chen et al. (2019), is the construction of a so-called super set, which contains the rare event of interest. The super set should be analytically tractable and be constructed with a probability that is of the same order as that of the rare event of interest. If the conditional distribution given being in the super set is accessible, this can be used as an efficient sampling scheme. The first part of this section simply articulates the elements involved in setting the stage for constructing such a set in the outcome space of L. Later, in Section 5, we will impose assumptions in order to ensure that the probability of the superset, which eventually we will denote by C_δ is suitably controlled as $\delta \to 0$. Simply collecting the elements necessary to construct C_δ requires introducing some super sets involving the decision space, since the optimal decision is unknown.

Let $F_\delta \subseteq \mathbb{R}^{d_x}$ denote the feasible region of the chance constraint optimization problem (CCP$_\delta$), that is,

$$F_\delta := \{x \in \mathbb{R}^{d_x} \,|\, \mathrm{P}(\phi(x, L) > 0) \le \delta\}. \tag{4}$$

Here, the subscript δ is involved to emphasize that the feasible region F_δ is parametrized by the risk level δ. For any fixed $x \in \mathbb{R}^{d_x}$, let $V_x := \{L \in \mathbb{R}^{d_l} \,|\, \phi(x, L) > 0\}$ denote the *violation event at x*.

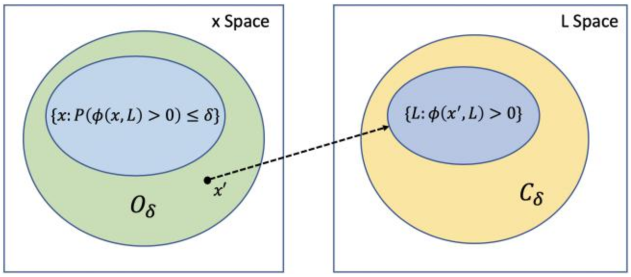
Property 1. *For any $\delta > 0$, there exist a set $O_\delta \subseteq \mathbb{R}^{d_x}$, and an event $C_\delta \subseteq \mathbb{R}^{d_l}$ that satisfy the following statements.*
(a) *The feasible set F_δ is a subset of O_δ.*
(b) *The event C_δ contains the violation event V_x for any $x \in O_\delta$.*
(c) *There exist a constant $M > 0$ independent of δ such that* $\mathrm{P}(L \in C_\delta) \le M \cdot \delta$.

To visualize our intent with Property 1, keep in mind a feature that is often present in heavy-tailed rare-event simulation. In particular, if L is a one-dimensional random variable with, for example, power-law tail decay, then $\mathrm{P}(L > b) \le M \times \mathrm{P}(L > b/2)$ for some $M < \infty$ for all b. For example, if $\mathrm{P}(L > b) = b^{-\alpha}, b \ge 1$ we can take $M = 2^\alpha$. In simple terms, "proportional enlargements" translate into "proportional likelihoods." This sort of feature can be used to motivate the intent of Property 1 and the selection of event C_δ, as it suggests the violation event V_x exhibits "proportional enlargements" when $\delta \to 0$. Specifically, suppose that for some specific $x' \in F_\delta$, we have that $V_{x'} = [b, \infty)$ and the safety constraint is active. That is, $P(L > b) = \delta$ and suppose that the enlarged region is of the form $C_\delta = [b/2, \infty)$. Then, if L is regularly varying we will have that (c) in Property 1 holds for all $\delta > 0$ (which corresponds to all b large). Generally speaking, if $x \in F_\delta$, the set V_x is the set of "bad" outcomes for such a decision. One can imagine that in situations of interest, as we will illustrate, the set of all possible bad outcomes, which is $\cup_{x \in F_\delta} V_x$, can be conveniently enclosed by a region which is a "proportional enlargement" of the set of bad outcomes of a suitable feasible decision (as illustrated in the previous one-dimensional situation). Property 1 implies that the likelihood of the set of all bad outcomes is proportional to the constraint parameter δ. In our algorithms, knowing the constant M will not be relevant, we just need to know that M exists. The sets O_δ and C_δ are auxiliary sets introduced to enclose the set of all possible bad outcomes. We will explain how to construct these sets in examples later.

In the rest of this paper, we will refer to O_δ as *the outer approximation set* and C_δ as *the uniform conditional event*. A graphical illustration of O_δ and C_δ is shown in Figure 1.

As hinted in our earlier discussion that motivates Property 1, we shall focus on the case that L follows a multivariate regularly varying distribution (i.e. a multidimensional version of a power-law-type distribution). The definition of multivariate regular variation is provided in Section 5. In this case, to illuminate how the sets O_δ and C_δ can be constructed for different problems, we provide explicit expressions of them for two running examples in Table 1. In Section 5, we will illustrate how to construct O_δ and C_δ for more general settings.

Figure 1. Illustration of O_δ and C_δ



Now, given O_δ and C_δ that satisfies Property 1, we define the conditionally sampled problem ($\mathrm{CSP}_{\delta,N'}$):

$$\begin{aligned} &\text{minimize} && c^\top x \\ &\text{subject to} && \phi(x, L_\delta^{(i)}) \le 0, \quad i = 1, \ldots, N'. \\ & && x \in O_\delta. \end{aligned} \tag{$\mathrm{CSP}_{\delta,N'}$}$$

Here, $L_\delta^{(i)}$ are independent and identically distributed (i.i.d.) samples generated from the conditional distribution $(L | L \in C_\delta)$.

We now present our main result of this section in Lemma 1, which validates ($\mathrm{CSP}_{\delta,N'}$) is an effective and sample efficient scenario approximation by incorporating (Calafiore and Campi 2006, theorem 2) and Property 1. The proof of Lemma 1 will be presented in Section 4.1.

Lemma 1. *Suppose that Property 1 is imposed and let $\beta > 0$ be a given confidence level.*

1. *Let $\delta' = \delta / \mathrm{P}(L \in C_\delta) \ge 1/M$ and N' be any integer that satisfies*

$$N' \ge \frac{2}{\delta'} \log \frac{1}{\beta} + 2d + \frac{2d}{\delta'} \log \frac{2}{\delta'}. \tag{5}$$

With probability at least $1 - \beta$, if the conditionally sampled problem ($\mathrm{CSP}_{\delta,N'}$) is feasible, then its optimal solution $x_N^* \in F_\delta$ and $\mathrm{Val}(\mathrm{CSP}_{\delta,N'}) \ge \mathrm{Val}(\mathrm{CCP}_\delta)$.

2. *Let N' be any integer such that $N' \le \beta \delta^{-1} \mathrm{P}(L \in C_\delta)$. Assume that the chance constraint problem (CCP_δ) is feasible. Then, with probability at least $1 - \beta$, ($\mathrm{CSP}_{\delta,N'}$) is feasible and $\mathrm{Val}(\mathrm{CCP}_\delta) \ge \mathrm{Val}(\mathrm{CSP}_{\delta,N'})$.*

Remark 1 (Size of Conditionally Sampled Problem). The lower bound of the sample size given in (5) is not greater than $2M \log\left(\frac{1}{\beta}\right) + 2d + 2dM \log(2M)$, which is independent of δ. Therefore, Lemma 1 shows that the chance constraint problem (CCP_δ) can be approximated by ($\mathrm{CSP}_{\delta,N'}$) with sample complexity bounded uniformly as $\delta \to 0$, as long as Property 1 is satisfied.

Remark 2 (Feasibility of Conditionally Sampled Problem). In Lemma 1, part 1, the conditionally sampled problem ($\mathrm{CSP}_{\delta,N'}$) is feasible with high probability if there exists small δ such that (CCP_δ) is feasible. In particular, we claim that

$$\mathrm{P}((\mathrm{CSP}_{\delta,N'}) \text{ is feasible}) \ge (1 - \delta_{\min}/\delta)^{N'}, \tag{6}$$

where $\delta_{\min} = \inf\{\delta \in \mathbb{R}_{++} : \text{Change Constrained problem } (\mathrm{CCP}_\delta) \text{ is feasible}\}$. Recall from Remark 1 that the N' can be chosen to be independent of δ; thus when $\delta_{\min}$ is small, we have ($\mathrm{CSP}_{\delta,N'}$) is feasible with high probability. For example, $\delta_{\min} = 0$ for both the minimal salvage fund problem and the portfolio optimization problem, which implies ($\mathrm{CSP}_{\delta,N'}$) is almost surely feasible for these two examples.

We next prove (6). For arbitrarily small $\epsilon > 0$, the feasible region $F_{\delta_{\min}+\epsilon}$ for problem ($\mathrm{CCP}_{\delta_{\min}+\epsilon}$) is nonempty, and thus we can pick $x \in F_{\delta_{\min}+\epsilon}$ such that $\mathrm{P}(L \in V_x) \le \delta_{\min} + \epsilon$. If $\delta > \delta_{\min} + \epsilon$, then $V_x \subseteq C_\delta$ and thus $\mathrm{P}(L \in V_x | L \in C_\delta) \le (\delta_{\min} + \epsilon)/\mathrm{P}(L \in C_\delta) \le (\delta_{\min} + \epsilon)/\delta$. Therefore, by the independence of samples,

$$\mathrm{P}(x \text{ is feasible for } (\mathrm{CSP}_{\delta,N'})) \ge \mathrm{P}(L \notin V_x | L \in C_\delta))^{N'} \ge (1 - (\delta_{\min} + \epsilon)/\delta)^{N'}.$$

Letting $\epsilon \to 0$, we conclude that (6) holds.

Remark 3 (Efficient Sampling Algorithm). Efficiently generating samples of $(L | L \in C_\delta)$ when $\delta \to 0$ requires rare event simulation techniques. For example, when L is light-tailed, exponential tilting can be applied to achieve $O(1)$ sample complexity uniformly in δ; when L is heavy-tailed, with the help of specific problem structure, one can apply importance sampling (Blanchet and Liu 2010) or Markov chain Monte Carlo (Gudmundsson and Hult 2014) to design an efficient sampling scheme. The specific structure of our salvage fund example results in C_δ being the complement of a box, which makes the sampling very tractable if the element of L are independent.

Even if the aforementioned rare event simulation techniques are hard to apply in practice, we can still apply a simple acceptance-rejection procedure to sample the conditional distribution $(L | L \in C_\delta)$. It costs $O(1/\delta)$ samples of L on average to get one sample of $(L | L \in C_\delta)$ because $\mathrm{P}(L \in C_\delta) = O(\delta)$. Consequently, the total complexity for generating $L_\delta^{(i)}, i = 1, \ldots, N'$ and solving ($\mathrm{CSP}_{\delta,N'}$) is $O(1/\delta)$, which is still much more efficient than the scenario approach in Calafiore and Campi (2006), because it requires computational complexity $O(((1/\delta)\log(1/\delta))^3)$ for solving a linear programming problem with $O((1/\delta)\log(1/\delta))$ sampled constraints by the interior point method.

Although Property 1 seems to be restrictive at first glance, we are still able to construct the sets O_δ and C_δ for a rich class of functions $\phi(x, L)$, including the constraint function for the minimal salvage fund problem. As we shall see in the proof of Lemma 1, once O_δ and C_δ are constructed the sampled problem ($\mathrm{CSP}_{\delta,N'}$) is a tractable approximation to the problem (CCP_δ). We explain how to construct the sets O_δ and C_δ in the next section under some additional assumptions. These assumptions relate in particular to the distribution of L. It turns out that, if L is heavy-tailed, the construction of O_δ and C_δ becomes tractable.

4.1. Proof of Lemma 1

If Property 1 is satisfied, (CCP_δ) is equivalent to

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & \mathrm{P}(\phi(x,L) > 0 \mid L \in C_\delta) \le \delta / \mathrm{P}(L \in C_\delta), \\ & x \in O_\delta \subseteq \mathbb{R}^{d_x}. \end{aligned} \tag{7}$$

Let $\delta' := \delta/\mathrm{P}(L \in C_\delta) \ge 1/M$ denote the risk level in the equivalent problem (7). The sampled optimization problem related to Problem (7) is given by

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & \phi(x, L_\delta^{(i)}) \le 0, \quad i = 1, \dots, N', \\ & x \in O_\delta, \end{aligned} \tag{$\mathrm{CSP}_{\delta,\mathrm{N}'}$}$$

where the $L_\delta^{(i)}$ are independently sampled from $\mathrm{P}(\cdot \mid L \in C_\delta)$. Notice that

$$N' \ge \frac{2}{\delta'} \log \frac{1}{\beta} + 2d + \frac{2d}{\delta'} \log \frac{2}{\delta'}.$$

According to Calafiore and Campi (2006, corollary 1 and theorem 2), with probability at least $1-\beta$, if the sampled problem ($\mathrm{CSP}_{\delta,N'}$) is feasible, then the optimal solution to problem ($\mathrm{CSP}_{\delta,N'}$) is feasible to the chance constraint problem (7). Because (7) and (CCP_δ) are equivalent, the optimal solution to problem ($\mathrm{CSP}_{\delta,N'}$) is also feasible to (CCP_δ). The proof of the first part of the lemma is complete.

Now we turn to prove the second part of the lemma. The equivalence between (CCP_δ) and (7) is still valid, so it is sufficient to compare the optimal values of (7) and ($\mathrm{CSP}_{\delta,N'}$). By applying Calafiore and Campi (2006, theorem 2) again, we have with probability at least $1-\beta$ ($\mathrm{CSP}_{\delta,N'}$) is feasible and the value of ($\mathrm{CSP}_{\delta,N'}$) is no larger than the optimal value of

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & \mathrm{P}(\phi(x,L) > 0 \mid L \in C_\delta) \le 1 - (1-\beta)^{1/N'}, \\ & x \in O_\delta \subseteq \mathbb{R}^{d_x}. \end{aligned} \tag{8}$$

The proof is complete by using $1-(1-\beta)^{1/N'} \ge \beta/N' \ge \frac{\delta}{\mathrm{P}(L \in C_\delta)}$. Therefore, using Val for "value of," $\mathrm{Val}(8) \le \mathrm{Val}(7) = \mathrm{Val}(\mathrm{CCP}_\delta)$.

5. Constructing Outer Approximations and Summary of the Algorithm

In this section, we come full circle with the intuition borrowed from rare event simulation explained at the beginning of Section 4. The scale-free properties of heavy-tailed distributions (to be reviewed momentarily) coupled with natural (polynomial) growth conditions (like the linear loss) given by the structure of the optimization problem, provide the necessary ingredients to show that the set C_δ has a probability that is of order $O(\delta)$.

In the discussion immediately following Property 1, we imagined that the uniform conditional set $L \in C_\delta$ was of the form $L \in [b/2, \infty)$ for $b \to \infty$ as $\delta \to 0$. However, Property 1 can still be enforced if this statement applies to L^2 or any power of L. This is because power law-type decay (and more generally regular variation) is preserved under power transformations. We will provide assumptions that will enforce that regular variation properties can be applied when estimating the likelihood of the uniform conditional event.

