Convex Approximations of a Probabilistic Bicriteria Model with Disruptions

Tara Rengarajan*, Nedialko Dimitrov†, David P. Morton*
* Graduate Program in Operations Research and Industrial Engineering, The University of Texas at Austin, Austin, TX 78712, USA,
† Operations Research Department, Naval Postgraduate School, Monterey, CA 93943, USA,
{trengan@yahoo.com, ned@nps.edu, morton@mail.utexas.edu}

We consider a multiperiod system operation problem with two conflicting objectives, minimizing cost and risk. Risk stems from uncertain disruptions to the system during operation. While a general model would hedge against disruptions in each time period, we study special cases in which only a modest number of disruptions occur. To optimize for risk, we employ a convex approximation based on constraint sampling. We develop a stratified sampling scheme based on distributional information on the time of disruption. We establish that our scheme yields significant savings in sampling costs—up to an order of magnitude in the number of time periods—over naive sampling. Moreover, in the absence of distributional information, we exhibit a sampling strategy that has comparable performance to optimal stratification. We numerically demonstrate that stratification improves cost over naive sampling, improving the solution's proximity to the efficient frontier of the bicriteria problem.

Key words: programming, stochastic: probabilistic constraints; simulation; programming: multiple criteria

---

1. Introduction

When optimizing large-scale stochastic systems, performance should be balanced against risk tolerance to random disruptions. Decision makers seek to understand trade-offs between these conflicting goals. From a practical perspective, it is of interest to the decision maker to find solutions wherein an improvement in one objective cannot be achieved without a detriment in the other, i.e., to optimize a bicriteria model. Bicriteria optimization is well-studied and has a rich history beginning with Markowitz (1952) trading off risk, as measured by variance of a financial portfolio's return, with that portfolio's mean return. Recent work in stochastic programming includes Ruszczyński and Vanderbei (2003), who construct the efficient frontier of mean-risk models, and Schultz and Tiedemann (2003), who propose and study stochastic programs with recourse that include a risk term in the objective.

We consider a probabilistic bicriteria program spanning finitely many time periods, involving random disruptions. In a significant departure from standard approaches, we model both the *time* and *magnitude* of the disruption as random. This is typical of events such as fires, storms and market crashes, which can seriously disrupt a system. The enormity of a disruption suggests that it occurs infrequently, possibly at most once to our finite planning horizon. Under such random-time disruptions, we study problems in which decisions are made in each time period subject to system dynamics and a bicriteria objective involving cost and a probabilistic risk measure. The risk is the probability of a bad, or undesirable, event consequent to the disruption.

The following perishable inventory management problem is a motivating example for models incorporating random disruptions. (See the reviews by Nahmias (1982), and Goyal and Giri (2001) for a review of perishable inventory theory.)

> **Example 1.** Consider a perishable inventory management problem over finitely many planning periods. In each time period, a firm manufactures a perishable product. That product may be made available for immediate use, i.e., "placed on the shelf" or may be stored for future use. Unconsumed product on the shelf perishes at the end of each period. A known nominal demand occurs in each time period. The random disruption appears in the form of excess demand, and occurs in at most one time period. Only product placed on the shelf can be used to meet nominal and random demand. Costs incurred by the firm include production costs, storage costs and penalty costs incurred due to unmet nominal demand. The firm hedges against the disruption by ensuring that excess demand is satisfied with high probability. Decisions are made prior to the start of the first period and under the assumption that the probability distribution governing the magnitude and time of the disruption is known. (This assumption is relaxed in Section 4.)

We now turn to a review of literature related to the problem we consider.

1.1. Related Work

There is a large literature on multicriteria optimization that goes much beyond what we have cited above. See, for example, the books of Ehrgott and Gandibleux (2002), Pardalos et al. (1995), Yu (1985) and Zeleny (1982). There is much less work on solving probabilistic bicriteria models, although we can point to Yang and Feng (2007), who study a bicriteria variant of a fixed charge transportation problem with probabilistic constraints, and Pagnoncelli et al. (2008), who computationally study the efficient frontier of a sampling-based approximation to an asset-allocation model.

The efficient frontier of a probabilistic bicriteria model can be determined by means of a parametric probabilistically-constrained program. Much of the computationally-oriented research in probabilistically-constrained programs has focused on tractable special cases and approximate solution methods. We outline below such work most closely related to ours.

Luedtke et al. (2010) formulate a probabilistically-constrained program as a mixed-integer program and strengthen that formulation. Nemirovski and Shapiro (2006b) use constraint sampling from an importance-sampling distribution, to construct a tractable approximation, under the assumption that the random constraints are bilinear in the decision variables and the random parameters. Nemirovski and Shapiro (2006a) develop a large deviation-type approximation that is convex and efficiently solvable. Luedtke and Ahmed (2008) approximate the probability distribution in a probabilistically-constrained program with the empirical distribution obtained from a sample, and obtain lower and upper confidence bounds on the model's optimal value; see Rengarajan and Morton (2009) for related work in the bicriteria setting. Calafiore and Campi (2005, 2006) provide probabilistic feasibility guarantees for probabilistically-constrained programs with convex objective and constraint functions, by the use of constraint sampling. Our work focuses on constraint sampling under a sampling budget, and builds on the results of Calafiore and Campi. For our random-time disruption model, we present sampling strategies that reduce the sample size requirement, as compared to direct application of the work of Calafiore and Campi (2005).

There has been little work in stochastic programming on the notion of the time of disruption being random. Salmerón et al. (2009) study the problem of transporting military cargo between seaports subject to a biological attack by an enemy. In contrast to the probabilistic model we consider, their model is a stochastic mixed-integer program with recourse. A substantial body of work has also been devoted to the area of "disruption management"

Yu and Qi (2004). The modeling approach in our work differs from disruption management in that decisions in our model are made initially to hedge against future disruption, while disruption management focuses on decisions made after the disruption.

1.2. Main Contributions

We approximately solve our probabilistic bicriteria model via constraint sampling subject to a sampling budget. In constraint sampling, we draw observations of the stochastic parameters and force a set of constraints generated by these observations to hold simultaneously. Thus, the original probabilistic program is replaced by a random, finitely constrained, sampled program.

In our bicriteria model, we seek to keep both cost and probabilistic risk low. The constraint sampling approximation minimizes cost subject to system dynamic constraints plus a set of constraints that ensures the bad event does not occur at the sampled observations. The constraint sampling approximation is a random program and hence, when optimized, yields a random decision vector. We call the expected value of the probabilistic risk function with respect to this random solution, the *expected probability of violation*.

Calafiore and Campi (2005) bound the expected probability of violation. Furthermore, they use this bound and Markov's inequality to bound the probability that the random solution is infeasible to an associated probabilistically-constrained program. An improved bound on this probability is derived in Calafiore and Campi (2006). When using constraint sampling to approximately solve a probabilistically-constrained program, the probability of obtaining a feasible solution is of foremost interest. In contrast, our goal is to approximately solve a probabilistic bicriteria model. Instead of focusing on the probability the chance constraint is violated for a specific level of risk tolerance, we study the trade-off between the expected probability of violation and cost. The bulk of this paper focuses on sampling in a manner that exploits the structure of the single-disruption model to tighten the bound on the expected probability of violation, tightened over that achieved by naive sampling. Apart from the resultant savings in sampling costs, the value of the tightened bound lies in potential improvement in cost for a given risk level. Our contributions are summarized below.

1. We first consider our bicriteria model with a single disruption and derive, through the use of stratified sampling, an improved bound on the expected probability of violation.

Using the new bound, we solve for the optimal sample allocation strategy, and show that it improves on naive sampling by as much as an order of magnitude in the number of time periods.

2. For a single disruption model, we establish the surprising result that even in the absence of exact knowledge of the probability distribution of the time of disruption, a polynomial sampling strategy is order optimal for most well-behaved distributions.

3. Using the perishable inventory problem of Example 1, we computationally demonstrate both a) that stratification gives savings in sampling costs and b) that stratification yields solutions closer on average to the efficient frontier of the bicriteria model than solutions derived from naive sampling.

4. We consider a setting in which the probability distribution on the time of the disruption is chosen by an adversary, subject to box constraints, whose goal is to make our improved bound as bad as possible. This leads to a two-person game with a convex-concave objective function. We use min-max theory to establish that the order of play is immaterial, and that the model can be solved by an intuitive water-filling algorithm.

5. Finally, we extend our improved bound to a setting in which there are two disruptions, and show that the order improvement over naive sampling realized in the single-disruption case holds in this setting as well. Our argument is easily seen to carry over to the general case of a modest number of disruptions.

The remainder of the paper is organized as follows. In Section 2, we develop our model for a single random-time disruption problem. Then, we give an overview of the naive sampling result of Calafiore and Campi adapted to our random-time disruption model. In Section 3.1, we derive our improved bound under stratified sampling. We discuss optimal and order-optimal sampling strategies in Section 3.2 through Section 3.5, highlighting the simultaneous improvement in expected probability of violation and cost. In Section 4, we present an adversarial model for hedging against unknown distributions on the time of disruption. Section 5 extends our results to handle multiple disruptions, and Section 6 summarizes.

2. A Model with a Single Disruption

2.1. Formulation and Assumptions

Consider the bicriteria minimization problem

$$\mathcal{BP}: \quad \operatorname*{vmin}_{x} \quad \left[\sum_{t=1}^{T} h_t(x_t), \, \mathbb{P}\big(f(x, \tilde{\xi}) > 0\big) \right] \tag{1a}$$

$$\text{s.t.} \quad \sum_{t=1}^{T} g_t(x_t) \le b, \tag{1b}$$

$$x_t \in D_t, \ t = 1, \ldots, T, \tag{1c}$$

where $x = (x_1, \ldots, x_T)$. Model $\mathcal{BP}$ has two objective functions, a cost function $\sum_{t=1}^{T} h_t(x_t)$ and a risk function $\mathbb{P}(f(x, \tilde{\xi}) > 0)$. The vector minimization in (1a) means that the set of optimal solutions to $\mathcal{BP}$ is the set of all *Pareto optimal* points for the two functions. In other words, we seek the efficient frontier of the bicriteria problem, namely, the set of all points $(\sum_{t=1}^{T} h_t(x_t), \mathbb{P}(f(x, \tilde{\xi}) > 0))$ with the property that there does not exist $y = (y_1, \ldots, y_T)$ satisfying (1b) and (1c) such that $\sum_{t=1}^{T} h_t(y_t) \le \sum_{t=1}^{T} h_t(x_t)$ and $\mathbb{P}\big(f(y, \tilde{\xi}) > 0\big) \le \mathbb{P}\big(f(x, \tilde{\xi}) > 0\big)$, and at least one of these inequalities is strict. We typically restrict the cost function or the risk function to some reasonable range rather than forming the entire efficient frontier of solutions.

We endow $\mathcal{BP}$ with the following structure. For every t, $h_t(\cdot)$ and $g_t(\cdot)$ are convex functions on D_t, which is a closed, convex set. The random vector $\tilde{\xi}$ has support Γ. We assume $f(\cdot, \xi)$ is convex on $\Pi_{t=1}^{T} D_t$, for every ξ in Γ. We interpret x_t as the decision vector for the t^{th} time period, and denote the dimension of x_t by $\dim(x_t)$.

We assume that the random vector $\tilde{\xi}$ has the form $\tilde{\xi} = (\tilde{I}, \tilde{\delta})$, where the binary vector $\tilde{I} = (\tilde{I}_1, \ldots, \tilde{I}_T)$ indicates the time of disruption, t, via $\tilde{I}_t = 1$, and $\tilde{\delta} = (\tilde{\delta}_1, \ldots, \tilde{\delta}_T)$ with $\tilde{\delta}_t$ being a random vector with support Γ_t. The notion that $\mathcal{BP}$ is a model with at most one disruption is captured by interpreting T as an artificial time period, in conjunction with $\sum_{t=1}^{T} \tilde{I}_t = 1$. The random vector $\tilde{\delta}_t$ represents the magnitude of the disruption given that it occurs in time period t. So, given a realization of $\tilde{\xi}$, we know when the disruption occurs as well as its magnitude. We let p_t denote the probability that the disruption occurs in time period t, i.e., $p_t = \mathbb{P}(\tilde{I}_t = 1), \ t = 1, \ldots, T$.

Assumption 1. The function f separates via $f(x, \tilde{\xi}) = \sum_{t=1}^{T} \tilde{I}_t \, f_t(x_t, \tilde{\delta}_t)$.

Given that the realizations of $\tilde{I}$ are unit vectors, Assumption 1 implies that convexity of $f(\cdot, \xi)$ for every $\xi \in \Gamma$ is equivalent to convexity of $f_t(\cdot, \delta_t)$ for every $\delta_t \in \Gamma_t,\ t = 1, \ldots, T$. Assumption 1 endows our model with additional structure that we exploit in deriving our main result in Section 3.1.

To illustrate the above model, we formulate the perishable inventory problem described in Example 1.

> **Example 1 (continued).** Let h_t and l_t respectively denote unit holding and penalty costs for unsatisfied nominal demand of the product, and let d_t denote this nominal demand, all in period $t,\ t = 1, \ldots, T$. Let $\tilde{\delta}_t$ denote the random demand given that the disruption occurs in period t. The production cost in each time period is a convex piecewise linear function with two pieces, joined at the cut-off level $c_t,\ t = 1 \ldots, T$, i.e., production in period t incurs a unit cost r_t up to c_t units and a unit cost $r'_t > r_t$ beyond. We now have the following model.
>
> $$\begin{aligned} \operatorname*{vmin}_{x,y,u,s,\nu} \quad & \left[\sum_{t=1}^{T} (r_t u_t^1 + r'_t u_t^2 + h_t s_t + l_t \nu_t),\ \mathbb{P}\left(\sum_{t=1}^{T} \tilde{I}_t (\tilde{\delta}_t - x_t) > 0 \right) \right] \\ \text{s.t.} \quad & s_{t-1} + u_t^1 + u_t^2 = y_t + x_t + s_t,\ t = 1, \ldots, T, \\ & s_0 = 0,\ s_T = 0, \\ & \nu_t \geq d_t - y_t,\ t = 1, \ldots, T, \\ & u_t^1 \leq c_t,\ t = 1, \ldots, T, \\ & x_t, y_t, u_t^1, u_t^2, s_t, \nu_t \geq 0,\ t = 1, \ldots, T. \end{aligned}$$
>
> Here, the primary decision variables are u_t^1 and u_t^2, respectively denoting production amounts capped by, and in excess of, c_t; y_t denoting the amount used to meet nominal demand; x_t denoting the amount used to meet random demand; and surplus s_t stored for use in period $t + 1$. We start with zero stock and end with zero stock. The amount of unsatisfied nominal demand is denoted by ν_t. We revisit this model in the context of computational results in Section 3.4.