We assume that the distribution of L is of multivariate regular variation. A definition that we now review. For background, we refer to Resnick (2013). Let $\mathscr{M}_+(\overline{\mathbb{R}}^{d_l} \backslash \{\mathbf{0}\})$ denote all Radon measures on the space $\overline{\mathbb{R}}^{d_l} \backslash \{\mathbf{0}\}$ (recall that a measure is Radon if it assigns finite mass to all compact sets). If $\mu_n(\cdot), \mu(\cdot) \in \mathscr{M}_+(\overline{\mathbb{R}}^{d_l} \backslash \{\mathbf{0}\})$, then μ_n converges to μ vaguely, denoted by $\mu_n \xrightarrow{v} \mu$, if for all compactly supported continuous functions $f : \overline{\mathbb{R}}^{d_l} \backslash \{\mathbf{0}\} \to \mathbb{R}_+$,

$$\lim_{n\to\infty} \int_{\overline{\mathbb{R}}^{d_l} \backslash \{\mathbf{0}\}} f(x) \mu_n(dx) = \int_{\overline{\mathbb{R}}^{d_l} \backslash \{\mathbf{0}\}} f(x) \mu(dx).$$

L is *multivariate regularly varying* with *limit measure* $\mu(\cdot) \in \mathscr{M}_+(\overline{\mathbb{R}}^{d_l} \backslash \{\mathbf{0}\})$ if

$$\frac{\mathrm{P}(x^{-1}L \in \cdot)}{\mathrm{P}(\|L\|_2 > x)} \xrightarrow{v} \mu(\cdot), \qquad \text{as } x \to \infty.$$

Assumption 1. *L is multivariate regularly varying with limit measure $\mu(\cdot) \in \mathscr{M}_+(\overline{\mathbb{R}}^{d_l} \backslash \{\mathbf{0}\})$.*

Here are some intuitions behind the definition of multivariate regularly varying. Suppose that L is written in terms of polar coordinates, with R being the radius and Θ being a random variable taking values on the unit sphere. The radius $R = \|L\|_2$ has a one-dimensional regularly varying tail (i.e., we can write $\mathrm{P}(R > x) = L(x)x^{-\alpha}$ for a slowly varying function L and $\alpha > 0$). The angle Θ, conditioned on R being large, converges weakly (as $R \to \infty$) to a limiting random variable. The distribution of this limit can be expressed in terms of the measure μ. For a recent application of multivariate regular variation in operations research, see Kley et al. (2016).

In this section, we present two methods for the construction of O_δ and C_δ satisfying Property 1. We mostly focus on our "scaling method" which is presented in Section 5.1, which is facilitated precisely by the scale-free property that we will impose on L. After showing the construction of the outer sets under the scaling method, we summarize the algorithm at the end of Section 5.1. We supply a lower bound guaranteeing a constant approximation for the output of the algorithm in Section 5.2. Our second method for outer approximation constructions is summarized in Section 5.3. This method is simpler to apply because is based on linear approximations; however, it is less general because it assumes that $\phi(x, L)$ is jointly convex.

5.1. Scaling Method

We start by analyzing the feasible region F_δ when $\delta \to 0$. Intuitively, if the violation probability $\mathrm{P}(\phi(x, L) > 0)$ has a strictly positive lower bound in any compact set, then F_δ will ultimately be disjoint with the compact set when $\delta \to 0$. Thus, the set F_δ is expelled to infinity when $\delta \to 0$ in this case. F_δ is moving toward the direction that $\phi(x, L)$ becomes small such that the violation probability becomes smaller. For instance, if x is one dimensional and $\phi(x, L)$ is increasing in x, then F_δ is moving toward the negative direction. Consider the portfolio optimization problem as another example, in which $\min_{j=1}^d x_j \to +\infty$ as $\delta \to 0$.

With such intuition, now we begin to construct the outer approximation set O_δ. To this end, we need to introduce an auxiliary function which we shall call a *level function*. We assume the existence of a level function in Assumption 2, and the level function needs to be explicitly computable to construct the outer approximation set O_δ.

Definition 1. We say that $\pi : \mathbb{R}^{d_x} \to [0, +\infty]$ is a level function if
1. For any $\alpha \geq 0$ and $x \in \mathbb{R}^{d_x}$, we have $\pi(\alpha \cdot x) = \alpha \cdot \pi(x)$,
2. The function $\pi(\mathrm{x})$ is coersive, i.e., $\lim_{\delta \to 0} \inf_{x \in F_\delta} \pi(x) + \infty$.

For a given level function π, we define its *unit level set* as $\Pi = \{x \in \mathbb{R}^{d_x} \,|\, \pi(x) = 1\}$.

Assumption 2. *There exists an explicitly computable level function π and its unit level set Π.*

The unit level set Π is used to characterize the moving direction of F_δ as $\delta \to 0$. The shape of Π is chosen in accordance with the moving direction of F_δ to reduce the size of O_δ to achieve better sample complexity. The outer approximation set O_δ is constructed as

$$O_\delta := \bigcup_{\alpha \geq \alpha_\delta} (\alpha \cdot \Pi) \supseteq F_\delta,$$

where α_δ characterizes the scaling rate of F_δ. We will explain how to choose α_δ in the proof of Lemma 2. Here are several examples of the level functions and unit level sets:

- Suppose that $\phi(x, L) = -\|x\|^2 - L$, then the level function π can be chosen as the Euclidean norm and Π can be chosen as the unit sphere in $\mathbb{R}^{d_x}$.
- For the portfolio optimization problem, the level function can be chosen as $\pi(x) = \min_{j=1}^d x_j + \infty \cdot I(x \notin \mathbb{R}_{++}^{d_x})$ in accordance with our intuition that $\min_{j=1}^d x_j \to \infty$, and the unit level set can be chosen as $\Pi = \{x \in \mathbb{R}^{d_x} \,|\, \min_{j=1}^d x_j = 1\}$.

To analyze the asymptotic shape of the uniform conditional event C_δ, we connect the asymptotic distribution of L to the asymptotic distribution of $\phi(x, L)$. Keep in mind that we wish to preserve the scaling property of the tail of L, when considering $\phi(x, L)$, so that Property 1 can be ensured. We pick a continuous nondecreasing function $h : \mathbb{R}_{++} \to \mathbb{R}_{++}$ such that $\lim_{\alpha \to +\infty} h(\alpha) = +\infty$ to characterize the scaling rate of L. In addition, we pick another positive function $r : \mathbb{R}_{++} \to \mathbb{R}_{++}$ to characterize the scaling rate of $\phi(\alpha \cdot x, h(\alpha) \cdot L)$. Intuitively, the scaling function $r(\cdot)$ and $h(\cdot)$ should ensure the condition that the collection of probability measures of $\left\{\frac{1}{r(\alpha)}\phi(\alpha \cdot x, h(\alpha) \cdot L)\right\}_{\alpha \geq 1}$ is tight. For the minimal salvage fund problem with fixed δ, as the deficit $\phi(x, L)$ is asymptotically linear with respect to the salvage fund x and the loss L, we can simply pick $r(\alpha) = h(\alpha) = \alpha$ in this problem. We next introduce two auxiliary functions Ψ_+ and Ψ_-.

Definition 2. Let $\Psi_+ : \mathbb{R}^{d_l} \to \mathbb{R}$, $\Psi_- : \mathbb{R}^{d_l} \to \mathbb{R}$ be two Borel measurable functions. We say Ψ_+ (respectively, c) is the asymptotic uniform upper (respectively, lower) bound of $\frac{1}{r(\alpha)}\phi(\alpha \cdot x, h(\alpha) \cdot l)$ over the unit level set $x \in \Pi$ if for any compact set $K \subseteq \mathbb{R}^{d_l}$,

$$\liminf_{\alpha\to\infty} \inf_{l\in K} \left(\Psi_+(l) - \sup_{x\in\Pi} \left[\frac{1}{r(\alpha)} \phi(\alpha \cdot x, h(\alpha) \cdot l) \right] \right) \geq 0, \tag{9a}$$

$$\limsup_{\alpha\to\infty} \sup_{l\in K} \left(\Psi_-(l) - \inf_{x\in\Pi} \left[\frac{1}{r(\alpha)} \phi(\alpha \cdot x, h(\alpha) \cdot l) \right] \right) \leq 0. \tag{9b}$$

We would like to have lower and upper bounds Ψ_- and Ψ_+ so that ϕ is of the order $r(\alpha)$ for every decision x, and in every direction l of the random vector L, whenever the norm of the latter is large (i.e., of the size $h(\alpha)$). A stronger assumption would have been to require an actual limit (rather than a liminf and a limsup), but this is not needed to provide big-O bounds for the complexity of our algorithm.

The functions Ψ_+ and Ψ_- are used to define the event $C_{\varepsilon,-}$ and $C_{\varepsilon,+}$, which serve as the inner and outer approximation of the event $\cup_{x\in\Pi} V_x$, where $V_x = \{l \in \mathbb{R}^{d_l} \mid \phi(x, l) > 0\}$ is the violation event at x.

Definition 3. For $\varepsilon > 0$, let $C_{\varepsilon,+}$ (respectively, $C_{\varepsilon,-}$) be the ε-outer (respectively, inner) approximation event:

$$C_{\varepsilon,+} := \{l \in \mathbb{R}^{d_l} \mid \Psi_+(l) \geq -\varepsilon\}, \tag{10a}$$

$$C_{\varepsilon,-} := \{l \in \mathbb{R}^{d_l} \mid \Psi_-(l) \geq +\varepsilon\}. \tag{10b}$$

We now define $O_\delta := \cup_{\alpha\geq\alpha_\delta} \alpha \cdot \Pi$. The following property ensures that the shape of Π is appropriate and α_δ is large enough; hence, O_δ is an outer approximation of F_δ.

Property 2. *There exist δ_0 such that for any $\delta < \delta_0$, we have an explicitly computable constant α_δ that satisfies*

$$\mathrm{P}(\|L\|_2 > h(\alpha_\delta)) = O(\delta) \quad \text{and} \quad F_\delta \subseteq \bigcup_{\alpha\geq\alpha_\delta} \alpha \cdot \Pi =: O_\delta.$$

If the violation probability is easy to analyze, we will directly derive the expression of α_δ and verify Property 2. Otherwise, we resort to Lemma 2, which provides a sufficient condition of Property 2 by analyzing the asymptotic probability of the violation event as $\delta \to 0$.

Lemma 2. *Suppose that Assumptions 1 and 2 hold. If there exists an asymptotic uniform lower bound function $\Psi_-(\cdot)$ as given in (9b) and $\varepsilon > 0$ such that $\mu(C_{\varepsilon,-}) > 0$, then Property 2 is satisfied.*

The high-level idea of Lemma 2 is to show that F_δ is disjoint from $\alpha \cdot \Pi$ for small α. To this end, the asymptotic scaling (9b) is used to demonstrate that the violation probability is no less than $\mathrm{P}(L \in h(\alpha) \cdot C_{\varepsilon,-})$, which is approximately equal to $\mathrm{P}(\|L\|_2 > h(\alpha_\delta)) \cdot \mu(C_{\varepsilon,-})$ according to the regularly varying property of L. The detailed proof of Lemma 2 is deferred to Appendix A.1.

We impose the following Assumption 3 on the asymptotic uniform upper bound $\Psi_+(\cdot)$ so that we can use the multivariate regular variation of L to estimate $\mathrm{P}(L \in \alpha \cdot C_{\varepsilon,+})$ for large scaling factor α.

Assumption 3. *There exist an event $S \subseteq \mathbb{R}^{d_l}$ with $\mu(S^c) < \infty$ such that*

$$S \subseteq \alpha \cdot S, \qquad \Psi_+(l) \leq \Psi_+(\alpha \cdot l), \qquad \forall l \in S,\ \alpha \geq 1.$$

In addition, there exist some $\varepsilon > 0$ such that $C_{\varepsilon,+}$ is bounded away from the origin, that is, $\inf_{l\in C_{\varepsilon,+}} \|l\|_2 > 0$.

Moreover, both S and $C_{\varepsilon,+}$ have explicit expressions.

For the minimal salvage fund problem, because the deficit function $\phi(x, L)$ is coordinate-wise nondecreasing with respect to the loss vector L, it is reasonable to assume that its asymptotic bound $\Psi_+(\cdot)$ is also coordinate-wise nondecreasing. For this example, the closed form expression of $\Psi_+(\cdot)$ and the detailed verification of all the assumptions are deferred to Proposition 2. Our next result summarizes the construction of the outer approximation sets.

Theorem 1. *Suppose that Property 2 and Assumption 3 are imposed. Then there exist $\delta_0 > 0$ such that the following sets*

$$O_\delta = \bigcup_{\alpha \geq \alpha_\delta} \alpha \cdot \Pi, \qquad C_\delta = h(\alpha_\delta) \cdot (C_{\varepsilon,+} \cup K^c \cup S^c) \tag{11}$$

satisfy Property 1 for all $\delta < \delta_0$. Here, S is given in Assumption 3 and K is a ball in $\mathbb{R}^{d_l}$ with $\mu(K^c) < \infty$.

The main idea for proving Theorem 1 is as follows: If L lies in a "well-behaved" compact region, then by applying Assumption 3 and the asymptotic uniform lower bound (9b), the violation events $\cup_{x \in O_\delta} V_x$ is uniformly enclosed in $h(\alpha_\delta) \cdot C_{\varepsilon,+}$. Otherwise L lies in the "ill-behaved" region $h(\alpha_\delta) \cdot (K^c \cup S^c)$. Combining these two cases inspires that definition of $C_\delta := h(\alpha_\delta) \cdot (C_{\varepsilon,+} \cup K^c \cup S^c)$, and the probability of $L \in C_\delta$ is $O(\delta)$ because of $\mathrm{P}(\|L\|_2 > h(\alpha_\delta)) = O(\delta)$ and L is regularly varying. The detailed proof of Theorem 1 is deferred to Appendix A.1.

With the aid of Lemma 1 and Theorem 1, we provide an algorithm for approximating (CCP$_\delta$) in which the sampled optimization problem is bounded in $1/\delta$.

Algorithm 1 (Scenario Approach with Optimal Scenario Generation)

input: Risk tolerance parameter δ, confidence level β, and all the elements and constants appearing in Property 2 and Assumption 3, including level function π or unit level set Π, constant α_δ, scaling function h, and explicit expression of $C_{\varepsilon,+}, K$ and S.

1 Compute the expression of sets O_δ and C_δ by (11);
2 Compute required number of samples N' by (5);
3 **for** $i = 1, \ldots, N'$ **do**
4 | Sample $L_\delta^{(i)}$ using acceptance-rejection or importance sampling.
5 end
6 Solve the conditionally sampled problem (CSP$_{\delta,N'}$).

5.2. Constant Approximation Guarantee

In Section 5.2, our objective is to show that the output of the previous algorithm is guaranteed to be within a constant factor of the optimal solution to (CCP$_\delta$) with high probability, uniformly in δ.

We shall work under the setting of Theorem 1, so we enforce Property 2 and Assumptions 3. We want to show that there exist some constant $\Lambda > 1$ independent of δ, such that $\mathrm{Val}(\mathrm{CCP}_\delta) \leq \mathrm{Val}(\mathrm{CSP}_{\delta,N'}) \leq \Lambda \times \mathrm{Val}(\mathrm{CCP}_\delta)$ with high probability. This indicates that our result guarantees a constant approximation to (CCP$_\delta$) for regularly varying distributions (under our assumptions) in $O(1)$ sample complexity when $\delta \to 0$ with high probability.

Note that $(\mathrm{CSP}_{\delta,N'}) \leq \Lambda \times \mathrm{Val}(\mathrm{CCP}_\delta)$ is meaningful only if $\mathrm{Val}(\mathrm{CCP}_\delta) > 0$. We assume that the outer approximation set is good enough such that the following natural assumption is valid.

Assumption 4. *There exist $\delta > 0$ such that* $\min_{x \in O_\delta} c^\top x > 0$.