Model $\mathcal{BP}$ can be used to describe stochastic programs with, and without, recourse. In both cases, the decisions x_t, $t = 1, \ldots, T$, are *static*, i.e., they are made at the beginning before the disruption occurs. When there is no recourse, the decisions x_t, $t = 1, \ldots, T$, are implemented in successive time periods regardless of the occurrence of a disruption. In this sense, the disruption is an event that does not affect the dynamics of decision making but only affects whether $f(x, \tilde{\xi}) \leq 0$ holds. When there is recourse, the decisions dictated by x are implemented until the time of disruption, say t. The function $f_t(\cdot, \cdot)$ can then represent a recourse function with the following form:

$$
\begin{aligned}
f_t(x_t, \delta_t) = \min_{y_t} \quad & q_t(y_t, \delta_t), && \text{(2a)} \\
\text{s.t.} \quad & r_t(y_t, \delta_t) \leq -w_t(x_t, \delta_t), && \text{(2b)} \\
& y_t \in B_t(\delta_t), && \text{(2c)}
\end{aligned}
$$

where $q_t(\cdot, \delta_t)$ and $r_t(\cdot, \delta_t)$ are convex functions on the closed convex set $B_t(\delta_t)$, and $w_t(\cdot, \delta_t)$ is convex on D_t, for every $\delta_t \in \Gamma_t$. The function $q_t(y_t, \delta_t)$ captures the cost of recourse decisions made after the disruption occurs. The functions $r_t(y_t, \delta_t)$ and $w_t(x_t, \delta_t)$ are vector valued with constraint (2b) restricting recourse decisions based on both the disruption and the first stage decision, x_t. The decision vector y_t could, for example, have dimension $\dim(x_{t+1}) + \cdots + \dim(x_T)$ and represent a recourse decision from period t to the horizon. In this case, when the disruption occurs in time period t, decisions $x_{t+1}, \ldots, x_T$ become irrelevant and their cost should not be counted. This can be captured by introducing probability masses p_t in $\mathcal{BP}$ and setting the cost objective function in (1a) to $\sum_{t=1}^{T} \big(\sum_{s=t}^{T} p_s\big) h_t(x_t)$, to capture the expected cost of the non-recourse decisions. If the constraint in (1b) represents a budget constraint, it can be a "hard" constraint, as it is currently stated, or it can similarly capture an expected value budget constraint by introducing probability masses p_t. In Section 3.4, and the online supplement, we develop Example 2, extending Example 1 to a recourse model and allowing us to make the above recourse formulation concrete.

2.2. A Convex Approximation by Constraint Sampling

Assume we can draw independent and identically distributed (i.i.d.) observations from the distributions of $\tilde{\xi}$ and $\tilde{\delta}_t$, $t = 1, \ldots, T$. An observation drawn from the distribution of $\tilde{\delta}_t$ is simply conditioned on $\tilde{I}_t = 1$. Drawing M i.i.d. observations of $\tilde{\xi}$, denoted by $\tilde{\xi}^j$, $j =$

$1, \ldots, M$, we can form the convex program

$$\begin{aligned}
\mathcal{RP}^M: \qquad \min_x \quad & \sum_{t=1}^{T} h_t(x_t) \\
\text{s.t.} \quad & \sum_{t=1}^{T} g_t(x_t) \leq b, \\
& f(x, \tilde{\xi}^j) \leq 0, \; j = 1, \ldots, M, \qquad (3) \\
& x_t \in D_t, \; t = 1, \ldots, T.
\end{aligned}$$

The convex program $\mathcal{RP}^M$ replaces the risk function in $\mathcal{BP}$ with the sampled constraints (3). Under the j^{th} observation $\tilde{\xi}^j = (\tilde{I}^j, \tilde{\delta}^j)$, the sampled constraint (3) has the form $f_t(x_t, \tilde{\delta}_t^j) \leq 0$, where $t = \sum_{t'=1}^{T} t' \tilde{I}_{t'}^j$ is the time period in which the disruption occurs. We assume that $\mathcal{RP}^M$ is feasible and attains its optimal value, w.p.1. (See Calafiore and Campi (2005) for a discussion of the case in which $\mathcal{RP}^M$ violates these conditions.) Note that when $f_t(\cdot, \cdot)$ is a recourse function as in (2), constraint (3) can be replaced by constraints (2b), (2c) and the constraint $q_t(y_t, \delta_t^j) \leq 0$ where $t = \sum_{t'=1}^{T} t' \tilde{I}_{t'}^j$.

Calafiore and Campi (2005) establish the following result, which we express in our setting.

Theorem 1. (Calafiore and Campi) *Let M be a non-negative integer. And, let $\tilde{\xi}^1, \ldots, \tilde{\xi}^M$, i.i.d. from the distribution of $\tilde{\xi}$, be used to define $\mathcal{RP}^M$. Assume that $\mathcal{RP}^M$ has a unique optimal solution $\tilde{x}^M$, or that an appropriate tie-breaking rule is applied to obtain $\tilde{x}^M$. Then,*

$$\mathbb{E}\left[\mathbb{P}\big(f(\tilde{x}^M, \tilde{\xi}) > 0\big)\right] \leq \frac{1}{M+1} \sum_{t=1}^{T} \dim(x_t), \qquad (4)$$

where the expectation is with respect to $(\tilde{\xi}^1, \ldots, \tilde{\xi}^M)$.

Theorem 1 specifies the sense in which we may view $\mathcal{RP}^M$ as an approximation of $\mathcal{BP}$: The bicriteria model has two objective functions, cost and risk which are simultaneously minimized in the Pareto sense. As the number of observations, M, grows large, greater emphasis is placed on minimizing risk. $\mathcal{RP}^M$ enables us to study the trade-off in cost in such an approach while enjoying the advantage of being computationally tractable.

We refer to $\tilde{x}_M$ as a *candidate solution* for $\mathcal{BP}$. For any candidate solution, $\tilde{x}_M$, we refer to $\mathbb{P}\big(f(\tilde{x}_M, \tilde{\xi}) > 0\big)$ as the *probability of violation* for $\tilde{x}_M$. Theorem 1 quantifies the expected probability of violation for the solution of $\mathcal{RP}^M$. By ranging the value of M in $\mathcal{RP}^M$, we can produce an approximation of $\mathcal{BP}$'s efficient frontier. Of course, $\mathcal{BP}$ is a nonconvex optimization model because $\mathbb{P}\big(f(x,\tilde{\xi})>0\big)$ is in general, not a convex function. So, there are limits to what one can expect from the convex approximating model $\mathcal{RP}^M$.

Drawing observation $\tilde{\xi}^j$ leads to a constraint that only involves the time period for which $\tilde{I}_t = 1$. So, with appropriate reindexing, constraints (3) can be rewritten as

$$f_t(x_t, \tilde{\delta}_t^j) \;\; \leq \;\; 0, \; t=1,\ldots,T, \; j \in J_t,$$

where $J_t = \{j \mid \tilde{I}_t^j = 1\}$. For each t, let N_t be the cardinality of J_t so that $\sum_{t=1}^T N_t = M$. Each N_t is a random variable that counts the number of observations of $\tilde{\xi}$ for which $\tilde{I}_t = 1$. We focus on the following question in the next section: Given the special structure of the single disruption model, can we improve on the bound (4) of Theorem 1 by controlling the sizes N_t? Or, equivalently, if we want the expected probability of violation to not exceed ϵ, can we achieve this with fewer observations than that prescribed by (4)?

We approach this issue as follows: Suppose that we draw N_t i.i.d. observations of $\tilde{\delta}_t$, $t = 1,\ldots,T$, with $\sum_{t=1}^T N_t = M$. In contrast to Theorem 1's setting, here we draw observations in a stratified manner so that N_t is deterministic. We show in Section 3.1 that the structure of the single disruption model $\mathcal{BP}$ enables derivation of an improved bound via stratification.

3. Stratified Sampling

3.1. An Improved Bound

The following result provides an analog of Theorem 1 when we draw samples in a stratified manner. This new bound is used in Section 3.2 to develop alternative sampling strategies. The proofs of all of our formal results, including that of Theorem 2, are provided in an online supplement to this paper.

Theorem 2. *Let* $\mathbf{N} = (N_1,\ldots,N_T)$ *be a vector of non-negative integers satisfying* $\sum_{t=1}^T N_t = M$. *And, let* $\tilde{\delta}_t^1,\ldots,\tilde{\delta}_t^{N_t}$, *i.i.d. from the distribution of* $\tilde{\delta}_t$, $t=1,\ldots,T$, *be used to define the convex program*

$$\begin{aligned}
\mathcal{RP}^{\mathbf{N}}: \qquad & \min_x \;\; \sum_{t=1}^T h_t(x_t) \\
& \text{s.t.} \;\; \sum_{t=1}^T g_t(x_t) \leq b, \\
& \qquad f_t(x_t, \tilde{\delta}_t^j) \leq 0, \; t=1,\ldots,T, \; j=1,\ldots,N_t, \\
& \qquad x_t \in D_t, \; t=1,\ldots,T.
\end{aligned}$$

Assume that $\mathcal{RP}^{\mathbf{N}}$ has a unique optimal solution $\tilde{x}^{\mathbf{N}}$, or that the two-norm tie-breaking rule is applied to obtain $\tilde{x}^{\mathbf{N}}$. Then,

$$\mathbb{E}\left[\mathbb{P}\big(f(\tilde{x}^{\mathbf{N}},\tilde{\xi}) > 0\big)\right] \leq \sum_{t=1}^{T} p_t \frac{\dim(x_t)}{N_t+1}, \tag{5}$$

where the expectation is with respect to $(\tilde{\delta}_1^1,\ldots,\tilde{\delta}_1^{N_1},\ldots,\tilde{\delta}_T^1,\ldots,\tilde{\delta}_T^{N_T})$.

The potential value of Theorem 2 lies in choosing the sample sizes N_t in a stratified manner so as to tighten the bound of (5) over that achieved by (4). This is discussed in the next section.

Theorem 2 uses a two-norm tie breaking rule when $\mathcal{RP}^{\mathbf{N}}$ has multiple optimal solutions, and Theorem 1 similarly points to a tie-breaking rule. The role of the two-norm tie-breaking rule is further discussed in the paper's online supplement.

3.2. Optimal Sampling Strategies

In this section, we design sampling allocations, $N_1,\ldots,N_T$, using bound (5) on the expected probability of violation from Theorem 2 to guide our sample allocation. Given a sampling budget M, we let $M = NT$ to indicate that individual time periods receive an average allocation of N observations. In this setting, we seek to solve

$$\begin{aligned}
\mathcal{SS}_{\text{int}}: \quad & \min_{N_1,\ldots,N_T} \quad \sum_{t=1}^{T} \frac{p_t n_t}{N_t+1} && \text{(6a)}\\
& \text{s.t.} \quad \sum_{t=1}^{T} N_t = NT, \\
& \qquad\quad N_t \in \mathbb{Z}_+, \quad t=1,\ldots,T, && \text{(6b)}
\end{aligned}$$

where $\mathbb{Z}_+$ is the set of non-negative integers, and $n_t = \dim(x_t)$, $t = 1,\ldots,T$. For small $M = NT$, the objective function in (6a) would involve terms of the form $p_t\min(1, n_t/(N_t{+}1))$ but for large M, (6a) applies, since the optimal value shrinks to zero as M grows. We denote by $\mathcal{SS}$ the continuous relaxation of $\mathcal{SS}_{\text{int}}$, i.e., $\mathcal{SS}$ is identical to model $\mathcal{SS}_{\text{int}}$ except that constraints (6b) are replaced by $N_t \geq 0$, $t = 1,\ldots,T$. The objective function for $\mathcal{SS}$ is strictly convex in $(N_1,\ldots,N_T)$ and its feasible region is convex and compact. The unique optimal solution to $\mathcal{SS}$ and its optimal value $V^*(N)$ are given by

$$N_t^* = \left(\frac{(p_t n_t)^{1/2}}{\sum_{t=1}^T (p_t n_t)^{1/2}}\right)(N+1)T - 1,\ t = 1, \dots, T, \tag{7}$$

$$V^*(N) = \frac{\left(\sum_{t=1}^T (p_t n_t)^{1/2}\right)^2}{(N+1)T}. \tag{8}$$

For practical implementation, we assume that the sample sizes are sufficiently large that we may round the values in (7) to achieve integer-valued allocations. For the analysis that follows, we neglect such issues and simply consider the continuous allocation.

3.3. Improvement over Naive Sampling

The proportion $\alpha_t = (p_t n_t)^{1/2} / \sum_{t=1}^T (p_t n_t)^{1/2}$ that appears in (7) is effectively the fraction of our sample budget that is allocated to period t. If we instead perform naive sampling and construct model $\mathcal{RP}^M$, then a random number of observations is allocated to each time period. However, for large M, the proportion of observations in $\mathcal{RP}^M$ allocated to period t converges (by the law of large numbers) to p_t. Hence, it is natural to consider what happens when we replace the optimal proportion in (7) with $\alpha_t = p_t,\ t = 1, \dots, T$. Theorem 3 compares the bound so obtained with the bound in Theorem 1 with $M = NT$, and characterizes the improvement attained by the optimal value of the bound (8).

Theorem 3. *Let $V^1(N) = \sum_{t=1}^T n_t/(NT+1)$ denote the bound (4) on the expected probability of violation with $M = NT$. Let $V^2(N)$ denote the bound (5) of Theorem 2 with $N_t = p_t(N+1)T - 1$. And, let $V^*(N)$ denote the bound (8) of Theorem 2 under sample size allocation (7). Then,*

$$(i) \quad V^*(N) \le V^2(N) \le V^1(N),$$

$$(ii) \quad \frac{V^2(N)}{V^1(N)} \to 1 \text{ as } N \to \infty, \tag{9a}$$

$$(iii) \quad \frac{V^*(N)}{V^2(N)} = \frac{\left(\sum_{t=1}^T (p_t n_t)^{1/2}\right)^2}{\sum_{t=1}^T n_t}. \tag{9b}$$

Moreover, if $n_t = n,\ t = 1, \dots, T$,

$$(iv) \quad \frac{1}{T} \le \frac{V^*(N)}{V^2(N)} \le 1, \tag{9c}$$

and these bounds are tight, i.e., they can be achieved.

Result (9a) of Theorem 3 shows that the bound of Theorem 1 and that of Theorem 2 under the proportional allocation, $N_t = p_t(N+1)T-1$, are essentially identical. Results (9b) and (9c) of the theorem characterize the improvement over these bounds provided by the optimal allocation (7). The factor of improvement can be T^{-1}, i.e., the improvement can be an order of magnitude in the number of time periods. In other words, if $n_t = n,\ t = 1, \ldots, T$, then $V^1(N)$ and $V^2(N)$ shrink to zero with N^{-1}, but $V^*(N)$ can shrink to zero as quickly as $N^{-1}T^{-1}$.

More generally, if $n_t = n,\ t = 1, \ldots, T$, then $V^2(N) = \Theta(1/N)$ as $T \to \infty$ and $V^*(N) = \Theta\left((\sum_{t=1}^T p_t^{1/2})^2/NT\right)$ as $T \to \infty$. For $V^*(N)$ to achieve the rate $N^{-1}T^{-1}$, we require $(\sum_{t=1}^T p_t^{1/2})^2 = \Theta(1)$ as $T \to \infty$. The degenerate distribution which achieves the rate in the proof of Theorem 3 is extreme and arguably not one that would arise in practice.

A more realistic alternative is the following: There is an underlying process that can independently yield a disruption event in each time period with common probability, say $1-q$. We await the first such disruption and neglect (the possibility of) subsequent disruptions. This is analogous to a continuous-time model of machine breakdown in which the arrival of a phantom customer induces a machine breakdown and these customers do not queue; see, e.g., Vinod and Solberg (1984). This argument leads to the following probability mass function for a disruption occurring in period t: $p_t = (1-q)q^{t-1}$, $t = 1, \ldots,$ for $q \in (0, 1)$. Of course, this distribution only applies for an infinite horizon, whereas our model has finite horizon T. One adaptation under a finite horizon is to introduce an additional artificial period to the model, $T+1$. With $p_t = (1-q)q^{t-1},\ t = 1, \ldots, T$, and $p_{T+1} = q^T$, we retain the memoryless property of the geometric distribution. Another natural adaptation is the truncated geometric distribution given by $p_t = (1-q)q^{t-1}/(1-q^T),\ t = 1, \ldots, T$. Note that these are indeed valid distributions—all probability masses are non-negative and sum to unity in both cases.

Part (ii) of Theorem 3, coupled with (7), implies that when the disruption is equally likely to occur in any of the T time periods the optimal allocation behaves similarly to naive sampling, i.e., stratified sampling does not improve the naive bound. This uniform distribution arises if we again posit that a disruption event can occur independently in each time period with common probability, and we then *condition* on one such event occurring up to time T. We view the geometric distribution as more natural: The conditioning that leads to the uniform distribution requires that the underlying "true" distribution governing the time of the first disruption be affected by our "modeling choice" of the time horizon, T.

The following corollary establishes that the $N^{-1}T^{-1}$ rate is achieved by these variants of the geometric distribution.

Corollary 4. *Let $V^*(N)/V^2(N)$ be given by equation (9b), assume $n_t = n,\ t = 1, \ldots, T$, and let $q \in (0, 1)$. If the probability mass function governing the time of the disruption satisfies*

$$p_t = (1-q)q^{t-1},\ t = 1, \ldots, T,\ p_{T+1} = q^T \tag{10a}$$

or

$$p_t = \frac{(1-q)q^{t-1}}{1-q^T},\ t = 1, \ldots, T, \tag{10b}$$

then $V^*(N)/V^2(N) = \Theta(1/T)$ as $T \to \infty$.

3.4. Computational Examples

Bounds (4) and (8) can be used to assess the proximity to the efficient frontier of solutions obtained under naive and optimal stratified sampling. Setting the naive sampling bound (4) equal to a risk level ϵ in the range of interest, with $M = NT$, we obtain

$$\frac{1}{NT+1}\sum_{t=1}^{T} n_t = \epsilon.$$

Similarly, setting the optimal allocation bound (8) to ϵ yields

$$\frac{\left(\sum_{t=1}^{T}(p_t n_t)^{1/2}\right)^2}{(N+1)T} = \epsilon.$$

In order to ensure that the expected probability of violation is bounded by ϵ, we require therefore that

$$N \geq \left\lceil \frac{1}{T}\left(\frac{1}{\epsilon}\sum_{t=1}^{T} n_t - 1\right)\right\rceil \tag{11}$$

for the naive strategy, and

$$N \geq \left\lceil \frac{1}{\epsilon T}\left(\sum_{t=1}^{T}(p_t n_t)^{1/2}\right)^2 - 1\right\rceil \tag{12}$$

for the optimal stratified strategy. The convex approximations $\mathcal{RP}^{NT}$ and $\mathcal{RP}^{\mathbf{N}}$ can be solved with N equal to the lower bounds (11) and (12), to yield candidate solutions. By repeating this procedure and averaging, we obtain estimates of the associated expected cost. We present two example problems.

Example 1. For our first computational example, we consider an instance of the inventory control model of Example 1 from Section 2.1 with $T = 20$. The time of disruption is assumed to obey a truncated geometric distribution as given in (10b), with $q = 0.8$. We assume that the random variables $\tilde{\delta}_t$ obey a Pareto distribution with a shape parameter, k, equal to 3, and a threshold ψ_0, equal to 1. This implies we have $\mathbb{P}(\tilde{\delta}_t > x) = (\psi_0/x)^{k-1}$ for $x \geq \psi_0$ and 1 otherwise, for $t = 1, \ldots, T$. Given this distribution, we impose the condition that the allocation x_t must be at least the threshold value of ψ_0. Thus, the model includes the additional constraints $x_t \geq \psi_0,\ t = 1, \ldots, T$. The nominal demand in each time period is set equal to 0.5. Costs decay geometrically over time; we set $h_1 = 3, r_1 = 1,\ r'_1 = 1.5, l_1 = 1000$, with a common decay rate of 0.9. The cut-off level c_t is chosen to be 3 for all t. We emphasize that the actual values of these parameters are unimportant as long as T is large enough to capture the asymptotic behavior discussed in Section 3.3.

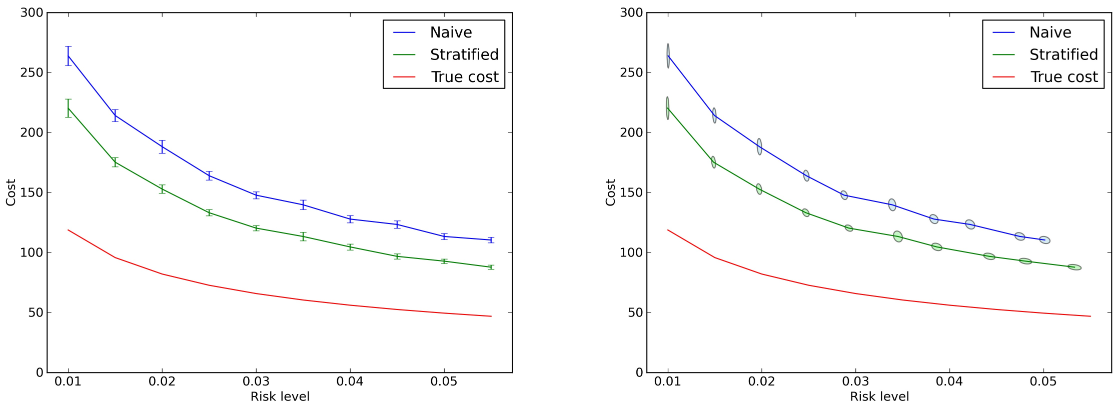
Figure 1 graphically depicts the results for Example 1. Both the nonlinear bicriteria model and the convex approximations were solved using GAMS, with calls to MINOS and CPLEX. Of primary interest is Figure 1a, wherein the axis labeled "Risk level" depicts the *a priori* bounds of Theorem 1 for naive sampling and Theorem 2 for stratified sampling. The axis labeled "Cost" depicts the expected objective function value of the (random) solution we obtain from solving model $\mathcal{RP}^M$ or $\mathcal{RP}^{\mathbf{N}}$, respectively, estimated by averaging over 2000 i.i.d. observations of $\tilde{x}^M$ and $\tilde{x}^{\mathbf{N}}$. Figure 1b depicts a *posterior* calculation of risk. Here, for naive sampling, the inner probability in $\mathbb{E}\left[\mathbb{P}\big(f(\tilde{x}^M, \tilde{\xi}) > 0\big)\right]$ is calculated exactly for a given $\tilde{x}^M$, and the outer expectation is estimated by a sample mean of 2000 i.i.d. observations of $\tilde{x}^M$. The same procedure is analogously carried out for stratified sampling, and Figure 1b also depicts 0.95-level cost–*posterior* risk confidence ellipsoids.

We report detailed results of the experiment in Table 1 for Figure 1a. For this example, we have $n_t = 1$ for $t = 1, \ldots, T$, because only a single (scalar) decision variable is required to define $f_t(x_t, \delta_t)$. Table 1 also reports statistics on the difference in cost between stratified and naive sampling when the samples are generated using common random numbers. The table shows that stratified sampling has smaller values of N in comparison to naive sampling. Figure 1 and the paired difference statistics, indicate a significant improvement when using optimal stratification over naive sampling. A one-sided paired-t test shows that these differences exceed the values listed under "95% CI Bound" in Table 1 at a 0.95 level of confidence. Comparing the confidence interval widths depicted in Figure 1a for the cost with those we can infer from Table 1 for the cost differences (the difference of the final two columns), we see substantial variance reduction from employing common random numbers. For this example, stratified sampling reduces the expected cost by 16-21%.

Example 2. To indicate further applicability of our model and the associated benefit from stratified sampling, we present a more complex example. At an intuitive level, the inventory control model of Example 1 describes the time-progression of a single facility serving a single customer base. In each time period t, the facility can produce up to c_t units of goods cheaply at a cost of r_t per unit, after which the facility can produce goods at a more expensive cost of r'_t per unit. The facility can also store goods from previous time periods with an associated holding cost. In each time period, the facility must stock its shelves for nominal and disruption demand. Unused goods on the shelf spoil at the end of period t. The random disruption takes the form of increased customer demand in a particular time period.

We extend this model by considering a network of two facilities and two customers as depicted in Figure 2. In each time period, each facility faces the same operational problem as in Example 1, but there are now two customers. Each time period, each facility decides the units of cheap production, expensive production, storage, and shipping to the two customers. Shipping is done without knowing the disruption demand and shipped units cannot be stored for subsequent time periods. In Figure 2, the arrows connecting the facility nodes to the customer nodes represent the possible transport of goods to the customers. The arrows connecting a facility in adjacent time periods represent the storage of product over time. Once again, the random disruption takes the form of increased customer demand, now a two-dimensional vector, in a particular time period.

The extended model allows us to demonstrate two aspects of stratified sampling absent from the first example. First, in Example 2, $n_t = 4$ for all time periods t because four first stage variables are required to define $f_t(x_t, \delta_t)$ in the recourse formulation (2a)-(2c). (This is further detailed in the context of the mathematical formulation in the online supplement.) Second, in the extended model, after a disruption occurs, we introduce recourse variables that take over the nominal production and transportation decisions of the facilities, using the method described at the end of Section 2.1. In particular, after a disruption occurs, the entire system must operate for an additional number (five) of time periods, using new production and transportation decisions and a fixed per-period budget. These recourse variables, and



(a) (b)

Figure 1: Example 1. The bottom line presents the true efficient frontier. The middle line presents the estimated expected cost of solutions found by stratified sampling. The top line presents the estimated expected cost of solutions found by naive sampling. Figure 1a plots the specified *a priori* risk level vs. the expected solution cost. Confidence intervals for the cost for the two top curves are also represented in the figure. Figure 1b plots the *posterior* risk level vs. the expected solution cost. Confidence ellipses for the cost and *posterior* risk level are also represented. More detailed data on this example are in Table 1.

Table 1: Example 1. For this example, stratified sampling reduces the expected cost by about 16-21%. The paired difference columns present point and one-sided interval estimates on the difference in cost between stratified and naive sampling generated using common random numbers.