The previous assumption will typically hold if c has strictly positive entries. Theorem 1 and the form of O_δ guarantee that the norm of the optimal solution of (CSP$_{\delta,N'}$) grows in proportion to α_δ, so we also assume the following scaling property for $\phi(x, l)$.

Assumption 5. *There exist a function $\phi_{\lim} : (\mathbb{R}^{d_x} \backslash \{\mathbf{0}\}) \times (\mathbb{R}^{d_l} \backslash \{\mathbf{0}\}) \to \mathbb{R}$ such that for every compact set $E \subseteq \mathbb{R}^{d_l} \backslash \{\mathbf{0}\}$, we have*

$$\lim_{\alpha \to \infty} \sup_{l \in E} \left| \frac{1}{r(\alpha)} \phi(\alpha \cdot x, h(\alpha) \cdot l) - \phi_{\lim}(x, l) \right| = 0.$$

In addition, $\phi_{\lim}(x, l)$ is continuous in one.

Assumption 5 is satisfied by both running examples. For the portfolio optimization problem, we have $\phi(x, l) = \sum_{j=1}^{d} (L_j / x_j) - \eta$, thus $\phi_{\lim}(x, l) = \phi(x, l)$. For the minimal salvage fund problem, we have $\phi_{\lim}(x, l) = \phi(x, l) - m$ such that $|\alpha^{-1}\phi(\alpha \cdot x, \alpha \cdot l) - \phi_{\lim}(x, l)| \leq \alpha^{-1} m$ and $|\phi_{\lim}(x, l) - \phi_{\lim}(x, l')| \leq \|l - l'\|_1$.

We define the following optimization problem, which will serve as an asymptotic upper bound of (CSP$_{\delta,N'}$) in stochastic order when $\delta \to 0$:

$$\begin{aligned} &\text{minimize} && c^\top x \\ &\text{subject to} && \phi_{\lim}(x, L_{\lim^{(i)}}) \leq 0, \quad i = 1, \ldots, N', \\ & && x \in \bigcup_{\alpha \geq 1} \alpha \cdot \Pi, \end{aligned} \tag{CSP$_{\mathrm{lim},N'}$}$$

where $L_{\lim^{(i)}}$ are i.i.d. samples from a random variable $L_{\lim}$, whose distribution is characterized by $P(L_{\lim} \in (C_{\varepsilon,+} \cup K^c \cup S^c)) = 1$ and $P(L_{\lim} \in E) = \mu(E)/\mu(C_{\varepsilon,+} \cup K^c \cup S^c)$ for all measurable set $E \subseteq C_{\varepsilon,+} \cup K^c \cup S^c$.

Theorem 2. *Let $\beta > 0$ be a given confidence level and N' be a fixed integer that satisfies (5). If Assumptions 4 and 5 are enforced, and ($\mathrm{CSP}_{\lim,N'}$) satisfies Slater's condition with probability one, then there exist $\delta_0 > 0$ and $\Lambda > 0$ such that*

$$P(\mathrm{Val}(\mathrm{CCP}_\delta) \le \mathrm{Val}(\mathrm{CSP}_{\delta,N'}) \le \Lambda \times \mathrm{Val}(\mathrm{CCP}_\delta)) \ge 1 - 2\beta, \qquad \forall \delta < \delta_0.$$

In Theorem 2, the Slater's condition (see section 5.2.3 in Boyd and Vandenberghe (2004) for reference) can be verified directly on the problem ($\mathrm{CSP}_{\lim,N'}$). This condition is satisfied in the salvage fund problem by standard linear programming duality. We also remark that Assumptions 4 and 5 only require the existence rather than the explicit knowledge of $\{\delta | \min_{x \in O_\delta} c^\top x > 0\}$ and function $\phi_{\lim}$.

5.3. Linear Approximation Method

Suppose that the constraint function $\phi(x,l)$ is jointly convex in (x, l), and L is multivariate regularly varying. We will develop a simpler method in this section to construct the outer approximation set O_δ and the uniform conditional event C_δ.

We first introduce a crucial assumption in the construction of O_δ and C_δ.

Assumption 6. *There exist a convex piecewise linear function $\phi\_(x,l) : \mathbb{R}^{d_x} \times \mathbb{R}^{d_l} \to \mathbb{R}$ of the form*

$$\phi\_(x,l) = \max_{j=1,\dots,N} a_j^\top l + b_j^\top x + c_j, \qquad a_j \in \mathbb{R}^{d_l}, b_j \in \mathbb{R}^{d_x} \text{ and } c_j \in \mathbb{R} \text{ for } j = 1,\dots,N.$$

such that

1. The inequality $\varphi\_(x,l) \le \varphi(x,l)$, holds for all $(x,l) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_l}$;
2. *There exist some constant $C \in \mathbb{R}_+$ such that $\phi(x,l) \le 0$ if $\phi\_(x,l) \le -C$.*

If $\phi(x,l)$ itself is a piecewise affine function, then Assumption 6 is satisfied by simply taking $\phi\_(x,l) = \phi(x,l)$. For general jointly convex functions, the following lemma verifies Assumption 6 if $\phi(x,l)$ has a compact zero sublevel set.

Lemma 3. *If the constraint function $\phi(x,L) : \mathbb{R}^{d_x} \times \mathbb{R}^{d_l} \to \mathbb{R}$ is convex and twice continuously differentiable, and it has a compact zero sublevel set $Z_\phi := \{(x,l) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_l} | \phi(x,l) \le 0\}$, then Assumption 6 is satisfied.*

With Assumption 6 enforced, we are now ready to provide our main result in this section to fully summarize the construction of O_δ and C_δ.

Theorem 3. *If Assumptions 1 and 6 hold, we can construct O_δ and C_δ that satisfy* Property 1 *as*

$$O_\delta := \bigcap_{j=1}^{N} \{x \in \mathbb{R}^{d_x} | b_j^\top x + c_j + \overline{F}^{-1}_{a_j^\top L}(\delta) \le 0\}, \qquad C_\delta := \bigcup_{j=1}^{N} \{L \in \mathbb{R}^{d_l} | a_j^\top L + C > \overline{F}^{-1}_{a_j^\top L}(\delta)\},$$

where $\overline{F}^{-1}_{a_j^\top L}(\delta) = \inf\{x \in \mathbb{R} | P(x > a_j^\top L) \le \delta\}$.

6. Verifying the Assumptions in Examples

In this section, we verify the elements required to apply our algorithm. We provide explicit expressions for sets O_δ and C_δ in the statement of the propositions. The detailed verification process and the steps for constructing sets O_δ and C_δ are presented as the proofs in Appendix A.2.

6.1. Portfolio Optimization with VaR Constraint

In this section, we will verify that Theorem 1 is applicable to an equivalent form of the portfolio optimization problem (2).

Proposition 1. *The portfolio optimization problem (1) satisfies all assumptions required by Theorem 1, such that the sets O_δ and C_δ admit the following explicit expressions:*

$$O_\delta = \{x \in \mathbb{R}^d_{++} | \eta \cdot x \ge \overline{F}^{-1}_{\mathbf{1}^\top L}(\delta)\}, \qquad C_\delta = \{l \in \mathbb{R}^d_{++} | 2 \cdot \mathbf{1}^\top l \ge \overline{F}^{-1}_{\mathbf{1}^\top L}(\delta)\}.$$

6.2. Minimal Salvage Fund

The key observation to solve the minimal salvage fund problem (3) is the following lemma, which provides a closed form piecewise linear expression for the constraint function $\phi(x,L)$.

Lemma 4. *In the minimal salvage fund problem* (3), *we have*

$$\phi(x, L) = \max_{j=1,\ldots,d} L_j - \mathbf{e}_j^\top (I - Q^\top)^{-1} x - m_j,$$

where $\mathbf{e}_j$ *denote the unit vector on the jth coordinate.*

Now we prove that Theorem 3 is applicable to the minimal salvage fund problem (4).

Proposition 2. *The minimal salvage fund problem* (3) *satisfies all assumptions required by Theorem* 3, *such that the sets* O_δ *and* C_δ *admit the following explicit expressions:*

$$O_\delta = \bigcap_{j=1}^{d} \{x \in \mathbb{R}^d \,|\, \overline{F}_{L_j}^{-1}(\delta) \le \mathbf{e}_j^\top (I - Q^\top)^{-1} x + m_j\}, \qquad C_\delta = \bigcup_{j=1}^{d} \{l \in \mathbb{R}^d \,|\, l_j > \overline{F}_{L_j}^{-1}(\delta)\}.$$

6.3. Quadratic Model

In this section, we consider a model with a quadratic control term in x as an additional example. Suppose that the constraint function $\phi(x, l) : \mathbb{R}^{d_x} \times \mathbb{R}^{d_l} \to \mathbb{R}$ is defined as

$$\phi(x, l) = x^\top Q x + x^\top A l, \tag{12}$$

where $Q \in \mathbb{R}^{d_x \times d_x}$ is a symmetric matrix and $A \in \mathbb{R}^{d_x \times d_l}$ is a matrix with $\mathrm{rank}(A) = d_x$; that is, there exists $\sigma > 0$ such that $\|A^\top x\|_2 \ge \sigma \|x\|_2$.

Proposition 3. *Consider the chance constraint optimization model with constraint function defined as* (12).

1. *If Q is a positive semidefinite matrix and L has a positive density, there exist some* δ *such that the problem is infeasible.*
2. *If Q has a negative eigenvalue and L is multivariate regularly varying, the model satisfies all the assumptions required by Theorem* 1.

7. Numerical Experiments

To empirically study the computational complexity and compare the quality of the solutions, in this section, we conduct numerical experiments for two scenario generation algorithms:

1. The efficient scenario generation approach proposed in this paper (abbreviated as Eff-Sc)
2. The scenario approach in Calafiore and Campi (2006) (abbreviated as CC-Sc)

In Section 7.1, we present the results for the portfolio optimization problem. In Section 7.2, we present the results for the minimal salvage fund problem. The numerical experiment is conducted using a Laptop with a 2.2-GHz Intel Core i7 CPU, and the sampled linear programming problem is solved using CVXPY (Diamond and Boyd 2016) with the MOSEK solver (MOSEK ApS 2020).

7.1. Portfolio Optimization with VaR Constraint

First, we present the parameter selection and the implement details for the numerical experiment of portfolio optimization problem (1). Suppose that there are $d = 10$ assets to invest, and the parameters of the problem are chosen as follows:

- The mean return vector is $\mu = (1.0, 1.5, 2.0, 2.5, 3, 1.6, 1.2, 1.1, 1.8, 2.2)$.
- The random variable L_j are i.i.d. with Pareto cumulative distribution function $\mathrm{P}(L_j > l) = (\ell_j / l)$, for $l \ge \ell_j$.
- The parameter $\ell = (\ell_1, \ldots, \ell_d) = (2.1, 1.3, 1.6, 2.5, 2.7, 1.3, 1.9, 1.5, 2.2, 2.3)$.
- The loss threshold $\eta = 1{,}000$.

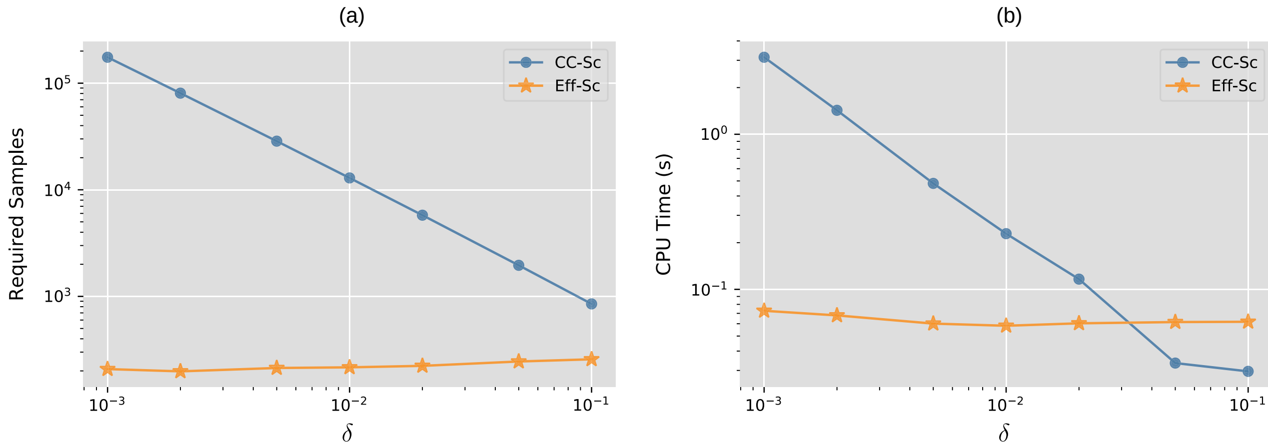
Now we explain the implementation detail of Eff-Sc. Recall the expression of O_δ and C_δ from Proposition 1, which involves the analytically unknown quantity $\overline{F}_{\mathbf{1}^\top L}^{-1}(\delta)$. Because quantile estimation is much more computationally efficient than solving the sampled optimization problem, we generate samples of L to estimate a confidence interval of $\overline{F}_{\mathbf{1}^\top L}^{-1}(\delta)$ with large enough confidence level $1 - o(\beta)$, and we denote the resulting confidence interval by $(\widehat{\mathrm{LB}}, \widehat{\mathrm{UB}})$. We replace the expressions of O_δ and C_δ by their sampled version conservative approximations, that is,

$$O_\delta = \{x \in \mathbb{R}_{++}^d \,|\, \eta \cdot x \ge \widehat{\mathrm{UB}}\}, \qquad C_\delta = \{l \in \mathbb{R}_{++}^d \,|\, 2 \cdot \mathbf{1}^\top l \ge \widehat{\mathrm{LB}}\}.$$

The value of $\mathrm{P}(L \in C_\delta)$ is also estimated using the generated samples. We compute the required number of samples N' using Lemma 1, and the samples of L_δ is generated via acceptance-rejection.

In Figure 2, we compare the efficiency between Eff-Sc and CC-Sc. Figure 2(a) presents the required number of samples for both algorithms, in which one can quickly remark that Eff-Sc requires significantly fewer samples than CC-Sc, especially for the problems with small δ. In Figure 2(b), we compare the running time for both models. Whereas Eff-Sc costs slightly more time for δ around 0.1 due to the overhead cost of computing O_δ and C_δ,

Figure 2. Comparison of Computational Efficiency for the Portfolio Optimization Problem



Notes. (a) Terms of the required number of samples. (b) Used CPU time. We test $\delta \in \{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1\}$.

the computational time stays nearly constant uniformly in δ, indicating that Eff-Sc is a substantially more efficient algorithm than CC-Sc.