Risk Level	Opt. Cost	Naive		Stratified		Cost Reduction	Paired Difference	
		N	$\mathbb{E}$[Opt. Cost]	N	$\mathbb{E}$[Opt. Cost]		Mean	95% CI Bound
0.010	118.77	100	263.86	72	220.30	16.51%	43.56	40.73
0.015	95.77	67	214.18	48	175.27	18.17%	38.91	35.96
0.020	82.05	50	188.19	36	152.88	18.77%	35.31	31.68
0.025	72.70	40	164.00	28	133.15	18.81%	30.85	28.54
0.030	65.79	34	147.77	24	120.32	18.57%	27.45	25.65
0.035	60.42	29	139.76	20	113.36	18.89%	26.40	24.77
0.040	56.09	25	127.86	18	104.65	18.15%	23.20	21.64
0.045	52.51	23	123.38	16	96.79	21.55%	26.59	24.63
0.050	49.47	20	113.38	14	92.80	18.15%	20.58	19.02
0.055	46.87	19	110.42	13	87.77	20.52%	22.65	21.19

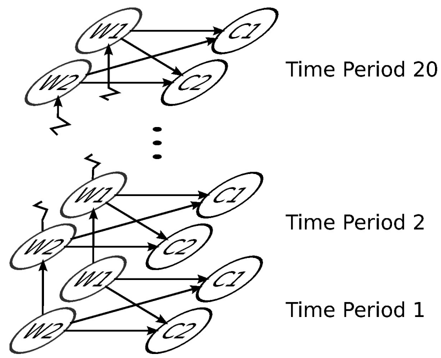


Figure 2: Example 2: Two-facility, two-customer model. The nodes labeled W represent two facilities (warehouses). The nodes labeled C represent two customers. In each time period, facilities produce goods, ship goods to meet customer demand and store goods for future time periods. Shipped goods cannot be stored for future time periods, even if unused by the customers. The arrows connecting facilities to customers represent possible transport from each facility to the customer to meet demand. The arrows connecting each facility to future time periods represent storage of product over time.

the requirement to satisfy a fixed per-period budget, represent the graceful shutdown of the system after a disruption. The paper's online supplement contains the full model and numerical values for the instance used to obtain the results below.

Analogous to the results presented above for Example 1, Figure 3 graphically depicts the optimized costs resulting from solving 200 replications of the stratified-sampling and naive-sampling models for Example 2. Table 2 provides detailed data from the computational experiment, including statistics on the paired differences in cost. As demonstrated by the paired difference statistics, stratified sampling provides a significant improvement in cost.

Choosing the time horizon T is an important modeling choice in the types of models we consider. Foremost, the horizon should be chosen so that focusing attention on a single (or small number of) disruption(s) is reasonable. To achieve the order of magnitude improvement characterized by Corollary 4, and subsequently by Theorems 5 and 6, T must be large enough so that these asymptotic results apply. Consider an infinite horizon problem and let p_t be given by a geometric distribution with $q = 0.8$. Then, $\sum_{t=21}^{\infty} p_t \approx 0.009$. Suppose we wish to enforce a bound on risk of $\epsilon = 0.05$ in the infinite horizon problem. Then, we can solve a model with a horizon of $T = 20$ (as in Examples 1 and 2) and $\epsilon = 0.041$, effectively "sacrificing" low-probability scenarios in periods $T = 21, 22, \ldots$. Such an approach bounds the size of the problem that we must solve for an infinite horizon case and can also provide guidance in choosing T.

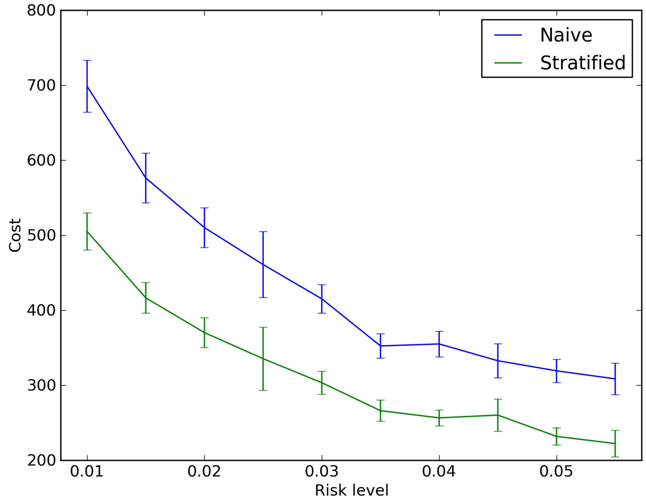


Figure 3: Example 2. The bottom line presents the estimated expected cost of solutions found by stratified sampling. The top line presents the estimated expected cost of solutions found by naive sampling. The vertical bars present 95% confidence intervals. More detailed data on this example are in Table 2.

Table 2: Example 2. For this example, stratified sampling reduces the expected cost by 21-28%. The paired difference columns present point and one-sided interval estimates on the difference in cost between stratified and naive sampling generated using common random numbers.

Risk Level	Naive		Stratified		Cost Reduction	Paired Difference	
	N	$\mathbb{E}$[Opt. Cost]	N	$\mathbb{E}$[Opt. Cost]		Mean	95% CI Bound
0.010	400	698.61	289	505.13	27.69%	193.48	170.16
0.015	267	576.35	192	416.49	27.74%	159.86	138.48
0.020	200	510.17	144	370.20	27.44%	139.97	125.22
0.025	160	460.92	115	335.41	27.23%	125.51	110.67
0.030	134	415.22	96	303.36	26.94%	111.87	100.64
0.035	115	352.39	82	266.11	24.48%	86.28	77.89
0.040	100	355.07	72	256.52	27.76%	98.56	86.29
0.045	89	332.68	64	260.20	21.79%	72.48	64.93
0.050	80	319.25	57	231.87	27.37%	87.38	77.45
0.055	73	308.53	52	222.23	27.97%	86.30	75.68

3.5. Comparing Allocation Strategies

In this section, we study how other allocation strategies compare with the optimal allocation (7) as T grows large. We consider sample allocations of the form $N_t = \alpha_t(N+1)T - 1,\ t = 1, \ldots, T$, satisfying $\sum_{t=1}^{T} \alpha_t = 1,\ \alpha_t \geq 0,\ t = 1, \ldots, T$. Under the assumption $n_t = n,\ t = 1, \ldots, T$, the bound (5) under such an allocation reduces to $\frac{n}{(N+1)T} \sum_{t=1}^{T} p_t/\alpha_t$. Letting $p = (p_1, \ldots, p_T)$ and defining $w(\alpha;\, p) = \sum_{t=1}^{T} p_t/\alpha_t$, the bound becomes $n\, w(\alpha; p)/\big((N+1)T\big)$. Observe that

1. The optimal allocation vector α^* is unique; α_t^* is proportional to the square root of p_t, and we have

$$\min_{\alpha \in \Delta} \; w(\alpha; p) = \left(\sum_{t=1}^{T} p_t^{1/2} \right)^2, \tag{13}$$

   where $\Delta = \{\alpha \in \mathbb{R}_+^T \mid \sum_{t=1}^{T} \alpha_t = 1\}$.

2. The uniform allocation vector $\alpha_{\text{unif}} = (1/T, \ldots, 1/T)$ and the proportional allocation vector $\alpha_{\text{prop}} = (p_1, \ldots, p_T)$ behave identically with respect to $w(\cdot\,;\, p)$, i.e., $w(\alpha_{\text{unif}};\, p) = w(\alpha_{\text{prop}};\, p) = T$. Thus, there is no difference between allocating samples proportionally (as is done in naive sampling) and allocating samples uniformly.

3. The inequality $w(\alpha_{\text{unif}}\,;\, p) = T \geq w(\alpha^*\,;\, p) = (\sum_{t=1}^{T} {p_t}^{1/2})^2$ is tight if and only if $p_t = 1/T,\ t = 1, \ldots, T$. In other words, the uniform allocation is optimal if and only if the distribution governing the time of occurrence of the disruption is also uniform.

The observation $w(\alpha_{\text{unif}};\, p) = w(\alpha_{\text{prop}};\, p)$ suggests $w(\cdot\,;\, p)$ is symmetric when α is proportional to a power of p, i.e., when $\alpha = \alpha(\gamma)$ with $\alpha_t(\gamma) = p_t^\gamma / \sum_{t=1}^{T} p_t^\gamma,\ t = 1, \ldots, T$, where $\gamma \in \mathbb{R}$. Let

$$\begin{aligned} u(\gamma) \; = w(\alpha(\gamma)\,;\, p) \;\; &= \;\; \sum_{t=1}^{T} (p_t)^\gamma \sum_{t=1}^{T} \frac{p_t}{(p_t)^\gamma} \\ &= \;\; \sum_{t=1}^{T} (p_t)^\gamma \sum_{t=1}^{T} (p_t)^{1-\gamma}. \end{aligned}$$

Thus, $u(\cdot)$ is symmetric about $\gamma = 1/2$. Further, as T grows large, we restrict attention to non-degenerate choices of p and γ such that (a) $(p_t)^\gamma$ decays fast enough to be summable in T, and (b) $(p_t)^\gamma$ decays slow enough that $p_t^{1-\gamma}$ is summable in T. Note that if conditions (a)

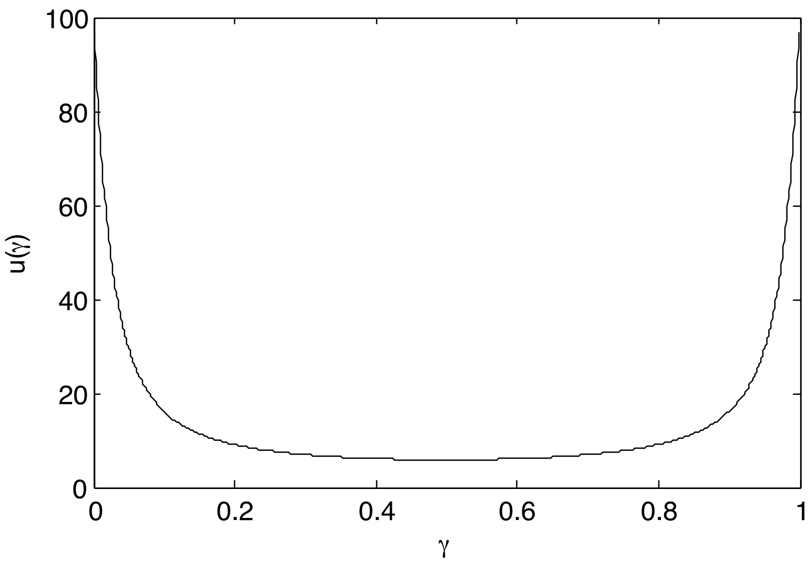


Figure 4: Plot of $u(\gamma)$ showing dependence of bound on the expected probability of violation on allocations N_t that are proportional to the γ^{th} power of the masses p_t, when $p_t = (1-q)q^{t-1}/(1-q^T)$, $t = 1\dots,T$, with $T = 100$ and $q = 0.5$.

and (b) hold for some vectors p and α, then $u(\gamma) = \Theta(1)$ as $T \to \infty$, and we achieve the rate of $N^{-1}T^{-1}$ achieved by α^* in bound (5). Recalling that $(\sum_{t=1}^T p_t^{1/2})^2 \geq 1$ for all choices of p, it follows that these values of γ are order-optimal in the sense of the optimization in (13). Now, if $\gamma \leq 0$, condition (a) is not satisfied, while if $\gamma \geq 1$, condition (b) is not satisfied. However, if the masses p_t decay geometrically, then, for any value of $\gamma \in (0,1)$, conditions (a) and (b) both hold. Figure 4 illustrates this behavior. However, if the masses p_t decay polynomially, then the range of γ values for which conditions (a) and (b) hold simultaneously is more restrictive. Theorem 5 formalizes this observation.

Theorem 5. *Assume $n_t = n$, $t = 1,\dots,T$. Further, assume that one of the following holds:*

(i) $p_t \propto t^{-r}$, $t = 1,\dots,T$ for some $r > 2$, and $\gamma \in (1/r, 1-1/r)$

(ii) $p_t \propto q^t$, $t = 1\dots,T$ for some $q \in (0,1)$, and $\gamma \in (0,1)$.

Let $\alpha_t = p_t^\gamma / \sum_{t=1}^T p_t^\gamma$, $t = 1,\dots,T$. Then,

$$\frac{w(\alpha\,;\,p)/\big((N+1)T\big)}{V^2(N)} \quad = \quad \Theta(1/T) \;\; \text{as} \;\; T \to \infty.$$

Theorem 5 suggests that for a more general probability mass function p, there may be conditions under which a polynomial allocation, α, yields similar results. In other words, we pose the question: When are the (generalized) conditions (a) $\sum_{t=1}^T \alpha_t$ is summable in T and (b) $\sum_{t=1}^T p_t/\alpha_t$ is summable in T, simultaneously true as T grows large? For example, suppose the distribution of the time of disruption has finite second moment in the limit $T \to \infty$, i.e., $\sum_{t=1}^T t^2 p_t$ is uniformly bounded. Under this assumption, letting $\alpha_t \propto t^{-2}$ satisfies conditions (a) and (b). More generally, both conditions hold when p has uniformly bounded k^{th} moment, and $\alpha_t \propto t^{-k}$, for some $k > 1$. We state this result in Theorem 6 below.

Theorem 6. *(Bounded Moment Condition) Let $k > 1$ and assume $n_t = n,\ t = 1, \dots, T$. Further, assume*

$$\limsup_{T\to\infty} \sum_{t=1}^{T} t^k p_t < \infty. \tag{14}$$

Let $\alpha_t = t^{-k}/\sum_{t=1}^T t^{-k},\ t = 1, \dots, T$. Then,

$$\frac{w(\alpha; p)/\big((N+1)T\big)}{V^2(N)} \ = \ \Theta(1/T) \ \text{ as } \ T \to \infty.$$

The import of Theorem 6 is the following: If the distribution governing the time of disruption is unknown, we cannot use the sample size allocation of (7). However, if that distribution can be assumed to satisfy the bounded moment condition (14), then the theorem specifies a sample-size allocation that is both order-optimal and improves on that of naive sampling by an order of magnitude in T. We compare the allocations in Theorems 5 and 6 with $V^2(N)$, our surrogate for naive sampling as in Theorem 3 and Corollary 4, but note that the conclusions of these two theorems can be recast as $\frac{w(\alpha\,;p)/\big((N+1)T\big)}{V^*(N)} = \Theta(1)$ as $T \to \infty$.