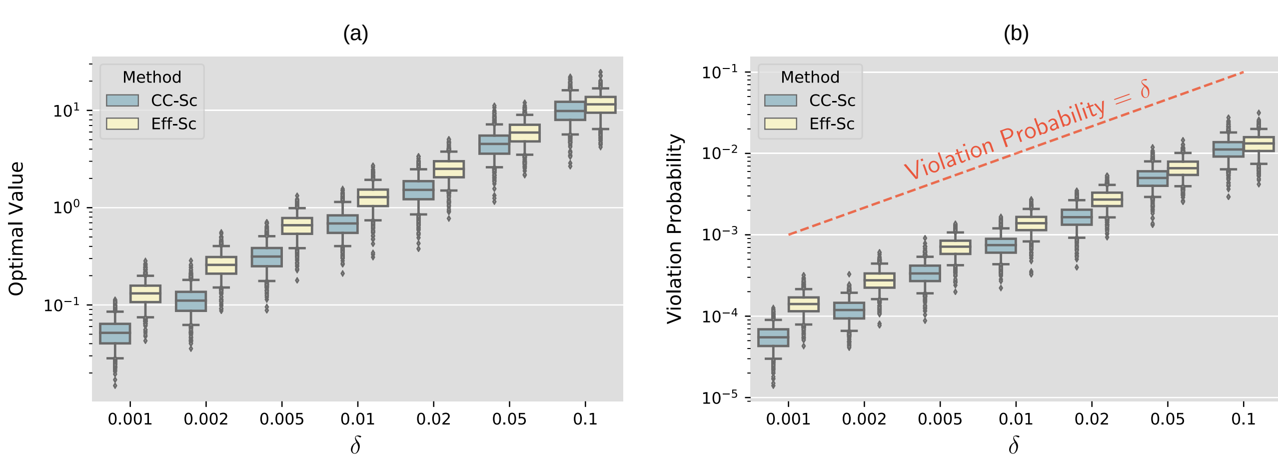
Finally, we compare Eff-Sc and CC-Sc for the optimal values of the sampled problems and the violation probabilities of the optimal solutions. Because both methods require generating random samples, the generated solutions are also random. Thus, the optimal values and the violation probabilities are also random. To compare the distributions of the random quantities, we conduct 10^3 independent experiments. In each experiment, we execute both algorithms and get two solutions, then we evaluate the solutions' violation probabilities using 10^6 samples of L. We use boxplots (McGill et al. 1978) to depict the samples' distribution through their quantiles. A boxplot is constructed of two parts: a box and a set of whiskers. The box is drawn from the 25% quantile to the 75% quantile, with a horizontal line drawn in the middle to denote the median. Two whiskers indicate 5% and 95% quantiles, respectively, and the scatters represent all the rest sample points beyond the whiskers.

In Figure 3, we present (a) the optimal values and (b) the violation probabilities. One can quickly remark from Figure 3(a) that the optimal value of Eff-Sc is stochastically larger than the optimal value of CC-Sc, whereas Figure 3(b) indicates that the optimal solutions produced by both methods are feasible for all the 10^3 experiments. Overall, with both methods successfully and conservatively approximating the probabilistic constraint, Eff-Sc is more computationally efficient and less conservative, producing solutions with better objective values than its counterpart.

7.2. Minimal Salvage Fund

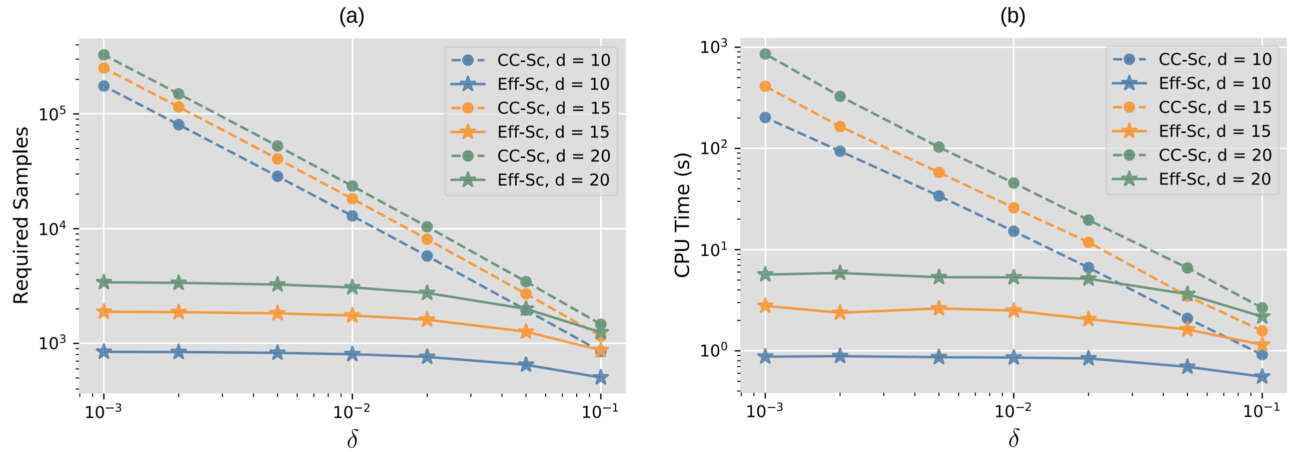
In this section, we conduct a numerical experiment for the minimal salvage fund problem (3). In the experiment, we pick $d \in \{10, 15, 20\}$ to test the performance of the problem in different dimensions.

Figure 3. Comparison of the Quality of Optimal Solutions for the Portfolio Optimization Problem



Notes. (a) Optimal value. (b) Solutions' violation probabilities. Here $\delta \in \{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1\}$, and the box plots are generated using 1,000 experiments.

Figure 4. Comparison of Computational Efficiency for the Minimal Salvage Fund Problem



Notes. (a) Required number of samples. (b) Used CPU time. We test $d \in \{10, 15, 20\}$ and $\delta \in \{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1\}$.

For each fixed d, the parameters of Problem (3) are chosen as follows:

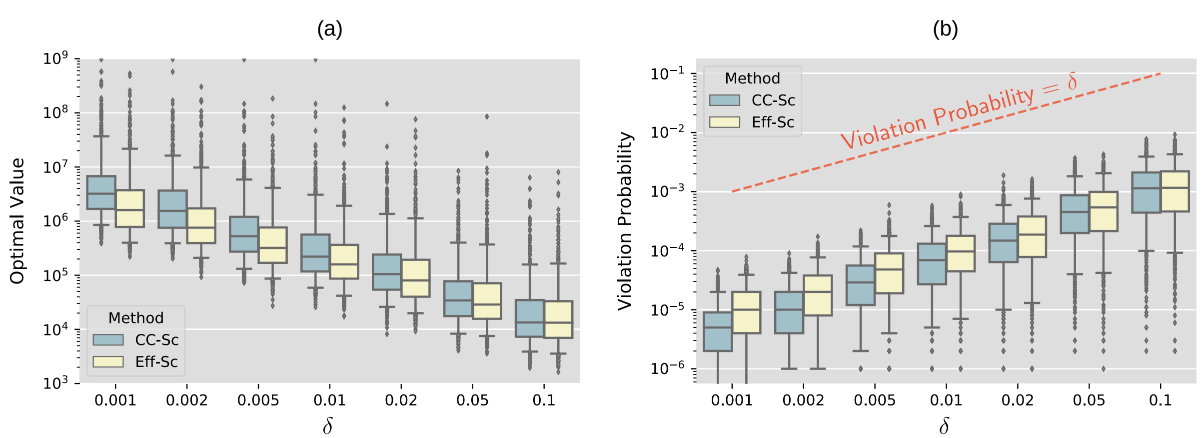
- The matrix $Q = (Q_{i,j} : i, j \in \{1, \ldots, d\})$ where $Q_{i,j} = 1/d$ if $i \neq j$ and otherwise $Q_{i,j} = 0$.
- The vector $m = (m_j : j \in \{1, \ldots, d\})$ where $m_j = 10$ for each j.
- The random variables L_j are i.i.d. with Pareto cumulative distribution function $\mathrm{P}(L_j > l) = (1/l)$, for $l \geq 1$.

Recall the explicit expressions for sets O_δ and C_δ from Proposition 2. To solve the conditionally sampled problem ($\mathrm{CSP}_{\delta,N'}$), it remains to sample $L_\delta^{(i)}$ and compute N', the required number of samples. When δ is small, when $\delta \leq 10^{-3}$, solving the optimization problem ($\mathrm{CSP}_{\delta,N'}$) costs much more time than simulating $L_\delta^{(i)}$, despite that a simple acceptance rejection scheme is applied to sample $L_\delta^{(i)}$ in our experiments. We fix the confidence level parameter $\beta = 10^{-5}$ and set $\delta' = \delta/P(L \in C_\delta) \geq d^{-1}$, and then we can compute N' by the first part of Lemma 1.

Similar to Figure 2 of the portfolio optimization problem, we compare the efficiency between Eff-Sc and CC-Sc for different d and δ in Figure 4, in terms of (a) the required number of samples and (b) the CPU time for solving the sampled approximation problem. We observe that the Eff-Sc has uniformly smaller sample complexity and computational complexity than CC-Sc, where the superiority becomes significant for small δ. In particular, the required number of samples and the used CPU time are bounded for Eff-Sc, whereas they quickly deteriorate for CC-Sc when δ becomes smaller. It is also worth noting that Eff-Sc is consistently more efficient than CC-Sc for all the tested dimensions.

Finally, we compare optimal values of the sampled problems and violation probabilities of the optimal solutions in Figure 5. We present in Figure 5(a) the optimal values and in Figure 5(b) the violation probabilities, with fixed dimension $d = 15$ (we provide additional results for $d = 5$ and $d = 10$ in Appendix C.3). One can quickly

Figure 5. Comparison of the Quality of Optimal Solutions for the Minimal Salvage Fund Problem



Notes. (a) Optimal value. (b) Solutions' violation probabilities. Here $d = 15$, $\delta \in \{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1\}$, and the box plots are generated using 1,000 experiments.

remark from Figure 5(a) that the optimal value of Eff-Sc is stochastically smaller than the optimal value of CC-Sc, whereas Figure 5(b) indicates that the optimal solutions produced by both methods are feasible for all the 10^3 experiments. Therefore, we are able to draw the same conclusion as we have from the portfolio optimization experiment: Eff-Sc efficiently produces less conservative solutions.

Acknowledgments

The authors thank Alexander Shapiro for helpful comments.

Appendix A. Proofs of Technical Results

A.1. Proofs for Section 5

Proof of Lemma 2. We will derive an expression of α_δ to ensure that $F_\delta \subseteq \cup_{\alpha\ge\alpha_\delta}\alpha\cdot\Pi$ for δ small enough. Because of Assumption 2, for any $\alpha_0 > 0$, there exist some δ small enough such that $F_\delta \subseteq \cup_{\alpha\ge\alpha_0}\alpha\cdot\Pi$. Therefore, it suffices to prove that F_δ and $\cup_{\alpha<\alpha_\delta}\alpha\cdot\Pi$ are disjoint. In other words,

$$P(\phi(\alpha\cdot x, L) > 0) > \delta, \quad \forall \alpha < \alpha_\delta,\ x\in\Pi,\ \delta<\delta_0. \tag{A.1}$$

Let ε be a positive number such that $\mu(C_{\varepsilon,-}) > 0$. Pick the set K in (9b) as a compact set such that $0 < \mu(K\cap C_{\varepsilon,-}) < \infty$. It follows from Inequality (9b) that there exist a constant α_1 such that

$$\Psi_-(l) - \varepsilon \le \inf_{x\in\Pi}\left[\frac{1}{r(\alpha)}\phi(\alpha\cdot x, h(\alpha)\cdot l)\right] \qquad \forall l\in K,\ \alpha>\alpha_1. \tag{A.2}$$

Therefore, for any $\alpha \ge \alpha_1$, we have

$$\begin{aligned} P\left(\min_{x\in\Pi}\phi(\alpha\cdot x, L) > 0\right) &= P\left(\min_{x\in\Pi}\frac{1}{r(\alpha)}\phi(\alpha\cdot x, L) > 0\right)\\ \text{(Due to (A.2))} &\ge P(\Psi_-(L/h(\alpha)) \ge \varepsilon; L/h(\alpha)\in K)\\ &= P(L\in h(\alpha)\cdot(K\cap C_{\varepsilon,-})). \end{aligned} \tag{A.3}$$

Recall that L is regularly varying from Assumption 1,

$$\lim_{\alpha\to\infty}\frac{P(L\in h(\alpha)\cdot(K\cap C_{\varepsilon,-}))}{P(\|L\|_2 > h(\alpha))} = \mu(K\cap C_{\varepsilon,-}).$$

Therefore, there exist a number α_2 such that

$$P(L\in h(\alpha)\cdot(K\cap C_{\varepsilon,-})) \ge \frac{1}{2}P(\|L\|_2 > h(\alpha))\mu(K\cap C_{\varepsilon,-}), \qquad \forall\alpha\ge\alpha_2. \tag{A.4}$$

The right-hand side of (A.4) is nondecreasing in α. Thus, if $\delta_1 := \frac{1}{2}P(\|L\|_2 > h(\alpha_2))\mu(K\cap C_{\varepsilon,-})$, for any $\delta\le\delta_1$, there exists α_δ satisfying

$$\frac{1}{2}P(\|L\|_2 > h(\alpha_\delta))\mu(K\cap C_{\varepsilon,-}) = \delta. \qquad \forall\alpha,\delta \quad s.t. \quad \alpha_2\le\alpha<\alpha_\delta,\ 0<\delta\le\delta_1. \tag{A.5}$$

Substituting (A.5) into (A.3), we have

$$P(\phi(x,L) > 0) \ge P\left(\min_{x\in\Pi}\phi(\alpha\cdot x, L) > 0\right) > \delta.$$
$$\forall\alpha, x, \delta \quad s.t. \quad \max(\alpha_1,\alpha_2)\le\alpha<\alpha_\delta, x\in\Pi, 0<\delta\le\delta_1.$$

Moreover, Assumption 2 guarantees the existence of δ_2 such that

$$P(\phi(\alpha\cdot x, L) > 0) > \delta, \qquad \forall\alpha<\max(\alpha_1,\alpha_2), x\in\Pi, \delta<\delta_2.$$

Consequently (A.1) is proved with $\delta_0 = \min(\delta_1,\delta_2)$. □

Proof of Theorem 1. We construct the uniform conditional event C_δ that contains all the V_x for $x\in O_\delta$. Because of Definition (9) and $\lim_{\delta\to 0}\alpha_\delta = \infty$, there exists δ_0 such that for all $\delta<\delta_0$,

$$\Psi_+(l) + \varepsilon \ge \sup_{x\in\Pi}\left[\frac{1}{r(\alpha)}\phi(\alpha\cdot x, h(\alpha)\cdot l)\right] \qquad \forall l\in K,\ \alpha>\alpha_\delta. \tag{A.6}$$

For any $x\in O_\delta$, there exists an $\alpha_x\ge\alpha_\delta$ such that $x\in\alpha_x\cdot\Pi$. Consequently, it follows from (A.6) that

$$\phi(x,l) > 0 \Rightarrow \Psi_+\left(\frac{l}{h(\alpha_x)}\right) \ge -\varepsilon, \qquad \forall x\in O_\delta,\ l\in h(\alpha_x)\cdot K.$$

Applying Assumption 3 yields that

$$\Psi_+\left(\frac{l}{h(\alpha_\delta)}\right) \geq \Psi_+\left(\frac{l}{h(\alpha_x)}\right) \geq -\varepsilon, \qquad \forall x \in O_\delta,\ l \in h(\alpha_x)\cdot(K\bigcap S).$$

Recall that K is a ball in $\mathbb{R}^{d_l}$ (thus, $K \subseteq (h(\alpha_x)/h(\alpha_\alpha))\cdot K$) and that $S \subseteq (h(\alpha_x)/h(\alpha_\alpha))\cdot S$ from Assumption 3, it follows that $h(\alpha_\delta)\cdot(K\cap S) \subseteq h(\alpha_x)\cdot(K\cap S)$. Consequently, whenever $l \in V_x$ for some $x \in O_\delta$, we either have $l \in h(\alpha_x)\cdot(K\cap S)$ implying $\Psi_+\left(\frac{l}{h(\alpha_\delta)}\right) \geq -\varepsilon$, or we have $l \in (h(\alpha_x)\cdot(K\cap S))^c \subseteq (h(\alpha_\delta)\cdot(K\cap S))^c$. Summarizing these two scenarios,

$$\begin{aligned}\bigcup_{x\in O_\delta} V_x &\subseteq \left\{ l \in \mathbb{R}^{d_l} \,|\, \Psi_+\left(\frac{l}{h(\alpha_\delta)}\right) \geq -\varepsilon \right\} \bigcup (h(\alpha_\delta)\cdot(K\bigcap S))^c \\ &= h(\alpha_\delta)\cdot(C_{\varepsilon,+}\bigcup K^c \bigcup S^c).\end{aligned}$$

Thus, we define the conditional set C_δ as

$$C_\delta := h(\alpha_\delta)\cdot(C_{\varepsilon,+}\bigcup K^c \bigcup S^c).$$

It remains to analyze the probability of the uniform conditional event C_δ. As L is multivariate regularly varying,

$$\lim_{\delta\to 0}\frac{\mathrm{P}(L\in C_\delta)}{\mathrm{P}(\|L\|_2 > h(\alpha_\delta))} = \mu(C_{\varepsilon,+}\bigcup K^c \bigcup S^c).$$

Recalling, $\mathrm{P}(\|L\|_2 > h(\alpha_\delta)) = O(\delta)$ and invoking Property 2, we get

$$\limsup_{\delta\to 0} \delta^{-1}\mathrm{P}(L\in C_\delta) < \infty.$$

Hence, the proof is complete. □

Proof of Theorem 2. Using Lemma 1, we immediately have $\mathrm{P}(\mathrm{Val}(\mathrm{CCP}_\delta) \leq \mathrm{Val}(\mathrm{CSP}_{\delta,N'})) \geq 1-\beta$, it remains to show that there exist $\Lambda > 0$ such that $\mathrm{P}(\mathrm{Val}(\mathrm{CSP}_{\delta,N'}) \leq \Lambda \times \mathrm{Val}(\mathrm{CCP}_\delta)) \geq 1-\beta$.

For simplicity, in the proof, we will use L_δ as a shorthand for $(L|L\in C_\delta)$, the random variable with conditional distribution of L given $L \in C_\delta$. By a scaling of x by a factor α_δ in $(\mathrm{CSP}_{\delta,N'})$, we have an equivalent optimization problem:

$$\begin{aligned} &\text{minimize} && c^\top x \\ &\text{subject to} && \frac{1}{r(\alpha_\delta)}\phi(\alpha_\delta\cdot x, L_\delta^{(i)}) \leq 0, \quad i=1,\ldots,N', \\ &&& x \in \bigcup_{\alpha\geq 1}\alpha\cdot\Pi. \end{aligned} \tag{A.7}$$

where $L_\delta^{(i)}$ are i.i.d. samples from L_δ. Notice that $\mathrm{Val}(\mathrm{CSP}_{\delta,N'}) = \alpha_\delta \times \mathrm{Val}(\mathrm{A.7})$.

For any compact set $E \subseteq C_\delta$, because L is multivariate regularly varying,

$$\lim_{\delta\to 0}\mathrm{P}((h(\alpha_\delta))^{-1}L_\delta \in E) = \lim_{\delta\to 0}\frac{\mathrm{P}(L\in(h(\alpha_\delta)\cdot E))}{\mathrm{P}(L\in C_\delta)} = \frac{\lim_{\delta\to 0}\frac{\mathrm{P}(L\in(h(\alpha_\delta)\cdot E))}{\mathrm{P}(\|L\|_2>h(\alpha_\delta))}}{\lim_{\delta\to 0}\frac{\mathrm{P}(L\in C_\delta)}{\mathrm{P}(\|L\|_2>h(\alpha_\delta))}} = \frac{\mu(E)}{\mu(C_{\varepsilon,+}\bigcup K^c\bigcup S^c)}.$$

Thus, $(h(\alpha_\delta))^{-1}L_\delta \xrightarrow{v} L_{\lim}$. As the limiting measure is a probability measure, the family $\{h(\alpha_\delta))^{-1}L_\delta | \delta > 0\}$ is tight and consequently $(h(\alpha_\delta))^{-1}L_\delta \xrightarrow{d} L_{\lim}$ follows directly from the vague convergence (Resnick 2013). Consequently, because all the samples are i.i.d, we also have

$$(h(\alpha_\delta))^{-1}\cdot(L_\delta^{(1)},\ldots,L_\delta^{(N')}) \xrightarrow{d} (L_{\lim^{(1)}},\ldots,L_{\lim^{(N')}}).$$

Now we define a family of deterministic optimization problem, denoted by $(DP(l_1,\ldots,l_{N'}))$, which is parameterized by $(l_1,\ldots,l_{N'})$ as follows:

$$\begin{aligned} &\text{minimize} && c^\top x \\ &\text{subject to} && \phi_{\lim}(x,l_i) \leq 0, \quad i=1,\ldots,N', \\ &&& x \in \bigcup_{\alpha\geq 1}\alpha\cdot\Pi. \end{aligned} \tag{$DP(l_1,\ldots,l_{N'})$}$$

Then, there exist a compact set $E_1 \subseteq \mathbb{R}^{d_l\times N'}$ such that

1. Problem $(DP(l_1,\ldots,l_{N'}))$ satisfies Slater's condition if $(l_1,\ldots,l_{N'}) \in E_1$;
2. The probability $\mathrm{P}((h(\alpha_\delta))^{-1}\cdot(L_\delta^{(1)},\ldots,L_\delta^{(N')}) \in E_1) \geq 1-\beta$ for all $\delta > 0$;

For every $(l_1,\ldots,l_{N'}) \in E_1$ and $\epsilon > 0$, due to the Slater's condition, there exists a feasible solution $x \in \cup_{\alpha\geq 1}\alpha$ such that $\sup_{j=1,\ldots,N'}\phi_{\lim}(x,l_j) < -\epsilon$. Because $\phi_{\lim}(x,l)$ is continuous in l, there exists an open neighborhood U around $(l_1,\ldots,l_{N'})$ such that $\sup_{(l_1,\ldots,l_{N'})\in U}\sup_{j=1,\ldots,N'}\phi_{\lim}(x,l_j) < -\epsilon/2$. Such a feasible solution x and neighborhood U exist for every $(l_1,\ldots,l_{N'}) \in E_1$.

There exists a finite open cover $\{U_i\}_{i=1}^m$ of E_1 due to its compactness. Let $\{x_i\}_{i=1}^m$ be the corresponding feasible solutions to the open cover $\{U_i\}_{i=1}^m$.

From Assumption 5, there exists $\delta_1 > 0$ such that for all $\delta < \delta_1$, we have

$$\sup_{(l_1,\ldots,l_{N'})\in E_1} \sup_{i=1,\ldots,m} \sup_{j=1,\ldots,N'} \left| \frac{1}{r(\alpha_\delta)} \phi(\alpha_\delta \cdot x_j, h(\alpha_\delta)\cdot l_j) - \phi_{\lim}(x_j, l_j) \right| < \epsilon/2. \tag{A.8}$$

Therefore, by the triangle inequality, it follows that if $\delta < \delta_1$,

$$\sup_{(l_1,\ldots,l_{N'})\in U_i} \sup_{j=1,\ldots,N'} \frac{1}{r(\alpha_\delta)} \phi(\alpha_\delta \cdot x_j, h(\alpha_\delta)\cdot l_j) < 0.$$

Consequently, x_j is a feasible solution for Optimization Problem (A.7) if $(h(\alpha_\delta))^{-1}\cdot(L_\delta^{(1)},\ldots,L_\delta^{(N')}) \in U_i$, which further implies that $\alpha_\delta^{-1} \times \mathrm{Val}(\mathrm{CSP}_{\delta,N'}) \le c^\top x_j$. As a result, we have

$$\mathrm{Val}(\mathrm{CSP}_{\delta,N'}) \le \alpha_\delta \times \max_{j=1,\ldots,m} c^\top x_j, \quad \text{if} \quad (h(\alpha_\delta))^{-1}\cdot(L_\delta^{(i)},\ldots,L_\delta^{(N')}) \in E_1.$$

Note that $\mathrm{Val}(\mathrm{CCP}_\delta) \ge \inf_{x\in O_\delta} c^\top x = \alpha_\delta \times \inf\{c^\top x | x \in \cup_{\alpha\ge 1}\alpha\cdot\Pi\}$. Therefore, let

$$\Lambda = \left(\inf\{c^\top x | x \in \bigcup_{\alpha\ge1} \alpha\cdot\Pi\}\right)^{-1} \times \left(\max_{j=1,\ldots,m} c^\top x_j\right) > 0.$$

It follows that

$$\mathrm{P}(\mathrm{Val}(\mathrm{CSP}_{\delta,N'}) \le \Lambda \times \mathrm{Val}(\mathrm{CCP}_\delta)) \ge \mathrm{P}((h(\alpha_\delta))^{-1}\cdot(L_\delta^{(1)},\ldots,L_\delta^{(N')}) \in E_1) \ge 1-\beta.$$

The statement is concluded by using the union bound, combining the lower bound together with the upper bound implied by Lemma 1 and Theorem 1, hence obtaining factor 2β. □

Proof of Lemma 3. Without loss of generality, assume that R is an integer such that

$$Z_\phi = \{(x,l) \in \mathbb{R}^{d_x}\times\mathbb{R}^{d_l} | \phi(x,l) \le 0\} \subseteq [-R,R]^{(d_x+d_l)}.$$

Let $N_1 = (2R+1)^{(d_x+d_l)}$, and let $(x^{(i)}, l^{(i)}), i = 1,\ldots,N_1$ be the integer lattice points in $[-R,R]^{(d_x+d_l)}$. In addition, let $a_j = \frac{\partial\phi}{\partial L}(x^{(i)}, l^{(i)})$, $b_j = \frac{\partial\phi}{\partial x}(x^{(i)}, l^{(i)})$ and $c_j = \phi(x^{(i)}, l^{(i)}) - \frac{\partial\phi}{\partial L}(x^{(i)}, l^{(i)})^\top l^{(i)} - \frac{\partial\phi}{\partial x}(x^{(i)}, l^{(i)})^\top x^{(i)}$ for $i = 1,\ldots,N_1$, then define $\phi_{1,-}(x,l) = \max_{j=1,\ldots,N_1} a_j^\top l + b_j^\top x + c_j$. Because the function $\phi(x,l)$ is convex, we can invoke the supporting hyperplane theorem to deduce that $a_j^\top l + b_j^\top x + c_j \le \phi(x,l)$ for $i = 1,\ldots,N_1$, and consequently $\phi_{1,-}(x,l) \le \phi(x,l)$. In addition, because $\phi(x,l) \ge 0$ at the boundary of the cube $[-R,R]^{(d_x+d_l)}$, there exist a constant C_1 such that $-C_1\cdot R \pm C_1\cdot x_i \le \phi(x,l)$ for $i = 1,\ldots,d_x$ and $-C_1\cdot R \pm C_1\cdot l_i \le \phi(x,l)$ for $i = 1,\ldots,d_l$, for all $(x,l) \in \mathbb{R}^{d_x}\times\mathbb{R}^{d_l}$. Therefore, with $\phi_{2,-}(x,l)$ being the maximum of the aforementioned $N_2 = 2(d_x+d_l)$ linear functions, we have $\phi_{2,-}(x,l) \le \phi(x,l)$, and we also have that $\phi_{2,-}(x,l) \le 0$ implies $(x,l) \in [-R,R]^{(d_x+d_l)}$.

Define $\phi_-(x,l) = \max\{\phi_{1,-}(x,l), \phi_{2,-}(x,l)\}$. We can conclude the property of $\phi_-(x,l)$ as follows: (1) $\phi_-(x,l)$ is a piecewise linear function of form $\max_{j=1,\ldots,N} a_j^\top l + b_j^\top x + c_j$, where $N = N_1 + N_2$; (2) $\phi_-(x,l) \le \phi(x,l)$; and (3) $\phi_-(x,l) \le 0$ implies $(x,l) \in [-R,R]^{(d_x+d_l)}$. To complete the proof, it remains to verify for $\phi_-(x,l)$ the second statement of Assumption 6.

As $\phi_-(x,l) \le 0$ implies $(x,l) \in [-R,R]^{(d_x+d_l)}$, it suffices to prove that there exist some universal constant $C \in \mathbb{R}_+$ such that $\phi(x,l) - \phi_-(x,l) \le C$ for all $(x,l) \in [-R,R]^{(d_x+d_l)}$. For an arbitrary point $(x,l) \in [-R,R]^{(d_x+d_l)}$, there exist a lattice point $(x^{(i)}, l^{(i)})$ such that $\|(x,l) - (x^{(i)}, l^{(i)})\|_2 \le \sqrt{d_x+d_l}/2$. Next, because $\phi(x,l)$ is twice continuously differentiable, the gradient $\nabla\phi(x,l)$ is Lipschitz over $[-R,R]^{(d_x+d_l)}$ with Lipschitz constant denoted by M_ϕ. Therefore, for any $(x,l) \in [-R,R]^{(d_x+d_l)}$,

$$\phi(x,l) - \phi_-(x,l) \le \phi(x,l) - \phi_{1,-}(x,l) \le \min_{j=1,\ldots,N_1} (\phi(x,L) - (a_j^\top L + b_j^\top x + c_j)) \le \frac{1}{4} M_\phi^2 \sqrt{d_x+d_l}.$$

The proof is now complete. □

Proof of Theorem 3. Because $\phi_-(x,L) \le \phi(x,L)$, the probability constraint $\mathrm{P}(\phi(x,L) > 0) \le \delta$ implies that $\mathrm{P}(\phi_-(x,L) > 0) \le \delta$, which further implies $\mathrm{P}(a_j^\top L + b_j^\top x + c_j > 0) \le \delta$ for $i = 1,\ldots,N$. Therefore, we have $-b_j^\top x - c_j \ge \overline{F}_{a_j^\top L}^{-1}(\delta)$ for $i = 1,\ldots,N$, which implies $F_\delta \subseteq O_\delta$.