Furthermore, from Theorems 5 and 6, if p_t decays geometrically, any geometric allocation that decays more slowly than p, as well as any summable polynomially-decaying allocation, is order optimal. Surprisingly, among such order-optimal allocations, a suitable polynomial allocation outperforms a geometric allocation that is far from optimal. For example, let p_t have a truncated geometric distribution, i.e., $p_t = (1-q)q^{t-1}/(1-q^T),\ t = 1, \dots, T$. Letting $q = 1/6$, $T = 100$ and $\gamma = 1/4$ gives $u(\gamma) \approx 3.123$, while letting $\alpha_t = t^{-2}/\sum_{t=1}^T t^{-2},\ t = 1, \dots, T$, gives $w(\alpha; p) \approx 2.747$. This dominance holds for all $\gamma \in (0, 1/4] \cup [3/4, 1)$. This is summarized in the remark below.

Remark. There exists a time horizon T and a truncated geometric distribution p over this horizon for which, under $n_t = n,\ t = 1, \dots, T$, a suitably chosen polynomial allocation outperforms a non-trivial set of geometric allocations in the sense of bound (5).

In summary, we observe: If p is unknown, or subject to perturbation, but decays fast enough to have a bounded k^{th} moment, Theorem 6 suggests a polynomial allocation. The next section considers the case when we cannot make this assumption, and the distribution of the time of disruption is selected to make our choice of α as bad as possible.

4. An Adversarial Problem

In this section, we consider perturbations of p that are more severe than those considered above. In particular, we assume an adversary selects the distribution governing the time of disruption. We model this by considering a game involving two players, the sampler and the adversary. The sampler picks the allocation vector α first, and then the adversary picks the distribution p. The sampler seeks to minimize the expected probability of violation while the adversary's goal is to maximize the same. For simplicity, we assume $n_t = n,\ t = 1, \ldots, T$. If the adversary is allowed to choose any probability mass function p_t, $t = 1, \ldots, T$, then the problem is devoid of interest since the adversary simply chooses $t' \in \operatorname{argmin}_{1 \le t \le T} \alpha_t$ and sets $p_{t'} = 1$. However, we assume we have some information concerning the distribution of the time of disruption that allows us to restrict the probability mass function to lie in a proper subset of the simplex Δ. Specifically, we consider the case in which intervals $[p_t^l, p_t^u]$ are specified, where $0 < p_t^l \le p_t^u < 1,\ t = 1, \ldots, T$. We therefore have the adversarial min-max formulation

$$\mathcal{AP}: \quad \min_{\alpha \in \Delta} \Big(\max_{p \in \Delta \cap B} \ \sum_{t=1}^{T} \frac{p_t}{\alpha_t} \Big),$$

where $B = [p_1^l, p_1^u] \times \cdots \times [p_T^l, p_T^u]$ and where we assume that $\Delta \cap B \neq \emptyset$.

The following proposition is based on the theory of min-max optimization and saddle points; see, for example, Bertsekas et al. (2003).

Theorem 7. *Consider problem $\mathcal{AP}$, a two-person zero-sum game involving a sampler who chooses α, and an adversary who chooses p. Then,*

(a) $\mathcal{AP}$ is equivalent to the problem

$$\max_{p \in \Delta \cap B} \Big(\min_{\alpha \in \Delta} \ \sum_{t=1}^{T} \frac{p_t}{\alpha_t} \Big) = \max_{p \in \Delta \cap B} \Big(\sum_{t=1}^{T} {p_t}^{1/2} \Big)^2, \tag{15}$$

i.e., the order of optimization is immaterial.

(b) *The sampler has a unique optimal strategy no matter who plays first. The adversary has a unique optimal solution when he plays first, but may have multiple optimal strategies when the sampler plays first.*

(c) *The uniform probability vector* $p_{\text{unif}} = (1/T, \cdots, 1/T)$ *lies in* B *and is optimal for the adversary if and only if* α_{unif} *is optimal for the sampler. Further,* p_{unif} *is optimal for the adversary if and only if all feasible solutions* $p \in \Delta \cap B$ *are optimal for the adversary when he plays second.*

Remark. Theorem 7 shows that the adversary has no advantage in playing second, i.e., we can solve the easier problem of optimal allocation after the adversary chooses p. It also shows that if the adversary can choose the uniform distribution, stratification is of no value to the sampler, at least in the sense of (5), for large sample sizes. See also the discussion prior to Corollary 4.

The concavity of the objective function on the right-hand side of (15) suggests an algorithm based on a "water-filling" type of argument. The algorithm, given below, initializes all p_t values at their lower bounds, p_t^l, then increases the values of the smallest p_t (increasing multiple components simultaneously if there is a tie) until the condition $\sum_{t=1}^T p_t = 1$ is satisfied, subject to the upper bounds, p_t^u.

Algorithm 1

Input: $p_t^l,\, p_t^u,\ t = 1, \ldots, T.$

Output: Optimal solution p^*.

Initial Step. Set $p_t = p_t^l,\ t = 1, \ldots, T,\ \ s = \sum_{t=1}^T p_t.$

If $s > 1$, the problem is infeasible.

If $s = 1$, set $p_t^* = p_t,\ t = 1, \ldots, T$, and terminate.

Iterative Step. Set $S = \operatorname{argmin}_{1 \le t \le T} p_t$. Simultaneously increase p_t for $t \in S$, updating s until one or more of the following happens:

(i) $s = 1$, upon which set $p_t^* = p_t,\ t = 1, \ldots, T$, and terminate.

(ii) The set S increases in cardinality, i.e., the value of p_t for $t \in S$ coincides with $p_{t'}^l$ for some $t' \notin S$. Let $S = S \cup \{t' \mid t' \notin S,\, p_{t'}^l = p_t \,\forall\, t \in S\}$.

(iii) The set S decreases in cardinality, i.e., the value of p_t coincides with p_t^u for some $t \in S$. Let $S' = \{t \in S \mid p_t = p_t^u\}$. Let $p_t^* = p_t,\ t \in S'$ and

$S = S \setminus S'$. Continue. $\square$

Given the output p^* from Algorithm 1, we now select sample sizes according to equation (7), i.e., the fraction of samples allocated to period t is $\alpha_t = [p_t^*]^{1/2} / \sum_{t=1}^T [p_t^*]^{1/2}$.

5. Multiple Disruptions

In this section, we allow multiple disruptions instead of just one, and show that the results of Section 2 extend to two or more disruptions. We derive bounds on the expected probability of violation when there are ℓ disruptions, where ℓ is small compared to T. We establish, analogous to the single disruption model, that stratified sampling yields a bound that is $\Theta(1/NT^{\ell})$, while naive sampling yields a bound that is $\Theta(1/NT^{\ell-1})$. Thus, when the number of disruptions is small when compared to the horizon T, stratification reduces the sample size requirement by an order of magnitude in T. For simplicity of presentation, our results below are for the case of two disruptions. Extensions to the case where there are more than two such events are straightforward.

5.1. A Two-Disruption Model

The model for two disruptions is identical to that of the $\mathcal{BP}$ model (1), except that instead of assuming $\sum_{t=1}^T \tilde{I}_t = 1$, we now assume $\sum_{t=1}^T \tilde{I}_t = 2$. The random vector $\tilde{\xi}$ has the form $\tilde{\xi} = (\tilde{I}, \tilde{\delta})$, with $\tilde{I} = (\tilde{I}_1, \dots, \tilde{I}_T)$ and $\tilde{\delta} = (\tilde{\delta}_1, \dots, \tilde{\delta}_T)$. We define

$$S = \{(t,s) \mid t = 1, \dots, T-1,\ s = t+1, \dots, T\},$$

and let $p_{t,s} = \mathbb{P}(\tilde{I}_t = 1, \tilde{I}_s = 1)$, $(t,s) \in S$, where $\sum_{(t,s)\in S} p_{t,s} = 1$. And, given that $\tilde{I}_t = \tilde{I}_s = 1$, $\tilde{\delta}_t$ and $\tilde{\delta}_s$ denote respectively, the magnitude of the disruptions in periods t and s, for $(s,t) \in S$. In the single-disruption model, we allow for the possibility that no disruption occurs by having period T as an artificial time period. The same idea allows the two-disruption model to capture having at most two disruptions. To do so, we simply introduce two artificial time periods, to capture the cases in which there is exactly one disruption and no disruptions. We allow the vectors $\tilde{\delta}_t$ and $\tilde{\delta}_s$ to be dependent.

In place of Assumption 1, we instead place the following structure on $f(x, \tilde{\xi})$:

Assumption 2. The function f separates via

$$f(x,\tilde{\xi}) = \sum_{(t,s)\in S} \tilde{I}_t \tilde{I}_s \max\{\, f_t(x_t,\tilde{\delta}_t), f_s(x_s,\tilde{\delta}_s)\,\}.$$

The "max" operator in the above assumption provides a notationally convenient way of expressing the joint pair of constraints $f_t(x_t,\tilde{\delta}_t) \le 0$ and $f_s(x_s,\tilde{\delta}_s) \le 0$, which we enforce in the sampled convex program $\mathcal{RP}^{\mathbf{N}}$ below.

We assume we can draw i.i.d. observations from the distribution of $\tilde{\xi}$ and the joint distribution of $(\tilde{\delta}_t,\tilde{\delta}_s)$, $(t,s)\in S$. When we sample $\tilde{\xi}^j = (\tilde{I}^j,\tilde{\delta}^j)$, $j=1,\ldots,M$, under the single-disruption assumption, the j^{th} constraint reduces to $f_t(x_t,\tilde{\delta}_t^j)\le 0$, where t is the single period indicated by $\tilde{I}^j$. Under the two-disruption assumption, that constraint becomes $\max(f_t(x_t,\tilde{\delta}_t^j), f_s(x_s,\tilde{\delta}_s^j)) \le 0$, where t and s are the two periods of disruption indicated by $\tilde{I}^j$. Of course, we remove the "max" by enforcing $f_t(x_t,\tilde{\delta}_t^j)\le 0$ and $f_s(x_s,\tilde{\delta}_s^j)\le 0$. Theorem 1 of Calafiore and Campi applies directly in the two-disruption setting, i.e., by solving $\mathcal{RP}^M$, we obtain $\tilde{x}^M$ which satisfies the bound in (4).

The following result generalizes Theorem 2. The proof is along the same lines and is omitted.

Theorem 8. *Let $\mathbf{N}=(N_{t,s})$ be a vector of $|S|$ non-negative integers such that $\sum_{(t,s)\in S} N_{t,s} = M$. And, let $(\tilde{\delta}_t^1,\tilde{\delta}_s^1),\ldots,(\tilde{\delta}_t^{N_{t,s}},\tilde{\delta}_s^{N_{t,s}})$, i.i.d. from the joint distribution of $(\tilde{\delta}_t,\tilde{\delta}_s)$, $(t,s)\in S$, be used to define the convex program*

$$\begin{aligned}
\mathcal{RP}^{\mathbf{N}}: \qquad & \min_x \ \sum_{t=1}^{T} h_t(x_t) \\
& \text{s.t.} \ \sum_{t=1}^{T} g_t(x_t) \le b, \\
& \qquad \max(f_t(x_t,\tilde{\delta}_t^j), f_s(x_s,\tilde{\delta}_s^j)) \le 0,\ (t,s)\in S,\ j=1,\ldots,N_{t,s}, \\
& \qquad x_t \in D_t,\ t=1,\ldots,T.
\end{aligned}$$

Assume that $\mathcal{RP}^{\mathbf{N}}$ has a unique optimal solution $\tilde{x}^{\mathbf{N}}$, or that the two-norm tie-breaking rule is applied to obtain $\tilde{x}^{\mathbf{N}}$. Then,

$$\mathbb{E}\left[\mathbb{P}\big(f(\tilde{x}^{\mathbf{N}},\tilde{\xi})>0\big)\right] \le \sum_{(t,s)\in S} p_{t,s}\frac{\dim(x_t)+\dim(x_s)}{N_{t,s}+1}, \tag{16}$$

where the expectation is with respect to the augmented random vector whose $(t,s)^{\text{th}}$ component is $(\tilde{\delta}_{t,s}^1,\ldots,\tilde{\delta}_{t,s}^{N_{t,s}})$, $(t,s)\in S$.

5.2. Optimal Allocation

We carry out the analysis for optimal sampling strategies analogously to that in Section 3.2. We have $|S| = T(T-1)/2$ and the role of p in the single-disruption case is played here by the vector $p = (p_{t,s})_{(t,s)\in S}$. Hence, we consider a budget of $M = NT(T-1)/2$ for constraint sampling in the two-disruption setting. All decision vectors x_t are assumed to have dimension n. The optimal allocation and optimal expected probability of violation are

$$\begin{aligned} N^*_{t,s} &= \frac{(p_{t,s})^{1/2}}{\sum_{(t,s)\in S}(p_{t,s})^{1/2}}(N+1)\Big(\frac{T(T-1)}{2}\Big) - 1,\ (t,s)\in S, \\ V^*(N) &= \frac{2n\Big(\sum_{(t,s)\in S}(p_{t,s})^{1/2}\Big)^2}{(N+1)\Big(\frac{T(T-1)}{2}\Big)}. \end{aligned} \tag{17}$$

We observe that with $M = NT(T-1)/2$ and $\dim(x_t) = n$ for all decision vectors x_t, the bound (4) decays as $N^{-1}T^{-1}$. When p has a nested geometric distribution, i.e., the waiting time for the first disruption is geometric, and the waiting time for the second disruption given the time of the first is also geometric, it can be established that $V^*(N)$, given by (17), decays as $N^{-1}T^{-2}$. Thus, we see that there can be an improvement of an order of magnitude in T with stratified sampling. In the case of ℓ disruptions where ℓ is $\Theta(1)$ as $T\to\infty$, bound (4) decays as $N^{-1}T^{-(\ell-1)}$ and when p has a nested geometric distribution, $V^*(N)$ decays as $\ell N^{-1}T^{-\ell}$. Thus, the improvement is again of an order of magnitude in T. We state this below for the two-disruption case.