Then, consider $x \in O_\delta$ and $L \in V_x = \{L \in \mathbb{R}^{d_l} | \phi(x,L) > 0\}$. It follows from the second statement of Assumption 6 that $\phi(x,L) > 0$ implies that $\phi_-(x,L) + C > 0$. Thus, there exist an index i such that $a_j^\top L + b_j^\top x + c_j + C > 0$. As $x \in O_\delta$ implies that $b_j^\top x + c_j + \overline{F}_{a_j^\top L}^{-1}(\delta) \le 0$, so

$$a_j^\top L - \overline{F}_{a_j^\top L}^{-1}(\delta) + C \ge a_j^\top L + b_j^\top x + c_j + C > 0.$$

Therefore, the condition set C_δ can be constructed as

$$C_\delta := \bigcup_{j=1}^N \{L \in \mathbb{R}^{d_l} | a_j^\top L + C > \overline{F}_{a_j^\top L}^{-1}(\delta)\}.$$

Thus, as the distribution $a_j^\top L$ is regularly varying in dimension one for each j, we have $\limsup_{\delta\to 0}\delta^{-1}\mathrm{P}(L\in C_\delta)\le N$, completing the proof. □

A.2. Proofs for Section 6

Proof of Proposition 1. Let $\phi(x,l)=\sum_{j=1}^d(l_j/x_j)-\eta$ and $\pi(x)=\min_{j=1}^d x_j$. The unit level set is $\Pi=\{x\in\mathbb{R}_{++}^d\,|\min_{j=1,\dots,d}x_j=1\}$. Let $h(\alpha)=\alpha$ and $r(\alpha)=1$, it follows that $\frac{1}{r(\alpha)}\phi(\alpha\cdot x,h(\alpha)\cdot l)=\phi(x,l)$. In view of the inequalities $\phi(x,l)\le\mathbf{1}^\top l-\eta$ and $\phi(x,l)\ge\min_{j=1,\dots,d}l_j-\eta$ when $x\in\Pi$, we choose the asymptotic uniform bounds as

$$\Phi_+(l)=\mathbf{1}^\top l-\eta,\qquad \Phi_-(l)=\min_{j=1,\dots,d}L_j-\eta.$$

Furthermore, by definition, we construct two approximation sets as

$$C_{\varepsilon,+}=\{l\in\mathbb{R}_{++}^d\,|\mathbf{1}^\top l\ge\eta-\varepsilon\},\qquad C_{\varepsilon,-}=\{l\in\mathbb{R}_{++}^d\,|\min_{j=1,\dots,d}L_j\ge\eta+\varepsilon\}.$$

With all the elements that we have already defined, Assumption 1 follows directly from the assumption on distribution of L. Now we turn to verify Assumption 2. As $\pi(\alpha\cdot x)=\alpha\cdot\pi(x)$ due to the definition of $\pi(x)$, it suffices to prove that $\lim_{\delta\to 0}\inf_{x\in F_\delta}\pi(x)=+\infty$. In view of $\phi(x,L)\le\mathbf{1}^\top L/\pi(x)-\eta$, we have

$$\begin{aligned}F_\delta=\{x\in\mathbb{R}_{++}^d\mid \mathrm{P}(\phi(x,L)>0)\le\delta\}&\subseteq\{x\in\mathbb{R}_{++}^d\mid \mathrm{P}\left(\mathbf{1}^\top L>\eta\cdot\pi(x)\right)\le\delta\}\\&=\{x\in\mathbb{R}_{++}^d\mid\eta\cdot\pi(x)\ge\overline{F}_{\mathbf{1}^\top L}^{-1}(\delta)\}.\end{aligned}$$

Consequently, we have $\inf_{x\in F_\delta}\pi(x)\ge\eta^{-1}\overline{F}_{\mathbf{1}^\top L}^{-1}(\delta)$. Taking limit for $\delta\to 0$, we conclude that $\lim_{\delta\to 0}\inf_{x\in F_\delta}\pi(x)=+\infty$.

As Assumptions 1 and 2 are both satisfied, and we also have $\mu(C_{\varepsilon,-})>0$, Property 2 is verified due to Lemma 2. In addition, if $\varepsilon\in(0,\eta)$, we have $C_{\varepsilon,+}$ is bounded away from the origin. Thus, Assumption 3 is verified with $S=\mathbb{R}^d$.

Finally, we provide closed form expressions for O_δ and C_δ. Define $\alpha_\delta=\eta^{-1}\cdot\overline{F}_{\mathbf{1}^\top L}^{-1}(\delta)$; then it follows that $O_\delta=\cup_{\alpha\ge\alpha_\delta}\alpha\cdot\Pi=\{x\in\mathbb{R}_{++}^d\,|\eta\cdot\pi(x)\ge\overline{F}_{\mathbf{1}^\top L}^{-1}(\delta)\}$, and $C_\delta=h(\alpha_\delta)\cdot(C_{\varepsilon,+}\cup K^c\cup S^c)=\alpha_\delta\cdot C_{\varepsilon,+}=\{l\in\mathbb{R}_{++}^d\,|\mathbf{1}^\top l\ge(1-\varepsilon/\eta)\cdot\overline{F}_{\mathbf{1}^\top L}^{-1}(\delta)\}$. By setting $\varepsilon=\eta/2$, we get the expression in the statement of the theorem. □

Proof of Lemma 4. We start by showing some properties of $I-Q^\top$. Because Q is a nonnegative matrix and the row sum is less than one, it is a substochastic matrix, and all of its eigenvalues must be less than one in magnitude. This further implies (1) $I-Q^\top$ is invertible, and (2) $(I-Q^\top)^{-1}=I+\sum_{n=1}^\infty(Q^\top)^n$ is a nonnegative matrix with strictly positive diagonal terms.

Notice that $y=(I-Q^\top)^{-1}x$ is the unique vector such that $(I-Q^\top)y=x$. Let (y',b') be the optimal solution of

$$\phi(x,L)=\min_{y,b}\{b\,|(L-y-m)\preceq b\cdot\mathbf{1},\,(I-Q^\top)y\preceq x,\,y\in\mathbb{R}_+^d,b\in\mathbb{R}\}.$$

We have $(I-Q^\top)y'\preceq(I-Q^\top)y=x$, and we multiply the nonnegative matrix $(I-Q^\top)^{-1}$ on both sides, yielding $y'\preceq y$. Obviously, let $b=\max_{j=1,\dots,d}(L_j-y_j)$ such that $(y,\,b)$ is a feasible solution to above problem. Obviously, it follows from $y'\preceq y$ that $b'=\max_{j=1,\dots,d}(L_j-y_j'-m_j)\ge\max_{j=1,\dots,d}(L_j-y_j-m_j)=b$; thus, $(y,\,b)$ is also optimal, which completes the proof. □

Proof of Proposition 2. Assumption 1 follows directly from the assumptions of the example. Now we turn to verify Assumption 6. Using Lemma 4, we define $\phi\_(x,l)=\phi(x,l)=\max_{j=1,\dots,d}L_j-\mathbf{e}_j^\top(I-Q^\top)^{-1}x-m_j$. Therefore, Assumption 6 is satisfied with $n=d$, $a_j=\mathbf{e}_j$, $b_j=-(I-Q)^{-1}\mathbf{e}_j$, $c_j=-m_j$ and $C=0$. Plugging the previous values into the expressions of O_δ and C_δ given in Theorem 3, we get the expressions shown in the statement of the proposition. □

The following lemma is used in the proof of Proposition 3.

Lemma A.1. *There exist sets $S_1,\dots,S_{2d_l}\subseteq\mathbb{R}^{d_l}$ with positive Lebesgue measure such that for any $z\in\mathbb{R}^{d_l}$ with $\|z\|_2=1$, there exist some $S_i\subseteq\{l\in\mathbb{R}^{d_l}\,|z^\top l>1\}$.*

Proof of Lemma A.1. Let $\mathbf{e}_j$ denote the unit vector on the jth coordinate in $\mathbb{R}^{d_l}$ for $j=1,\dots,d_l$. Fix $z=(z_1,\dots,z_{d_l})\in\mathbb{R}^{d_l}$ with $\|z\|_2=1$, define θ_j be the angle between z and $\mathbf{e}_j$, which satisfies $\cos(\theta_j)=z^\top e_j$. Because we have $\sum_{j=1}^n\cos(\theta_j)^2=1$, so there exist some i such that $\cos(\theta_j)^2\ge 1/n$; thus, $z_j\in[-1,-1/\sqrt{n}]\cup[1/\sqrt{n},1]$. Then, define

$$S_{2i-1}=\left\{l=(l_1,\dots,l_{d_l})\in\mathbb{R}^{d_l}\,|L_j>0,\,l_i^2\ge(n-1)\sum_{j\neq i}l_j^2\right\},$$

$$S_{2i}=\left\{l=(l_1,\dots,l_{d_l})\in\mathbb{R}^{d_l}\,|L_j<0,\,l_i^2\ge(n-1)\sum_{j\neq i}l_j^2\right\}.$$

We have either $S_{2i-1}\subset\{l\in\mathbb{R}^{d_l}\,|z^\top l>1\}$ or $S_{2i}\subset\{l\in\mathbb{R}^{d_l}\,|z^\top l>1\}$. Thus, the proof is complete. □

Proof of Proposition 3. For the first statement, because $x^\top Qx\ge 0$ and $A^\top x\in\mathbb{R}^{d_l}$, and invoking the assumption that L has a positive density:

$$\min_{y\in\mathbb{R}^{d_l}\setminus\{\mathbf{0}\}}\mathrm{P}(y^\top L>0)\ge\min_{y:\|y\|_2=1}\mathrm{P}(y^\top L>0)>0.$$

For the second statement, Assumption 1 is easy to verify. Notice that $\alpha^{-2}\phi(\alpha\cdot x,\alpha\cdot L)=\phi(x,L)$ for all $\alpha>0$, so we pick the scaling rate function as $h(\alpha)=\alpha$ and $r(\alpha)=\alpha^2$. Let $\lambda_{\max}$ denote the maximal eigenvalue of Q, and $\lambda_{\min}$ denote the minimal eigenvalue of Q. The rest of the proof will be divided into two cases.

Case 1 ($\lambda_{\max}<0$): We pick the unit level set as $\Pi=\{x\in\mathbb{R}^{d_x}\,|\,\|x\|_2=1\}$. Because $\lim_{\delta\to 0}\inf_{x\in F_\delta}\|x\|_2=\infty$, Assumption 2 is verified. Next, we directly show Property 2 instead of using Lemma 2. For any $x\in\alpha\cdot\Pi$, we have

$$\begin{aligned}\min_{x\in\alpha\cdot\Pi}\mathrm{P}(x^\top Qx+x^\top AL>0)&\ge\min_{x\in\Pi}\mathrm{P}(\alpha\lambda_{\min}+x^\top AL>0)\\&=\min_{x\in\Pi}\mathrm{P}\left(\frac{x^\top AL}{\|A^\top x\|_2}>\frac{-\alpha\lambda_{\min}}{\|A^\top x\|_2}\right)\\&\ge\min_{z:\|z\|=1}\mathrm{P}(z^\top L>-\alpha\sigma^{-1}\lambda_{\min})\\(\text{Apply Lemma A.1})&\ge\min_{i=1,\dots,2d_l}\mathrm{P}(L\in-\alpha\sigma^{-1}\lambda_{\min}S_i).\end{aligned}$$

Thus, α_δ can be chosen such that $\alpha_\delta=O(\delta)$, and $\min_{i=1,\dots,2d_l}\mathrm{P}(L\in-\alpha_\delta\sigma^{-1}\lambda_{\min}S_i)>\delta$. As a result, Property 2 is verified. We next turn to derive the asymptotic uniform bound Ψ_+. Observing that

$$\sup_{x\in\Pi}\phi(x,L)\le\lambda_{\max}+\|A\|_F\|L\|_2,$$

we define $\Psi_+(L):=\lambda_{\max}+\|A\|_F\|L\|_2$. Assumption 3 now follows from the definition of Ψ_+.

Case 2 ($\lambda_{\max}\ge 0$): The unit level set Π is chosen as an unbounded set $\Pi=\{x\in\mathbb{R}^{d_x}\,|\,x^\top Qx=-\|x\|_2\}$, and we have $\min_{x\in\Pi}\|x\|_2=1/|\lambda_{\min}|$. For any $x\in\alpha\cdot\Pi$, we have

$$\begin{aligned}\min_{x\in\alpha\cdot\Pi}\mathrm{P}(x^\top Qx+x^\top AL>0)&\ge\min_{x\in\Pi}\mathrm{P}(x^\top AL>\alpha),\\&=\min_{x\in\Pi}\mathrm{P}\left(\frac{x^\top AL}{\|A^\top x\|_2}>\frac{\alpha}{\|A^\top x\|_2}\right)\\&\ge\min_{z:\|z\|=1}\mathrm{P}(z^\top L>-\alpha\sigma^{-1}\lambda_{\min})\\(\text{Apply Lemma A.1})&\ge\min_{i=1,\dots,2d_l}\mathrm{P}(L\in-\alpha\sigma^{-1}\lambda_{\min}S_i).\end{aligned}$$

Thus, we can pick an α_δ that satisfies Property 2. Now, $\sup_{x\in\Pi}\phi(x,L)$ is bounded by

$$\sup_{x\in\Pi}\phi(x,L)\le\sup_{x\in\Pi}\|x\|_2(\|AL\|_2-1)\le-\frac{1}{2}|\lambda_{\min}|^{-1}\cdot I(\|AL\|_2\le 1/2)+\infty\cdot I(\|AL\|_2>1),$$

so we can pick $\Psi_+(L):=-\frac{1}{2}|\lambda_{\min}|^{-1}\cdot I(\|AL\|_2\le 1/2)+\infty\cdot I(\|AL\|_2>1)$. Consequently Assumption 3 follows immediately. □

Appendix B. Importance Sampling for Multivariate Regularly Varying Distribution

B.1. Multivariate Regularly Varying Distribution with Gaussian Copula

In this section, we assume that the correlation structure of the random vector L is characterized by Gaussian Copula.

Let $\Phi:\mathbb{R}\to[0,1]$ be the standard univariate Gaussian cumulative distribution function (CDF), and $\Phi_\Sigma:\mathbb{R}^d\to[0,1]$ be the joint CDF of multivariate Gaussian CDF with mean of zero, variance of one, and covariance matrix of Σ. The Gaussian Copula $C_\Sigma:[0,1]^d\to[0,1]$ is defined as $C_\Sigma(u_1,\dots,u_d)=\Phi_\Sigma(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_d))$. Suppose that the random vector L has marginal CDF $F_{L_i}:\mathbb{R}\to[0,1]$ for $i=1,\dots,d$, we assume that $U:=(U_1,\dots,U_d)=(F_{L_1}(L_1),\dots,F_{L_d}(L_d))$ has joint CDF C_Σ.

Algorithm B.1 (Sampling of Multivariate Regularly Varying Distribution with Gaussian Copula)

Input: The covariance matrix of Gaussian Copula Σ, the marginal CDFs F_{L_i}.

1 Apply Cholesky decomposition or singular value decomposition to compute the matrix A such that $\Sigma=A^\top A$.
2 Sample a d-dimensional multivariate standard normal vector Z, compute the linear transform $X=A^\top Z$.
3 For each $i=1,\dots,d$, compute $L_i=F_{L_i}^{-1}(\Phi(X_i))$.

In the importance sampling algorithm developed later, we need to sample the random vector L conditional on the value of one coordinate, for example, $L_i=l_i$. Without loss of generality, we assume that the value of the last coordinate is given, and the covariance matrix Σ admits the blockwise representation $\Sigma=\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&1\end{pmatrix}$, where $\Sigma_{11}\in\mathbb{R}^{(d-1)\times(d-1)}$, $\Sigma_{12}\in\mathbb{R}^{(d-1)\times 1}$, $\Sigma_{21}\in\mathbb{R}^{1\times(d-1)}$ and $\Sigma_{12}=\Sigma_{21}^\top$. Suppose that X is a random vector with normal distribution $N(\mathbf{0},\Sigma)$, then the conditional distribution of $(X_1,\dots,X_{d-1})$ given $X_d=x_d$ is jointly normal distributed with mean $x_d\cdot\Sigma_{12}$ and covariance $\Sigma_{11}-\Sigma_{12}\Sigma_{21}$. In Algorithm B.2, we describe the conditional sampling method for L.

Algorithm B.2 (Sampling of Multivariate Regularly Varying Distribution with Gaussian Copula Conditional on $L_d = l_d$)

Input: The covariance matrix of Gaussian Copula Σ, the marginal CDFs F_{L_i}, the conditional value of the last coordinate $L_d = l_d$.