Theorem 9. *Assume* $\dim(x_t) = n,\ t = 1,\dots,T$. *Let* $V^1(N)$ *denote the bound (4) on the expected probability of violation with* $M = NT(T-1)/2$. *And, let* $V^*(N)$ *denote the bound (17). Then,*

$$(i)\quad \frac{V^*(N)}{V^1(N)} = \Theta\left(\frac{\Big(\sum_{(t,s)\in S} p_{t,s}^{1/2}\Big)^2}{T}\right) \quad \text{for large } T, \tag{18a}$$

$$(ii)\quad \frac{2N}{T(N+1)} + \Theta(1/T^3) \le \frac{V^*(N)}{V^1(N)} \le \frac{N(T-1)}{N+1} + \Theta(1/T). \tag{18b}$$

Furthermore, if $p_{t,s} = 2/T(T-1),\ (t,s)\in S$, *the upper bound in (18b) is tight; and if* p *is a nested geometric distribution, then* $V^*(N)/V^1(N) = \Theta(1/T)$ *as* $T\to\infty$.

Remark. The form of the requisite sample size in the ℓ-disruption case is $M = N\binom{T}{\ell}$ and hence the computational effort to solve the optimization model $\mathcal{RP}^{\mathbf{N}}$ in Theorem 8 grows exponentially in ℓ. As a practical matter this means we are computationally limited to a small number of disruptions, say, $\ell = 2$ or 3.

6. Conclusion

In this paper, we consider convex approximations of a multiperiod bicriteria minimization model with cost and risk as objectives. Towards constructing the efficient frontier for our model, we resort to constraint sampling to ensure that the risk is low. Our model incorporates random disruptions and is endowed with special structure that suggests a stratification strategy in sampling. We show that optimal stratification can provide improvements in sampling cost up to an order of magnitude in the number of time periods over a naive strategy when the number of disruptions is small. We also illustrate, using two examples from perishable inventory theory, that stratification produces better proximity to the efficient frontier on average. We attribute this to greater tightness of the bound on the expected probability of violation with stratification than naive sampling.

We pursue order-optimal stratification strategies, motivated by the fact that in practice, waiting time distributions are often unknown. Assuming that the unknown distribution is "well-behaved," we demonstrate that polynomially decaying allocations yield order-wise improvement identical to optimal allocation. To examine the case when this assumption is removed, we consider a worst-case setting that casts the model in a game-theoretic framework. Our significant result in this setting is that there is no advantage to the adversary in playing second. In other words, the problem is easily solvable using our earlier analyses, and by means of an intuitive "water-filling" algorithm.

Acknowledgements

This work has been supported by the National Science Foundation through grants CMMI-0653916 and CMMI-0800676 and the Defense Threat Reduction Agency through grant HDTRA1-08-1-0029. We thank two anonymous references for suggestions that improved the paper.

References

Bertsekas, D. P., A. Nedic, A. E. Ozdaglar. 2003. *Convex Analysis and Optimization*. Athena Scientific, Belmont.

Calafiore, G., M. C. Campi. 2005. Uncertain convex programs: Randomized solutions and confidence levels. *Math. Programming* **102** 25–46.

Calafiore, G., M. C. Campi. 2006. The scenario approach to robust control design. *IEEE Trans. on Automatic Control* **51** 742–753.

Ehrgott, M., X. Gandibleux. 2002. *Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys*. Kluwer Academic Publishers, Norwell.

Goyal, S. K., B. C. Giri. 2001. Recent trends in modeling of deteriorating inventory. *Eur. J. of Oper. Res.* **134** 1–16.

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

Luedtke, J., S. Ahmed, G. Nemhauser. 2010. An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming* **122** 247–272.

Markowitz, H. 1952. Portfolio selection. *J. of Finance* **7** 77–91.

Nahmias, S. 1982. Perishable inventory theory: A review. *Oper. Res.* **30** 680–708.

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

Nemirovski, A., A. Shapiro. 2006b. Scenario approximations of chance constraints. *Probabilistic and Randomized Methods for Design Under Uncertainty*. Springer, London, 3–48.

Pagnoncelli, B., S. Ahmed, A. Shapiro. 2008. Computational study of a chance constrained portfolio selection problem. `http://www.optimization-online.org`.

Pardalos, P. M., Y. Siskos, C. Zopounidis. 1995. *Advances in Multicriteria Analysis*. Kluwer Academic Publishers, Dordrecht.

Rengarajan, T., D. P. Morton. 2009. Estimating the efficient frontier of a probabilistic bicriteria model. *Proceedings of the Winter Simulation Conference*.

Ruszczyński, A., R. J. Vanderbei. 2003. Frontiers of stochastically nondominated portfolios. *Econometrica* **71** 1287–1297.

Salmerón, J., R. K. Wood, D. P. Morton. 2009. A stochastic program for optimizing military sealift subject to attack. *Military Oper. Res.* **14** 19–39.

Schultz, R., S. Tiedemann. 2003. Risk aversion via excess probabilities in stochastic programs with mixed-integer recourse. *SIAM J. on Optim.* **14** 115–138.

Vinod, B., J. J. Solberg. 1984. Performance models for unreliable flexible manufacturing systems. *Omega* **12** 299–308.

Yang, L., Y. Feng. 2007. A bicriteria solid transportation problem with fixed charge under stochastic environment. *Appl. Math. Model.* **31** 2668–2683.

Yu, G., X. Qi. 2004. *Disruption Management: Framework, Models and Applications*. World Scientific, Singapore.

Yu, P. L. 1985. *Multiple-Criteria Decision Making - Concepts, Techniques, and Extensions*. Plenum Press, New York.

Zeleny, M. 1982. *Multiple Criteria Decision Making*. McGraw-Hill, New York.

Online Supplement for:

Convex Approximations of a Probabilistic Bicriteria Model with Disruptions

Tara Rengarajan, Nedialko Dimitrov, David P. Morton

Graduate Program in Operations Research and Industrial Engineering, The University of Texas at Austin, Austin, TX 78712, USA, {trengan@yahoo.com, ned.dimitrov@gmail.com, morton@mail.utexas.edu}

Before turning to the proof of Theorem 2, we point to a technical lemma leading to inequality (4), that is proved in Calafiore and Campi (2005) and Calafiore and Campi (2006). Both proofs work by contradiction and are based on the same idea. In what follows, we construct a direct proof.

Proof of Technical Lemma

We begin by stating the classical Helly's theorem from convex analysis, which can, for example, be found in Bertsekas et al. (2003).

Theorem 10. (Helly) *Consider a finite collection of convex subsets of $\mathbb{R}^n$, and assume that the intersection of every subcollection of $n+1$ (or fewer) sets has a nonempty intersection. Then, the entire collection has a nonempty intersection.*

Consider the convex program

$$\begin{aligned} \mathcal{P}: \quad z^* = \min_{x \in D} \quad & h(x) \\ \text{s.t.} \quad & f_i(x) \leq 0,\ i = 1, \ldots, m, \end{aligned}$$

where $h(\cdot)$ and $f_i(\cdot)$, $i = 1, \ldots, m$, are convex functions on D, a closed, convex subset of $\mathbb{R}^n$. Define $\mathcal{P}_k$ as the problem obtained by dropping the k^{th} constraint in $\mathcal{P}$:

$$\begin{aligned} \mathcal{P}_k: \quad z_k^* = \min_{x \in D} \quad & h(x) \\ \text{s.t.} \quad & f_i(x) \leq 0,\ i = 1, \ldots, k-1, k+1, \ldots, m. \end{aligned}$$

When solving the above models, we adopt the convention that if there exist multiple optimal solutions, we choose the one with the smallest $\|\cdot\|_2$ norm. Since all sets and functions are convex, this two-norm tie-breaking rule extracts a unique optimal solution.

Definition 1. The k^{th} constraint of $\mathcal{P}$ is called a *support constraint* for $\mathcal{P}$ if the optimal solutions to $\mathcal{P}_k$ and $\mathcal{P}$ are different under the two-norm tie-breaking rule.

We use the following lemma to derive our improved bound on the expected probability of violation.

Lemma 11. *The number of support constraints for $\mathcal{P}$ is at most n.*

Proof. Let x^* be the optimal solution to $\mathcal{P}$, and x_k^* be the optimal solution to $\mathcal{P}_k$, under the two-norm tie-breaking rule. Let $S \neq \emptyset$ be the index set of support constraints. Define $\mathcal{X}' = \text{co}(\{x_k^*\}_{k \in S})$ and $\mathcal{X}_i = \{x \mid f_i(x) \leq 0\},\ i = 1, \ldots, m$. If the k^{th} constraint is a support constraint for $\mathcal{P}$, then we have $\mathcal{X}' \cap \bigcap_{\substack{i=1 \\ i \neq k}}^m \mathcal{X}_i \neq \emptyset$.

We claim that $\mathcal{X}' \cap \bigcap_{i=1}^m \mathcal{X}_i = \emptyset$. Suppose, by way of contradiction, that $w \in \mathcal{X}' \cap \bigcap_{i=1}^m \mathcal{X}_i$. Then, w is feasible for $\mathcal{P}$. Since dropping the k^{th} constraint gives a relaxation of $\mathcal{P}$, we have $z_k^* \leq z^*$ for all k in S. Let S' be a minimal subset of indices of S such that $w \in \text{co}(\{x_k^*\}_{k \in S'})$. If $z_k^* < z^*$ for some k in S', then, by the definition of S' and the convexity of h, it follows that $h(w) < z^*$, contradicting the optimality of x^*. Hence, $z_k^* = z^*$ for all k in S'. This implies $\|x_k^*\|_2 < \|x^*\|_2$ for all k in S' which in turn, by the convexity of $\| \cdot \|_2$, implies that $\|w\|_2 < \|x^*\|_2$, a contradiction to the tie-breaking rule. Thus, the claim is established.

Applying the contrapositive of Helly's theorem to the collection $\{\mathcal{X}', \mathcal{X}_1, \ldots, \mathcal{X}_m\}$, there exists a subcollection of at most $n+1$ sets that has empty intersection. Since problem $\mathcal{P}$ is feasible, i.e., $\bigcap_{i=1}^m \mathcal{X}_i$ is nonempty, the collection of sets with empty intersection includes $\mathcal{X}'$. Reindexing the sets $\mathcal{X}_i$ if necessary, let the collection of sets with empty intersection be given by $\{\mathcal{X}', \mathcal{X}_1, \ldots, \mathcal{X}_\ell\}$, where $\ell \leq n$. By the definitions of S and support constraint, the k^{th} constraint is a support constraint if and only if $\mathcal{X}' \cap \bigcap_{\substack{i=1 \\ i \neq k}}^m \mathcal{X}_i \neq \emptyset$. For $k \notin \{1, \ldots, \ell\}$, we have $\mathcal{X}' \cap \bigcap_{\substack{i=1 \\ i \neq k}}^m \mathcal{X}_i \subseteq \mathcal{X}' \cap \bigcap_{i=1}^\ell \mathcal{X}_i = \emptyset$, and hence for such k, the k^{th} constraint is not a support constraint. Therefore, the k^{th} constraint is a support constraint only if $k \in \{1, \ldots, \ell\}$, i.e., the number of support constraints is at most ℓ, which, in turn, is at most n. $\square$

Proof of Theorem 2

Proof. Let $\tilde{\delta}^{\mathbf{N}} = (\tilde{\delta}_1^1, \ldots, \tilde{\delta}_1^{N_1}, \ldots, \tilde{\delta}_T^1, \ldots, \tilde{\delta}_T^{N_T})$. The expected probability of violation is given by

$$\begin{aligned}
\mathbb{E}\left[\mathbb{P}\big(f(\tilde{x}^{\mathbf{N}}, \tilde{\xi}) > 0\big)\right] &= \mathbb{E}\left[\mathbb{E}_{\tilde{\xi}}\left[\mathbb{I}\big(f(\tilde{x}^{\mathbf{N}}, \tilde{\xi}) > 0\big) \mid \tilde{\delta}^{\mathbf{N}}\right]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} p_t \mathbb{E}_{\tilde{\xi}}\left[\mathbb{I}\big(f(\tilde{x}^{\mathbf{N}}, \tilde{\xi}) > 0\big) \mid \tilde{I}_t = 1,\, \tilde{\delta}^{\mathbf{N}}\right]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} p_t \mathbb{E}_{\tilde{\delta}_t}\left[\mathbb{I}\big(f_t(\tilde{x}_t^{\mathbf{N}}, \tilde{\delta}_t) > 0\big) \mid \tilde{\delta}^{\mathbf{N}}\right]\right] \\
&= \sum_{t=1}^{T} p_t \underbrace{\mathbb{E}\left[\mathbb{E}_{\tilde{\delta}_t}\left[\mathbb{I}\big(f_t(\tilde{x}_t^{\mathbf{N}}, \tilde{\delta}_t) > 0\big) \mid \tilde{\delta}^{\mathbf{N}}\right]\right]}_{V(\mathbf{N},t)},
\end{aligned} \tag{19}$$

where $\mathbb{I}(A)$ is an indicator random variable on the event A. Observe that the inner expectation in (19), $\mathbb{E}_{\tilde{\delta}_t}\left[\mathbb{I}(f_t(\tilde{x}_t^{\mathbf{N}}, \tilde{\delta}_t) > 0) \mid \tilde{\delta}^{\mathbf{N}}\right]$ is the conditional probability of violation, given that the disruption occurs in period t, and given $\tilde{\delta}^{\mathbf{N}}$ which determines $\tilde{x}_t^{\mathbf{N}}$. Defining $V(\mathbf{N}, t)$ as in (19), $V(\mathbf{N}, t)$ is then the conditional expected probability of violation given that the disruption occurs in period t.

To derive an upper bound for $V(\mathbf{N}, t)$, we fix the time period t and argue as follows: Let

$$\mathbf{N}^+ = (N_1, \ldots, N_{t-1}, N_t + 1, N_{t+1}, \ldots, N_T)$$

and define

$$\begin{aligned}
\mathcal{RP}^{\mathbf{N}^+}: \qquad \min_x \quad & \sum_{s=1}^{T} h_s(x_s) \\
\text{s.t.} \quad & \sum_{s=1}^{T} g_s(x_s) \le b, \\
& f_s(x_s, \tilde{\delta}_s^j) \le 0,\ s = 1, \ldots, T,\ j = 1, \ldots, \mathbf{N}_s^+, \\
& x_s \in D_s,\ s = 1, \ldots, T.
\end{aligned} \tag{20}$$

Problem $\mathcal{RP}^{\mathbf{N}^+}$ is identical to $\mathcal{RP}^{\mathbf{N}}$ with an additional constraint in period t, from an i.i.d. observation $\tilde{\delta}_t^{N_t+1}$ from the distribution of $\tilde{\delta}_t$. For $j \in \{1, \ldots, N_t + 1\}$, let $\mathcal{RP}_j^{\mathbf{N}^+}$ be identical to $\mathcal{RP}^{\mathbf{N}^+}$ except that the constraint associated with observation $\tilde{\delta}_t^j$ in (20) is dropped, and let $\tilde{x}_j^{\mathbf{N}^+}$ be the optimal solution to $\mathcal{RP}_j^{\mathbf{N}^+}$ under the two-norm tie-breaking rule. Let $\tilde{\delta}^{\mathbf{N}^+}(j)$ be the sample used to define $\mathcal{RP}_j^{\mathbf{N}^+}$ and $\tilde{\delta}^{\mathbf{N}^+}$ be that for $\mathcal{RP}^{\mathbf{N}^+}$, where we continue to suppress dependence on t. Then,

$$\begin{aligned} V(\mathbf{N},t) &= \mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}(N_t+1)}\left[\mathbb{E}_{\tilde{\delta}_t^{N_t+1}}\left[\mathbb{I}(f_t(\tilde{x}^{\mathbf{N}^+}_{N_t+1,t},\tilde{\delta}_t^{N_t+1})>0)\,|\,\tilde{\delta}^{\mathbf{N}^+}(N_t+1)\right]\right] \\ &= \mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}}\left[\mathbb{I}(f_t(\tilde{x}^{\mathbf{N}^+}_{N_t+1,t},\tilde{\delta}_t^{N_t+1})>0)\right], \end{aligned}$$

where $\tilde{x}^{\mathbf{N}^+}_{N_t+1,t}$ is the period t subvector of $\tilde{x}^{\mathbf{N}^+}_{N_t+1}$. Furthermore, for $j=1,\ldots,N_t+1$,

$$\begin{aligned} \mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}(j)}\left[\mathbb{E}_{\tilde{\delta}_t^{j}}\left[\mathbb{I}(f_t(\tilde{x}^{\mathbf{N}^+}_{j,t},\tilde{\delta}_t^{j})>0)\,|\,\tilde{\delta}^{\mathbf{N}^+}(j)\right]\right] &= \mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}}\left[\mathbb{I}(f_t(\tilde{x}^{\mathbf{N}^+}_{j,t},\tilde{\delta}_t^{j})>0)\right] \\ &= V(\mathbf{N},t), \end{aligned}$$

since $(\tilde{\delta}_t^1,\ldots,\tilde{\delta}_t^{N_t+1})$ are i.i.d. Hence,

$$\begin{aligned} V(\mathbf{N},t) &= \frac{1}{N_t+1}\sum_{j=1}^{N_t+1}\mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}}\left[\mathbb{I}(f_t(\tilde{x}^{\mathbf{N}^+}_{j,t},\tilde{\delta}_t^{j})>0)\right] \\ &= \frac{1}{N_t+1}\mathbb{E}_{\tilde{\delta}^{\mathbf{N}^+}}\Big[\underbrace{\sum_{j=1}^{N_t+1}\mathbb{I}(f_t(\tilde{x}^{\mathbf{N}^+}_{j,t},\tilde{\delta}_t^{j})>0)}_{\tilde{u}_t^{\mathbf{N}}}\Big]. \end{aligned} \tag{21}$$

Defining $\tilde{u}_t^{\mathbf{N}}$ as in (21), we see that $\tilde{u}_t^{\mathbf{N}}$ counts the number of observations in the sample $\tilde{\delta}^{\mathbf{N}^+}$, which are support constraints for the problem $\mathcal{RP}^{\mathbf{N}^+}$.

We now rewrite $\mathcal{RP}^{\mathbf{N}^+}$ in the form

$$\begin{aligned} \min_x \quad & h_t(x_t)+G^{\mathbf{N}^+}(x_t) \\ \text{s.t.} \quad & f_t(x_t,\tilde{\delta}_t^j)\le 0,\ j=1,\ldots,N_t+1, \\ & x_t\in H^{\mathbf{N}^+}, \end{aligned} \tag{22}$$

where $H^{\mathbf{N}^+}=\{x_t\in D_t\mid \exists\, x_s\in D_s, s=1,\ldots,T, s\neq t,\ \text{satisfying}\ \sum_{s=1}^T g_s(x_s)\le b\}$, and

$$\begin{aligned} G^{\mathbf{N}^+}(x_t) = \quad & \min_{\substack{x_s,s=1,\ldots,T\\ s\neq t}} \quad \sum_{\substack{s=1\\ s\neq t}}^{T} h_s(x_s) \\ & \text{s.t.} \qquad \sum_{\substack{s=1\\ s\neq t}}^{T} g_s(x_s)\le b-g_t(x_t), \\ & \qquad\qquad f_s(x_s,\tilde{\delta}_s^j)\le 0,\ s=1,\ldots,T,\ s\neq t,\ j=1,\ldots,N_s, \\ & \qquad\qquad x_s\in D_s,\ s=1,\ldots,T,\ s\neq t. \end{aligned}$$

Convexity of $f_s(\cdot, \tilde{\delta}_s^j), g_s(\cdot), h_s(\cdot)$ and $D_s,\ s = 1, \dots, T$, implies $H^{\mathbf{N}^+}$ is convex, and $G^{\mathbf{N}^+}(\cdot)$ is convex on $H^{\mathbf{N}^+}$. By Lemma 11, the number of support constraints for (22), or equivalently $\tilde{u}_t^{\mathbf{N}}$, is at most $\dim(x_t)$. Thus,

$$V(\mathbf{N}, t) \leq \frac{\dim(x_t)}{N_t + 1},$$

and using this in equation (19) yields

$$\mathbb{E}[\mathbb{P}(f(\tilde{x}^{\mathbf{N}}, \tilde{\xi}) > 0)] \leq \sum_{t=1}^{T} p_t \frac{\dim(x_t)}{N_t + 1},$$

as desired. □

Proof of Theorem 3

Proof. Substituting $N_t = p_t(N+1)T - 1$ into bound (5) yields

$$V^2(N) = \frac{1}{(N+1)T} \sum_{t=1}^{T} n_t, \tag{23}$$

and so $V^2(N) \leq V^1(N)$ is immediate. $V^*(N) \leq V^2(N)$ follows from the fact that $N_t = p_t(N+1)T - 1$ is a feasible solution to $\mathcal{SS}$ and $V^*(N)$ is its optimal value. Given $V^1(N) = \sum_{t=1}^{T} n_t/(NT+1)$ and $V^2(N)$ in equation (23), result (9a) follows, and equation (9b) is simply the ratio of (8) to (23). With $n_t = n,\ t = 1, \dots, T$, equation (9b) reduces to $(\sum_{t=1}^{T} p_t^{1/2})^2/T$. If we minimize $(\sum_{t=1}^{T} p_t^{1/2})^2/T$ subject to $\sum_{t=1}^{T} p_t = 1,\ p_t \geq 0,\ t = 1, \dots, T$, then the optimal p is a degenerate distribution, with all the probability mass in a single time period; maximizing $(\sum_{t=1}^{T} p_t^{1/2})^2/T$ subject to the same constraints yields the equal allocation solution $p = (1/T, \dots, 1/T)$. These two distributions achieve the lower and upper bounds, respectively, in (9c). □

Proof of Corollary 4

Proof. Under (10a) and $n_t = n,\ t = 1, \dots, T$,

$$\frac{V^*(N)}{V^2(N)} = \frac{\left((1-q)^{1/2} \sum_{t=1}^{T} q^{(t-1)/2} + q^{T/2}\right)^2}{T} = \Theta\left(\frac{1}{T}\right).$$

Under (10b) and $n_t = n,\ t = 1, \ldots, T$,

$$\frac{V^*(N)}{V^2(N)} = \frac{(1-q)\left(\sum_{t=1}^{T} q^{(t-1)/2}\right)^2}{T(1-q^T)} = \Theta\left(\frac{1}{T}\right).$$

□

Proof of Theorem 5

Proof. It suffices to establish that $u(\gamma) = \Theta(1)$ as $T \to \infty$. Suppose (i) holds. Then,

$$\begin{aligned} u(\gamma) &= \sum_{t=1}^{T} 1/t^{\gamma r} \sum_{t=1}^{T} 1/t^{(1-\gamma)r} / \sum_{t=1}^{T} 1/t^r \\ &= \Theta(1) \quad \text{as } T \to \infty, \end{aligned}$$

since $\gamma r > 1$, $(1-\gamma)r > 1$ and $r > 2$.

Next, suppose (ii) holds. Then,

$$\begin{aligned} u(\gamma) &= \sum_{t=1}^{T} q^{\gamma t} \sum_{t=1}^{T} q^{(1-\gamma)t} / \sum_{t=1}^{T} q^t \\ &= \Theta(1) \quad \text{as } T \to \infty, \end{aligned}$$

since $0 < q,\ \gamma < 1$. □

Proof of Theorem 6

Proof. As in Theorem 5, it suffices to show that $w(\alpha; p) = \Theta(1)$ as $T \to \infty$. From (14), there exists a finite constant c with $\sum_{t=1}^{T} t^k p_t < c$ for all T; we also have $\sum_{t=1}^{T} t^{-k} < \zeta(k) = \sum_{t=1}^{\infty} t^{-k}$, where $\zeta(k)$ is finite since $k > 1$. Thus,

$$1 \le (\sum_{t=1}^{T} p_t^{1/2})^2 \le w(\alpha; p) = \sum_{t=1}^{T} \frac{p_t}{\alpha_t} = \sum_{t=1}^{T} t^{-k} \sum_{t=1}^{T} t^k p_t < c\,\zeta(k) \ \ \forall T$$

establishing that $w(\alpha; p) = \Theta(1)$ as desired. □

Proof of Theorem 7

Proof. (a) The function $w(\cdot;p)$ is convex and closed for each $p \in \Delta \cap B$, and $w(\alpha;\cdot)$ is concave and closed for each $\alpha \in \Delta$. These facts in conjunction with the compactness of Δ and B imply (for example, see Bertsekas et al. (2003) - Prop. 2.6.4) that the minmax equality holds, i.e.,

$$\min_{\alpha\in\Delta}\Big(\max_{p\in\Delta\cap B} \sum_{t=1}^{T} \frac{p_t}{\alpha_t}\Big) = \max_{p\in\Delta\cap B}\Big(\min_{\alpha\in\Delta} \sum_{t=1}^{T} \frac{p_t}{\alpha_t}\Big).$$

From (13), the right-hand side of (15) follows.

(b) Strict concavity of $\sum_{t=1}^{T} p_t^{1/2}$ along with convexity and compactness of $\Delta \cap B$ imply that a unique optimal solution, p^*, exists for the problem on the right-hand side of (15). This shows that the adversary has a unique strategy when he plays first. Further, the inner minimization on the left-hand side of (15) is simply (13) with a unique optimal solution, α^*. Hence, the sampler has a unique optimal solution as the second player.

Consider the inner maximization over p in $\mathcal{AP}$. This is a simply a linear program subject to a convexity constraint and upper and lower bounds on the variables. Since the lower bounds are all nonzero, using translation, the objective of the linear program can be rewritten as $\sum_{t=1}^{T} p_t^l/\alpha_t + \sum_{t=1}^{T} p_t'/\alpha_t$, where $p_t' = p_t - p_t^l$. Thus, the outer minimization over α has an objective function that is the sum of a strictly convex function and a convex function, and is therefore strictly convex. Hence, there exists a unique optimal solution, say $\hat{\alpha}$, to the minimization problem in $\mathcal{AP}$, i.e., the sampler has a unique optimal strategy, $\hat{\alpha}$, as the first player.

It now follows that $(\hat{\alpha}, p^*)$ is a saddle point for $w(\cdot,\cdot)$. This implies that $\hat{\alpha} \in \operatorname{argmin}_{\alpha\in\Delta} w(\alpha, p^*)$. By uniqueness of the sampler's strategy as the second player, we have $\hat{\alpha} = \alpha^*$. The sampler's strategy is therefore identical, regardless of the order of play.

Now, suppose $\alpha^*_{t_1} = \alpha^*_{t_2}$ for some $t_1 \neq t_2$. Then, $p^*_{t_1} = p^*_{t_2}$, where p^* is again the unique optimal solution when the adversary plays first. Any vector $q \in \Delta \cap B$ such that $q_t = p^*_t, t \notin \{t_1, t_2\}$ and $q_{t_1} + q_{t_2} = p^*_{t_1} + p^*_{t_2}$, is also optimal for the adversary after the sampler has chosen α^*. Thus, in the event of a tie in the components of α^*, there may exist multiple optimal strategies for the adversary as the second player.

(c) We have

$$\begin{aligned} \max_{p \in \Delta \cap B} w(\alpha_{\text{unif}}, p) &= w(\alpha_{\text{unif}}, p_{\text{unif}}) \\ &\leq w(\alpha, p_{\text{unif}}), \end{aligned}$$

where the inequality follows from our optimal allocation results. If $p_{\text{unif}} \in \Delta \cap B$, then we also have $w(\alpha, p_{\text{unif}}) \leq \max_{p \in \Delta \cap B} w(\alpha, p)$. Hence,

$$\max_{p \in \Delta \cap B} w(\alpha_{\text{unif}}, p) \leq \max_{p \in \Delta \cap B} w(\alpha, p),$$

which establishes that α_{unif} is optimal for the sampler. Conversely, if α_{unif} is optimal for the sampler, then the uniqueness of the sampler's strategy, together with (a), implies that $p_{\text{unif}} \in \Delta \cap B$ and is optimal for the adversary. The second statement in (c) follows from the above inequality.