1 Map the observation into the Gaussian space: $x_d = \Phi^{-1}(F_{L_d}(l_d))$.
2 Sample a $d-1$-dimensional multivariate normal vector $(X_1, \dots, X_{d-1})$, with mean $x_d \cdot \Sigma_{12}$ and covariance $\Sigma_{11} - \Sigma_{12}\Sigma_{21}$.
3 For each $i = 1, \dots, d-1$, compute $L_i = F_{L_i}^{-1}(\Phi(X_i))$.

B.2. Importance Sampling

In this section, we present an importance sampling method to sample from the conditional distribution $(L|L \in C_\delta)$, where C_δ is the uniform conditional event.

B.2.1. Minimal Salvage Fund. Recall from Proposition 2 that the uniform conditional event for the minimal salvage fund problem is $C_\delta = \cup_{i=1}^d \{l \in \mathbb{R}^d | l_i > \overline{F}_{L_i}^{-1}(\delta)\}$. To simplify the notation, let us define $\omega_i(\delta) = \overline{F}_{L_i}^{-1}(\delta)$ for $i = 1, \dots, d$. It follows that $\mathrm{P}(L_i > \omega_i(\delta)) = \delta$ for $i = 1, \dots, d$.

The algorithm is a combination of acceptance rejection and importance sampling. Suppose that $\mathrm{P}(\mathrm{d}l)$ is the probability measure corresponding to the random vector L, then the target measure $\mathrm{P}_{\text{target}}(\mathrm{d}l)$ corresponding to the conditional distribution $(L|L \in C_\delta)$ can be expressed as

$$\mathrm{P}_{\text{target}}(\mathrm{d}l) = \frac{\mathbb{I}\{l \in C_\delta\}}{\mathrm{P}(L \in C_\delta)} \mathrm{P}(\mathrm{d}l).$$

Now we describe how to sample from the proposal distribution with importance sampling. For each fixed $i \in \{1, \dots, d\}$, the conditional distribution $(L|L_i > \omega_i(\delta))$ can be sampled using the importance sampling: We first sample L_i conditional on $L_i > \omega_i(\delta)$ by the inverse CDF method and then apply Algorithm B.2 to sample L_j for $j \neq i$ conditional on L_i. The resulting random vector L has probability measure $\frac{\mathbb{I}\{l_i > \omega_i(\delta)\}}{\mathrm{P}(L_i > \omega_i(\delta))} \mathrm{P}(\mathrm{d}l)$. Then, if we uniformly sample the random index i from $\{1, \dots, d\}$ instead of using the fixed index, the proposal distribution becomes

$$\mathrm{P}_{\text{proposal}}(\mathrm{d}l) = \frac{1}{d\delta} \sum_{i=1}^d \mathbb{I}\{l_i > \omega_i(\delta)\} \mathrm{P}(\mathrm{d}l).$$

The likelihood ratio is

$$\frac{\mathrm{P}_{\text{target}}(\mathrm{d}l)}{\mathrm{P}_{\text{proposal}}(\mathrm{d}l)} = \frac{d\delta}{\mathrm{P}(L \in C_\delta)} \frac{\mathbb{I}\{l \in C_\delta\}}{\sum_{i=1}^d \mathbb{I}\{l_i > \omega_i(\delta)\}}.$$

The proposal distribution guarantees that there exist at least an index i such that $L_i > \omega_i(\delta)$; thus, we have $\mathbb{I}\{l \in C_\delta\} = 1$ and $\sum_{i=1}^d \mathbb{I}\{l_i > \omega_i(\delta)\} \geq 1$. In addition, the definition of C_δ implies that $\mathrm{P}(L \in C_\delta) \geq \mathrm{P}(L_i > \omega_i(\delta)) = \delta$. Consequently, the likelihood ratio is upper bounded by d.

To conclude this section, we summarize the detail of the importance sampling in Algorithm B.3.

Algorithm B.3 (Importance Sampling Algorithm for Minimal Salvage Fund Problem)

Input: The covariance matrix of Gaussian Copula Σ, the marginal CDFs F_{L_i}, the risk level of tolerance δ.

1 Uniformly sample a random index $i \in \{1, \dots, d\}$.
2 Sample a uniform random variable $U_1 \sim \mathrm{Unif}(0,1)$. Set $L_i = F_{L_i}^{-1}((1-\delta) + \delta \cdot U)$.
3 Apply Algorithm B.2 to sample the rest coordinates L_j for $j \neq i$ conditional on the value of L_i.
4 Sample a uniform random variable $U_2 \sim \mathrm{Unif}(0,1)$.
5 If $U_2 > (\sum_{k=1}^d \mathbb{I}\{L_k > \omega_k(\delta)\})^{-1}$, output $L = (L_1, \dots, L_d)$; otherwise, return to step 1.

B.2.2. Portfolio Optimization with VaR Constraint. Recall from Proposition 1 that the uniform conditional event for the portfolio optimization problem is $C_\delta = \{l \in \mathbb{R}_{++}^d | 2 \cdot \mathbf{1}^\top l \geq \overline{F}_{\mathbf{1}^\top L}^{-1}(\delta)\}$. It is not hard to see that

$$C_\delta \subseteq \bigcup_{i=1}^d \{l \in \mathbb{R}^d | l_i > (2d)^{-1} \cdot \overline{F}_{\mathbf{1}^\top L}^{-1}(\delta)\},$$

where the right-hand side has a similar form to C_δ in the minimal salvage fund problem. Define $\varpi_i(\delta) := (2d)^{-1} \cdot \overline{F}_{\mathbf{1}^\top L}^{-1}(\delta)$, and we construct the proposal distribution as

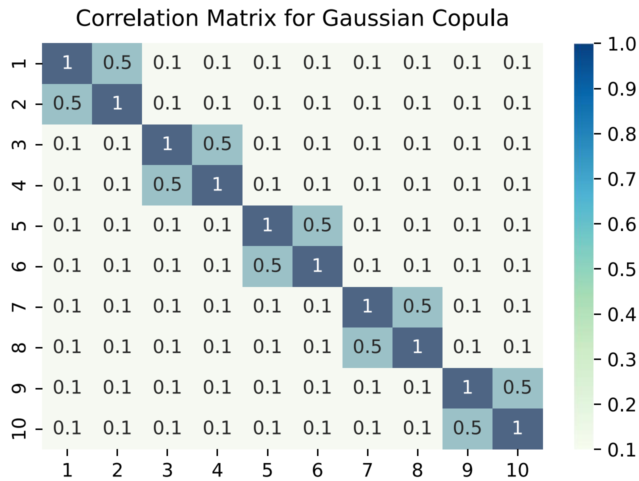
$$\mathrm{P}_{\text{proposal}}(\mathrm{d}l) \propto \sum_{i=1}^d \mathbb{I}\{l_i > \varpi_i(\delta)\} \mathrm{P}(\mathrm{d}l).$$

Because the target distribution is still $\mathrm{P}_{\text{target}}(\mathrm{d}l) = \frac{\mathbb{I}\{l \in C_\delta\}}{\mathrm{P}(L \in C_\delta)} \mathrm{P}(\mathrm{d}l)$, the likelihood ratio is

$$\frac{\mathrm{P}_{\text{target}}(\mathrm{d}l)}{\mathrm{P}_{\text{proposal}}(\mathrm{d}l)} \propto \frac{\mathbb{I}\{l \in C_\delta\}}{\sum_{i=1}^d \mathbb{I}\{l_i > \varpi_i(\delta)\}} \leq 1.$$

To conclude this section, we summarize the detail of the importance sampling in Algorithm B.4.

Figure C.1. Correlation Matrix for Gaussian Copula



Algorithm B.4 (Importance Sampling Algorithm for Portfolio Optimization with VaR Constraint)

Input: The covariance matrix of Gaussian Copula Σ, the marginal CDFs F_{L_i}, the risk level of tolerance δ.

1 Sample a random index $i \in \{1, \ldots, d\}$ with probability propotional to $\mathrm{P}(L_i > \varpi_i(\delta))$, where $\varpi_i(\delta) = (2d)^{-1} \cdot \overline{F}^{-1}_{\mathbf{1}^\top L}(\delta)$.
2 Sample a uniform random variable $U_1 \sim \mathrm{Unif}(0,1)$. Set $L_i = F^{-1}_{L_i}(F_{L_i}(\varpi_i(\delta)) + (1 - F_{L_i}(\varpi_i(\delta))) \cdot U_1)$.
3 Apply Algorithm B.2 to sample the rest coordinates L_j for $j \neq i$ conditional on the value of L_i.
4 Sample a uniform random variable $U_2 \sim \mathrm{Unif}(0,1)$.
5 If $L \in C_\delta$ (i.e., $2 \cdot \mathbf{1}^\top L \geq \overline{F}^{-1}_{\mathbf{1}^\top L}(\delta)$) and $U_2 > (\sum_{k=1}^{d} \mathbb{I}\{L_k > \omega_k(\delta)\})^{-1}$, output $L = (L_1, \ldots, L_d)$; otherwise, return to step 1.

Appendix C. Additional Numerical Results

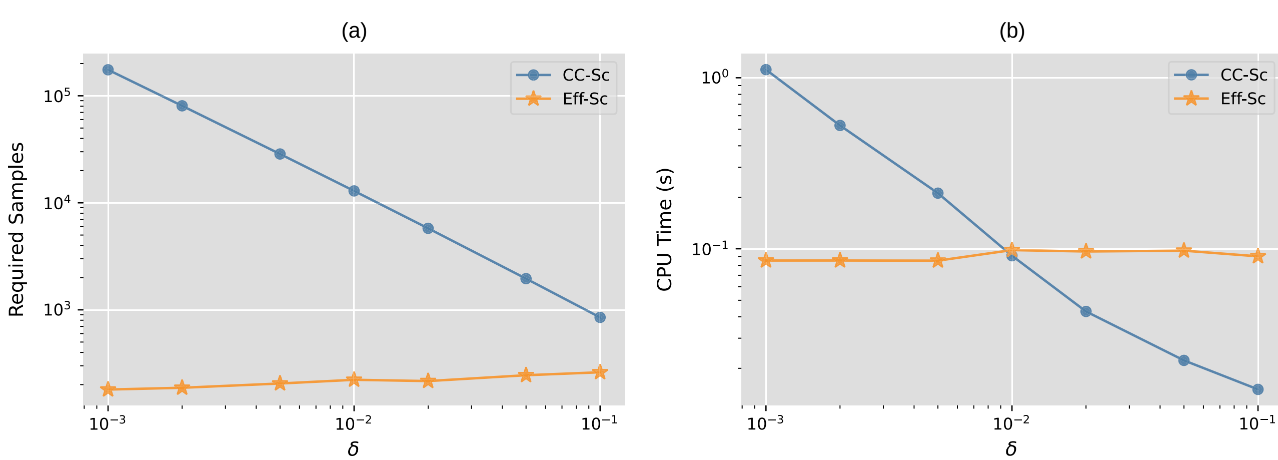
C.1. Portfolio Optimization with Dependent Loss

In this section, we conduct additional numerical experiments for the portfolio optimization problem (2). We still consider the portfolio optimization problem with $d = 10$ assets and use the same mean return vector μ and loss threshold η as Section 7.1. Although we also assume the same marginal distribution for the loss vector L, we apply the Gaussian Copula (see Appendix B.1) to impose the dependence structure between different coordinates of L. In particular, we assume the correlation matrix of Gaussian Copula as in Figure C.1.

To solve the change constraint problem using Eff-Sc. We adopt the same construction of O_δ and C_δ as Section 7.1 and compute the required number of samples N' using Lemma 1. The samples of L_δ is generated via the importance sample method (see Algorithm B.4 for detail).

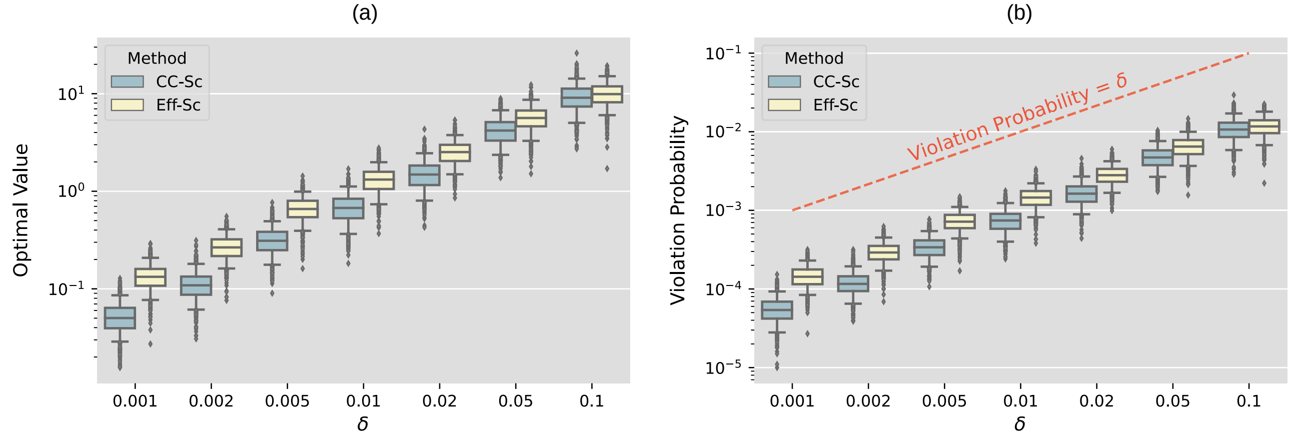
In Figure C.2, we compare the efficiency between Eff-Sc and CC-Sc. As shown in Figure C.2(a), the required number of samples for Eff-Sc is substantially less than CC-Sc, especially when δ is small. In Figure C.2(b), we compare the running

Figure C.2. Comparison of Computational Efficiency for the Portfolio Optimization Problem



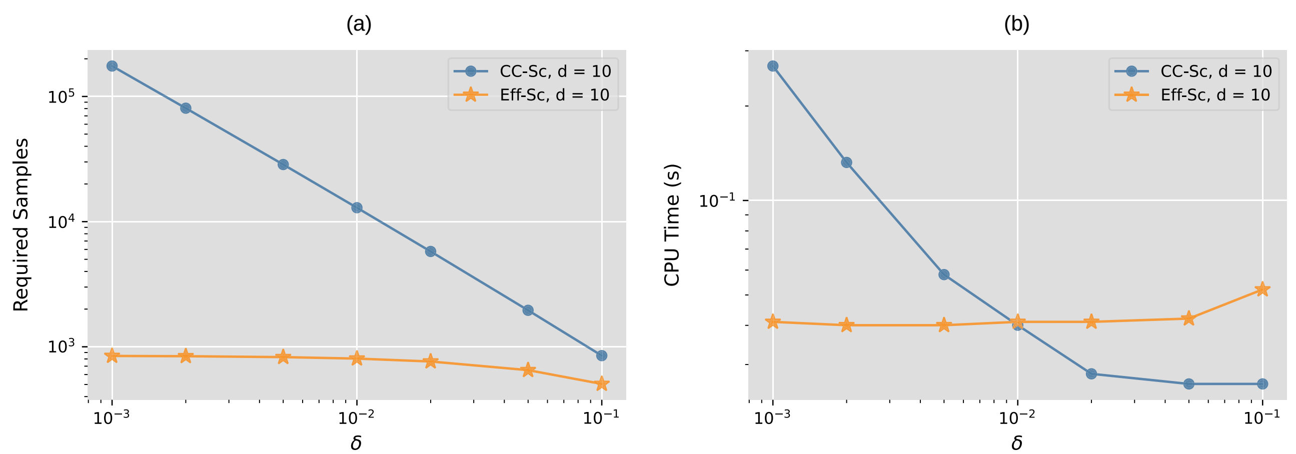
Notes. (a) Required number of samples. (b) Used CPU time. We test $\delta \in \{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1\}$.