□

Proof of Theorem 9

Proof. We have

$$V^1(N) = \frac{nT}{NT(T-1)/2 + 1}. \tag{24}$$

from which, using (17), we have

$$\frac{V^*(N)}{V^1(N)} = \frac{2(NT(T-1)+2)\left(\sum_{(t,s) \in S} p_{t,s}^{1/2}\right)^2}{(N+1)T^2(T-1)}. \tag{25}$$

The equality in (18a) now follows, as does (18b). Tightness of the upper bound is obtained by setting p to the discrete uniform distribution. To prove the final result, let $0 < q, w < 1$ and

$$p_{t,s} = \frac{(1-q)q^{t-1}}{1-q^{T-1}} \cdot \frac{(1-w)w^{s-t-1}}{1-w^{T-t}},$$

i.e., the components of p are probability masses corresponding to a "nested" geometric distribution. Letting $u = 1 - w$ and $p = 1 - q$ for notational convenience, we have

$$\begin{aligned}\left(\sum_{(t,s)\in S} p_{t,s}^{1/2}\right)^2 &= \left(\sum_{(t,s)\in S} \frac{u^{1/2}p^{1/2}w^{(s-t-1)/2}q^{(t-1)/2}}{(1-w^{T-t})^{1/2}(1-q^{T-1})^{1/2}}\right)^2 \\ &= \frac{u\,p}{(1-q^{T-1})}\left(\sum_{(t,s)\in S} \frac{w^{(s-t-1)/2}q^{(t-1)/2}}{(1-w^{T-t})^{1/2}}\right)^2.\end{aligned}$$

After some simplification, we get

$$\left(\sum_{(t,s)\in S} p_{t,s}^{1/2}\right)^2 = \frac{(1-w)(1-q)}{(1-q^{T-1})(1-w^{1/2})^2}\left(\sum_{t=1}^{T-1} \frac{q^{(t-1)/2}(1-w^{(T-t)/2})}{(1-w^{T-t})^{1/2}}\right)^2. \tag{26}$$

The term on the right hand side of (26) is $\Theta(1)$ as $T \to \infty$, which completes the proof. $\square$

Detailed Model for Example 2

We describe the mathematical model of Example 2, explain the meaning of each construct, then relate the model to the paper's general recourse formulation (1)-(2), and finally give the parameter values that define the model instance used in the computations of Section 3.4. We specify the convex approximation form of the model, i.e., in the form of $\mathcal{RP}^{NT}$ or $\mathcal{RP}^{\mathbf{N}}$.

Sets and indices:

- $t = 1, \ldots, T$ indexes time periods.
- $w = 1, \ldots, W$ indexes facilities (warehouses).
- $c = 1, \ldots, C$ indexes customers.
- $d = 1, \ldots, D$ indexes sampled disruptions.
- t_d, time period of sampled disruption d.
- τ, number of time periods for graceful shut down of system after a disruption.
- $\mathcal{T}_d = \{t_d + 1, \ldots, \min\{t_d + \tau, T\}\}$, set of time periods in which the system operates after a disruption.

Parameters:

- $\mathrm{lowc}(t, w)$, per-unit cost of cheap production in period t in facility w.
- $\mathrm{highc}(t, w)$, per-unit cost of expensive production in period t in facility w.
- $\mathrm{holdc}(t, w)$, per-unit storage cost in period t in facility w.

- lowb(t, w), bound on the maximum number of cheap units that can be produced in period t in facility w.
- unmetc(t, c), penalty for unmet demand in period t for customer c.
- nomd(t, c), nominal demand in period t for customer c.
- transc(t, w, c), per-unit transportation cost for period t from facility w to customer c.
- budg(t), recourse budget for period t. After a disruption occurs, each period of operation must stay under this recourse budget for τ additional time periods.
- $\tilde{\delta}_c^d$, magnitude of the random disruption demand for customer c.
- p_t, probability mass function governing the time of disruption.
- $\bar{p}_t = \sum_{s=t}^{T} p_s$.

Decision variables:

- LPROD(t, w), units of cheap production at facility w in period t.
- HPROD(t, w), units of expensive production at facility w in period t.
- STO(t, w), units stored at facility w from period t to period $t + 1$.
- SHELF(t, w), units of placed on the shelf at facility w in period t.
- TRANS(t, w, c), units of transported from facility w to customer c in period t.
- UDEM(t, c), unmet demand of customer c in period t.

In addition to these variables, for each sampled disruption d, we have recourse variables that describe the change in operation of the system after the disruption occurs. For example, variables dSTO(d, t, w) denote units stored at facility w from period t to period $t + 1$ after disruption d occurs. The model includes similar decision variables dLPROD(d, t, w), dHPROD(d, t, w), dSHELF(d, t, w), dTRANS(d, t, w, c), and dUDEM(d, t, c).

Boundary Conditions:

- STO$(0, w) \equiv 0, \forall w$; there is no initial inventory of goods.
- STO$(T, w) \equiv 0, \forall w$; there is no final period inventory of goods.
- dSTO$(d, \min(T, t_d + \tau), w) \equiv 0, \forall d, w$; there is no final period inventory of good under recourse, τ periods after the disruption.

The formulation of the model follows. Its specialization to $\mathcal{RP}^{NT}$ (naive sampling) or $\mathcal{RP}^{\mathbf{N}}$ (stratified sampling) is carried out by how the disruptions $d = 1, \ldots, D$ are sampled.

$$
\begin{aligned}
\min \quad & \sum_{t=1}^{T} \overline{p}_t \Bigg(\sum_{w=1}^{W} \Bigg(\quad \mathrm{lowc}(t,w)\mathrm{LPROD}(t,w) + \mathrm{highc}(t,w)\mathrm{HPROD}(t,w) + && \text{(27a)} \\
& \qquad \mathrm{holdc}(t,w)\mathrm{STO}(t,w) + \sum_{c=1}^{C} \mathrm{transc}(t,w,c)\mathrm{TRANS}(t,w,c) \Bigg) + \\
& \quad \sum_{c=1}^{C} \mathrm{unmetc}(t,c)\mathrm{UDEM}(t,c) \Bigg) \\
\text{s.t.} \quad & \mathrm{LPROD}(t,w) \le \mathrm{lowb}(t,w) \quad \forall t, w, && \text{(27b)} \\
& \mathrm{dLPROD}(d,t,w) \le \mathrm{lowb}(t,w) \quad \forall d, t \in \mathcal{T}_d, w && \text{(27c)} \\
& \mathrm{STO}(t-1,w) + \mathrm{LPROD}(t,w) + \mathrm{HPROD}(t,w) = \mathrm{STO}(t,w) + \mathrm{SHELF}(t,w) \quad \forall t, w && \text{(27d)} \\
& \mathrm{dSTO}(d,t-1,w) + \mathrm{dLPROD}(d,t,w) + \mathrm{dHPROD}(d,t,w) = && \text{(27e)} \\
& \quad \mathrm{dSTO}(d,t,w) + \mathrm{dSHELF}(d,t,w) \quad \forall d, t \in \mathcal{T}_d, w \\
& \sum_{c=1}^{C} \mathrm{TRANS}(t,w,c) \le \mathrm{SHELF}(t,w) \quad \forall t, w && \text{(27f)} \\
& \sum_{c=1}^{C} \mathrm{dTRANS}(d,t,w,c) \le \mathrm{dSHELF}(d,t,w) \quad \forall d, t \in \mathcal{T}_d, w && \text{(27g)} \\
& \sum_{w=1}^{W} \mathrm{TRANS}(t,w,c) + \mathrm{UDEM}(t,c) \ge \mathrm{nomd}(t,c) \quad \forall t, c && \text{(27h)} \\
& \sum_{w=1}^{W} \mathrm{TRANS}(t_d,w,c) + \mathrm{UDEM}(t_d,c) \ge \mathrm{nomd}(t_d,c) + \tilde{\delta}_c^d \quad \forall d, c && \text{(27i)} \\
& \sum_{w=1}^{W} \mathrm{dTRANS}(d,t,w,c) + \mathrm{dUDEM}(d,t,c) \ge \mathrm{nomd}(t,c) \quad \forall d, t \in \mathcal{T}_d, c && \text{(27j)} \\
& \mathrm{dSTO}(d,t_d,w) = \mathrm{STO}(t_d,w) \quad \forall d, w && \text{(27k)} \\
& \Bigg(\sum_{w=1}^{W} \Bigg(\quad \mathrm{lowc}(t,w)\mathrm{dLPROD}(d,t,w) + \mathrm{highc}(t,w)\mathrm{dHPROD}(d,t,w) + && \text{(27l)} \\
& \qquad \mathrm{holdc}(t,w)\mathrm{dSTO}(d,t,w) + \sum_{c=1}^{C} \mathrm{transc}(t,w,c)\mathrm{dTRANS}(d,t,w,c) \Bigg) + \\
& \quad \sum_{c=1}^{C} \mathrm{unmetc}(t,c)\mathrm{dUDEM}(d,t,c) \Bigg) \le \mathrm{budg}(t) \quad \forall d, t \in \mathcal{T}_d \\
& \text{All decision variables are non-negative.} && \text{(27m)}
\end{aligned}
$$

The objective function in (27a) follows the general method described at the end of Section 2.1, and minimizes the expected cost of the non-recourse decisions, including costs for production, storage, shipping and unmet demand. Constraint (27b) ensures that a facility cannot produce more goods at a cheap rate than the upper bound, $\text{lowb}(t, w)$, and constraint (27c) does the same for the related recourse variables. Constraint (27d) ensures the conservation of goods through time at every facility. Constraint (27e) is the analog under recourse, and in this case, the constraint is expressed only for the τ time periods after the disruption occurs, to ensure a graceful shutdown of the system after the disruption. Constraints (27f) and (27g) ensure that goods can only be transported to customers if they have been put on the shelf. Constraint (27h) ensures that nominal demand at each customer is either met or accounted for as being unmet. We include similar constraints, (27i) and (27j), to capture the increase in demand during disruptions, and the requirement to meet nominal demand during recourse. Constraint (27k) links the first stage storage variables at the time of disruption to the recourse storage variables, which appear in constraint (27e). Finally, constraint (27l) ensures that during each of the periods of recourse, no more than the budget, $\text{budg}(t)$, can be used to operate the system.

With respect to our general model (1) with $f_t(x_t, \delta_t)$ defined in the recourse function (2), the decision variables x_t correspond to variables $\text{LPROD}(t, w)$, $\text{HPROD}(t, w)$, $\text{STO}(t, w)$, $\text{SHELF}(t, w)$, $\text{TRANS}(t, w, c)$, and $\text{UDEM}(t, c)$. And, the recourse variables variables y_t correspond to their counterparts after a disruption occurs: $\text{dLPROD}(d, t, w)$, $\text{dHPROD}(d, t, w)$, $\text{dSTO}(d, t, w)$, $\text{dSHELF}(d, t, w)$, $\text{dTRANS}(d, t, w, c)$, and $\text{dUDEM}(d, t, c)$. For the y_t variables, the index t on $\text{dLPROD}(d, t, w), \ldots, \text{dUDEM}(d, t, c)$ includes all $t \in \mathcal{T}_d$, and variable $\text{dSTO}(d, t, w)$ is also defined for $t = t_d$. The objective function in (27a) corresponds to $\sum_{t=1}^{T} \left(\sum_{s=t}^{T} p_s\right) h_t(x_t)$. The inventory constraints (27d) are of type (1b) and constraints (27b), (27f) and (27h) correspond to (1c). Constraints (27c), (27e), (27g), and (27j) are of type (2c). Constraints (27i) and (27k) are of type (2b). If $\tau = 1$ so that constraint (27l) is enforced for exactly one time period for each disruption then the cost function in (2a) is the left-hand side of constraint (27l) minus its right-hand side, $\text{budg}(t_d + 1)$, because we seek $f_t(x_t, \delta_t) \leq 0$. More generally, when constraint (27l) is enforced for multiple time periods, the objective function in (2a) is the maximum of all the values just described, over $t \in \mathcal{T}_d$.

The dimension parameter n_t is used in determining our sample sizes—see (11) and (12)—and is the number of x_t variables required to define $f_t(x_t, \delta_t)$. In the context of a recourse formulation this is the number of first stage variables that appear on the right-hand side of constraints of type (2b) in the recourse model. In the case of model (27), this involves the number of first stage variables that appear in constraints (27i) and (27k), $n_t = W \cdot C + W + C$. However, we can reduce this to $W + C$ by introducing a new first stage decision variable, TRANSUDEM(t,c), defined to be the left-hand side of constraint (27i) in a constraint of type (1c) and replacing the left-hand side of constraint (27i) with TRANSUDEM(t,c).

In Example 2, paralleling as closely as possible Example 1, we set $T = 20$, $W = 2$, $C = 2$, and D is determined by our choice of naive or stratified sampling, through inequalities (11) and (12), respectively. We set

$$\begin{aligned}
&\mathrm{lowc}(t,w) = 1.0 \cdot 0.9^{t-1}\\
&\mathrm{highc}(t,w) = 1.5 \cdot 0.9^{t-1}\\
&\mathrm{holdc}(t,w) = 3.0 \cdot 0.9^{t-1}\\
&\mathrm{lowb}(t,w) = 3.0\\
&\mathrm{unmetc}(t,c) = 1000 \cdot 0.9^{t-1}\\
&\mathrm{nomd}(t,c) = 0.5\\
&\mathrm{transc}(t,w,c) = \frac{w}{4.0}\\
&\mathrm{budg}(t) = 2.0 \cdot 0.95^{t-1}.
\end{aligned}$$

And, $\tau = 5$ is the number of additional periods the system operates after the disruption has occurred. The probability mass function governing the time of the disruption is given by (10b) with $q = 0.8$, and the magnitude of the disruption at each customer is independent and given by the same Pareto distribution assumed in Example 1. As described above, we can take $n_t = W + C = 4$.