Figure C.3. Comparison of the Quality of Optimal Solutions for the Portfolio Optimization Problem with Dependent Loss Generated Using Gaussian Copula



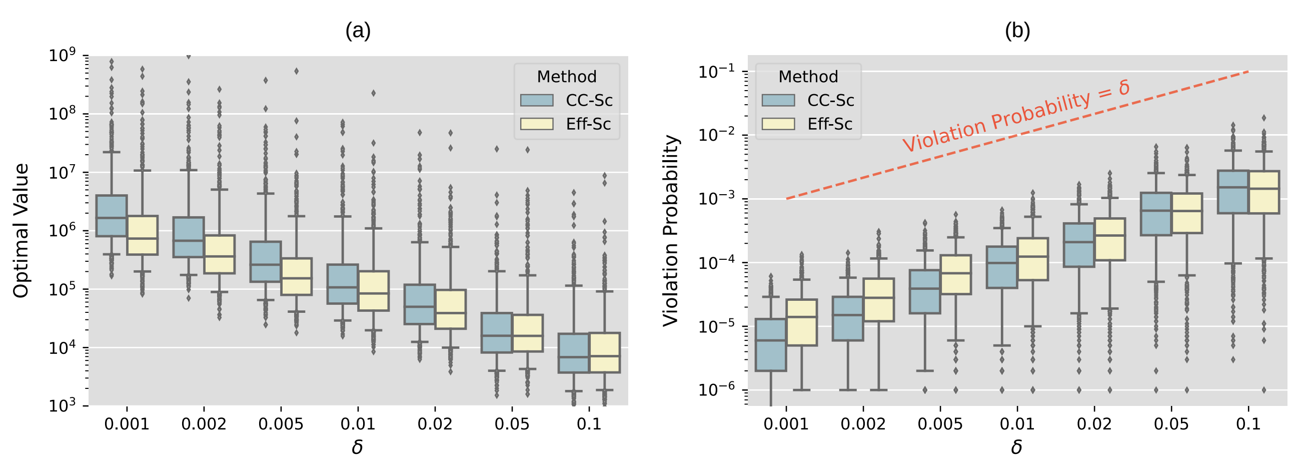
Notes. (a) Optimal value. (b) Solutions' violation probabilities. Here $\delta \in \{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1\}$, and the box plots are generated using 1,000 experiments.

Figure C.4. Comparison of Computational Efficiency for the Minimal Salvage Fund Problem with Dependent Loss



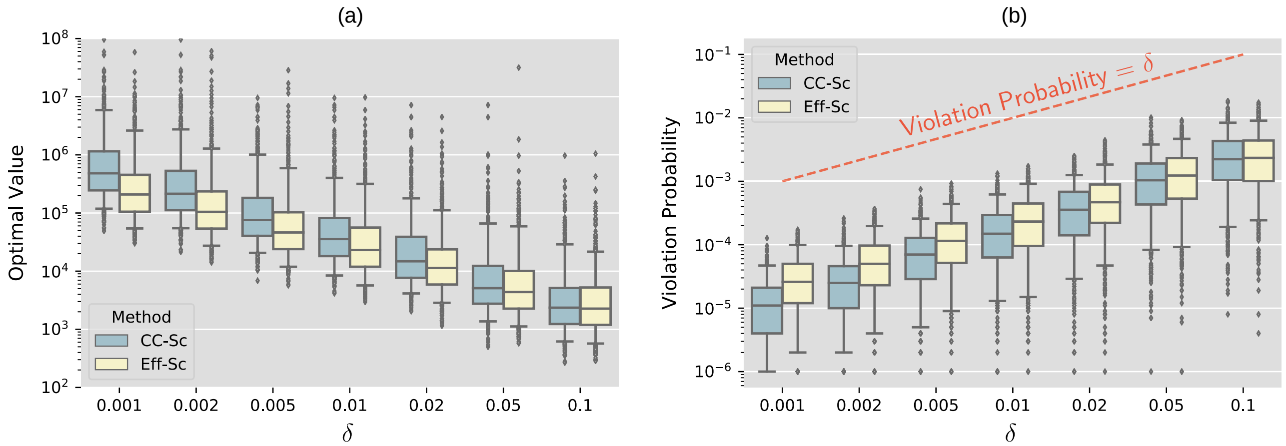
Notes. (a) Required number of samples. (b) Used CPU time. We test $d = 10$ and $\delta \in \{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1\}$.

Figure C.5. Comparison of the Quality of Optimal Solutions for the Minimal Salvage Fund Problem with Dependent Loss



Notes. (a) Optimal value. (b) Solutions' violation probabilities. Here $d = 10$, $\delta \in \{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1\}$, and the box plots are generated using 1,000 experiments.

Figure C.6. Comparison of the Quality of Optimal Solutions Given by Eff-Sc and CC-Sc for $d = 5$



Notes. (a) Optimal value. (b) Solutions' violation probabilities.

time for both models. We remark that the computational time for Eff-Sc stays nearly constant for different δ, and that Eff-Sc needs less time to solve than CC-Sc for small δ.

In Figure C.3, we compare the optimal value and the conservativeness of the solutions generated by Eff-Sc and CC-Sc. From the figure, we can conclude that the solutions for Eff-Sc and CC-Sc are both feasible, and the Eff-Sc solution is less conservative with better optimal value.

C.2. Minimal Salvage Fund with Dependent Loss

In this section, we test the performance of the minimal salvage fund problem (2) in which the loss vector L has dependent structure characterized by the Gaussian copula.

In the experiment, we fixed $d = 10$, and use the same parameters Q and m and the same marginal distribution of L_j as introduced in Section 7.2. We assume that the dependence structure of different coordinates of L is prescribed by the Gaussian Copula with correlation matrix shown in Figure C.1.

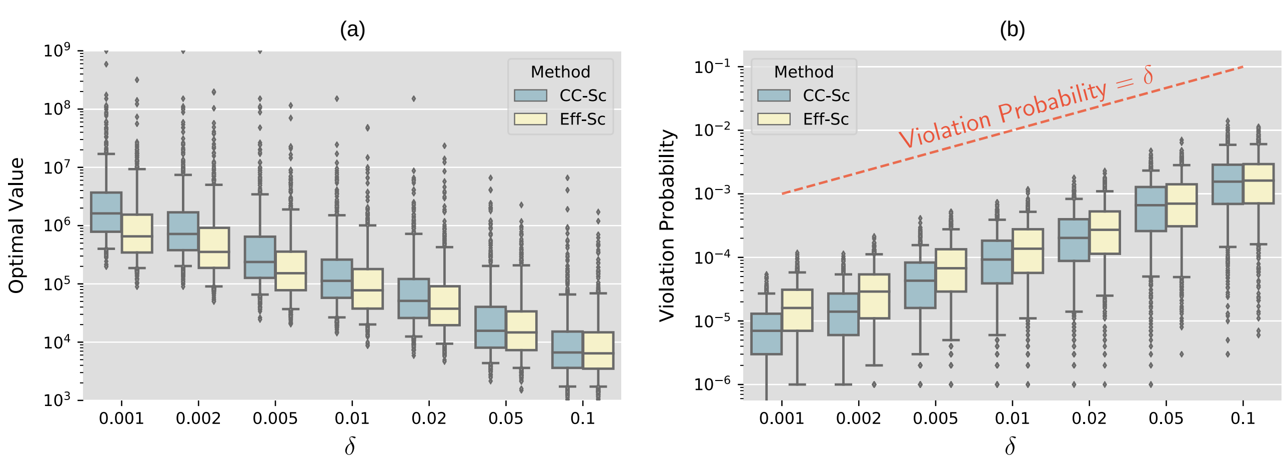
In Figure C.4, we compare the efficiency between Eff-Sc and CC-Sc for solving the minimal salvage fund problem. In particular, we compare the required number of samples in Figure C.4(a) and the total required CPU time in Figure C.4(b). Despite slightly larger CPU time for Eff-Sc for large δ, the CPU time for Eff-Sc becomes significantly smaller than CC-Sc when $\delta < 0.01$, and the required number of samples for Eff-Sc is also universally smaller.

In Figure C.5, we also compare the quality of the solutions generated by Eff-Sc and CC-Sc. Once again, we found that the solutions generated by Eff-Sc are less conservative with better optimal value on average.

C.3. Minimal Salvage Fund for $d = 5$ and $d = 10$

In this section, we demonstrate the quality of the solutions produced by Eff-Sc is better than CC-Sc when the dimension of the problem is $d = 5$ or $d = 10$. See Figure C.6 for dimension $d = 5$ and Figure C.7 for $d = 10$.

Figure C.7. Comparison of the Quality of Optimal Solutions Given by Eff-Sc and CC-Sc for $d = 10$



Notes. (a) Optimal value. (b) Solutions' violation probabilities.

References

Ahmed S, Shapiro A (2008) Solving chance-constrained stochastic programs via sampling and integer programming. Chen Z-L, Raghavan S, eds. *State-of-the-Art Decision-Making Tools in the Information-Intensive Age* (INFORMS), 261–269.

Alsenwi M, Pandey SR, Tun YK, Kim KT, Hong CS (2019) A chance constrained based formulation for dynamic multiplexing of eMBB-URLLC traffics in 5G new radio. *Proc. Internat. Conf. on Information Networking* (IEEE, New York), 108–113.

Andrieu L, Henrion R, Römisch W (2010) A model for dynamic chance constraints in hydro power reservoir management. *Eur. J. Oper. Res.* 207(2):579–589.

Barrera J, Homem-de Mello T, Moreno E, Pagnoncelli BK, Canessa G (2016) Chance-constrained problems and rare events: An importance sampling approach. *Math. Programming* 157(1):153–189.

Ben-Tal A, Nemirovski A (2000) Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming* 88(3):411–424.

Ben-Tal A, Nemirovski A (2002) Robust optimization: Methodology and applications. *Math. Programming* 92(3):453–480.

Bertsimas D, Sim M (2004) The price of robustness. *Oper. Res.* 52(1):35–53.

Blanchet J, Liu J (2010) Efficient importance sampling in ruin problems for multidimensional regularly varying random walks. *J. Appl. Probability* 47(2):301–322.

Bonami P, Lejeune MA (2009) An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Oper. Res.* 57(3):650–670.

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).

Calafiore G, Campi MC (2005) Uncertain convex programs: Randomized solutions and confidence levels. *Math. Programming* 102(1):25–46.

Calafiore G, Campi MC (2006) The scenario approach to robust control design. *IEEE Trans. Automated Control* 51(5):742–753.

Charnes A, Cooper WW, Symonds GH (1958) Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Management Sci.* 4(3):235–263.

Chen B, Blanchet J, Rhee CH, Zwart B (2019) Efficient rare-event simulation for multiple jump events in regularly varying random walks and compound Poisson processes. *Math. Oper. Res.* 44(3):919–942.

Chen W, Sim M, Sun J, Teo CP (2010) From CVaR to uncertainty set: Implications in joint chance-constrained optimization. *Oper. Res.* 58(2):470–485.

Diamond S, Boyd S (2016) CVXPY: A Python-embedded modeling language for convex optimization. *J. Machine Learn. Res.* 17(83):1–5.

Eisenberg L, Noe TH (2001) Systemic risk in financial systems. *Management Sci.* 47(2):236–249.

Embrechts P, Klüppelberg C, Mikosch T (2013) *Modelling Extremal Events: For Insurance and Finance*, vol. 33 (Springer Science & Business Media, Boston).

Frank B (2008) Municipal bond fairness act. *110th Congress, 2d Session, House of Representatives, Report*, 110–835.

Gudmundsson T, Hult H (2014) Markov chain Monte Carlo for computing rare-event probabilities for a heavy-tailed random walk. *J. Appl. Probability* 51(2):359–376.

Hillier FS (1967) Chance-constrained programming with 0-1 or bounded continuous decision variables. *Management Sci.* 14(1):34–57.

Hong LJ, Huang Z, Lam H (2021) Learning-based robust optimization: Procedures and statistical guarantees. *Management Sci.* 67(6):3447–3467.

Hong LJ, Yang Y, Zhang L (2011) Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach. *Oper. Res.* 59(3):617–630.

Kley O, Klüppelberg C, Reinert G (2016) Risk in a large claims insurance market with bipartite graph structure. *Oper. Res.* 64(5):1159–1176.

Küçükyavuz S (2012) On mixing sets arising in chance-constrained programming. *Math. Programming* 132(1–2):31–56.

Lagoa CM, Li X, Sznaier M (2005) Probabilistically constrained linear programs and risk-adjusted controller design. *SIAM J. Optim.* 15(3):938–951.

Lejeune MA, Margot F (2016) Solving chance-constrained optimization problems with stochastic quadratic inequalities. *Oper. Res.* 64(4):939–957.

Luedtke J (2014) A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. *Math. Programming* 146(1–2):219–244.

Luedtke J, Ahmed S (2008) A sample approximation approach for optimization with probabilistic constraints. *SIAM J. Optim.* 19(2):674–699.

Luedtke J, Ahmed S, Nemhauser GL (2010) An integer programming approach for linear programs with probabilistic constraints. *Math. Programming* 122(2):247–272.

McGill R, Tukey JW, Larsen WA (1978) Variations of box plots. *Amer. Statist.* 32(1):12–16.

MOSEK ApS (2020) MOSEK fusion API for Python. Retrieved September 10, 2020, https://docs.mosek.com/9.2/pythonfusion.pdf.

Nemirovski A, Shapiro A (2006a) Convex approximations of chance constrained programs. *SIAM J. Optim.* 17(4):969–996.

Nemirovski A, Shapiro A (2006b) Scenario approximations of chance constraints. *Probabilistic and Randomized Methods for Design Under Uncertainty* (Springer, Berlin), 3–47.

Peña-Ordieres A, Luedtke JR, Wächter A (2020) Solving chance-constrained problems via a smooth sample-based nonlinear approximation. *SIAM J. Optim.* 30(3):2221–2250.

Prekopa A (1970) On probabilistic constrained programming. William Kuhn H, ed. *Proc. Princeton Sympos. on Math. Programming*, vol. 113 (Princeton University Press, Princeton), 138.

Prékopa A (2003) Probabilistic programming. *Handbook Oper. Res. Management Sci.* 10:267–351.

Resnick SI (2013) *Extreme Values, Regular Variation and Point Processes* (Springer, Berlin).

Seppälä Y (1971) Constructing sets of uniformly tighter linear approximations for a chance constraint. *Management Sci.* 17(11):736–749.

Tong S, Subramanyam A, Rao V (2022) Optimization under rare chance constraints. *SIAM J. Optim.* 32(2):930–958.

Wierman A, Zwart B (2012) Is tail-optimal scheduling possible? *Oper. Res.* 60(5):1249–1257.

Zhang M, Küçükyavuz S, Goel S (2014) A branch-and-cut method for dynamic decision making under joint chance constraints. *Management Sci.* 60(5):1317–1333.