An Optimal Path Model for the Risk-Averse Traveler

Leilei Zhang [*1] and Tito Homem-de-Mello [†2]

[1]Department of Industrial and Manufacturing Systems Engineering, Iowa State University

[2]School of Business, Universidad Adolfo Ibáñez, Santiago, Chile

February 13, 2015

Abstract

In this paper we discuss stochastic optimal path problems where the goal is to find a path that has minimal expected cost and at the same time is less risky (in terms of travel time) than a given benchmark path. The model is suitable for a risk-averse traveler, who prefers a path with a more guaranteed travel time to another path which could be faster but could also be slower. Such risk attitude is incorporated using the concept of second order stochastic dominance constraints. Recently developed theory for optimization problems with stochastic dominance constraints ensures that the resulting problem can be written as a large linear integer program with binary variables; for networks of realistic size, however, such a direct approach is not practical due to the size of the resulting optimization problem. Moreover, the solution returned by the model may contain cycles, which is clearly undesirable from a practical perspective.

A number of strategies are explored to solve the problem. First, we prove that cycles can be prevented by a simple modification of the model if the arc travel times are mutually independent. We then propose a sample average approximation (SAA) approach to the problem using samples from the distribution of travel times. Because of the randomness resulting from sampling, it is important that statistical guarantees for the solution returned by algorithm be given, and we provide heuristic procedures to deal with stochastic constraints. We also incorporate a branch-and-cut approach that exploits the structure of the problem in order to deal with the integrality constraints more efficiently.

We present some numerical experiments for a 1,522-arc system that corresponds to a large portion of the Chicago area network. The results show that our approach can solve the problem very effectively, producing solutions with statistical guarantees of optimality within reasonable computational time.

Keywords: Optimal path problem; stochastic dominance; sample average approximation; branch-and-cut

[*]leileizh@iastate.edu

[†]tito.hmello@uai.cl

1 Introduction

The optimal path problem with random travel times is fundamental in transportation networks. Travelers need to decide which route to take to go from point A to point B, knowing that the travel time on each link is subject to variability due to the presence of traffic, accidents, or other unforeseen reason. Consequently, the total time of each potential path is random as well. The choice of an appropriate route requires then a criterion to compare random quantities. In the classical formulation of the stochastic shortest path problem, the goal is to find the path with smallest *expected* total travel time. While such a criterion has the advantage of yielding efficient algorithms, it has a major drawback: by construction, it finds the path that is the best on the average, which means that if one considers multiple instances of the problem — i.e., multiple realizations of the travel times — the chosen path will tend to perform well on the average but, for a particular realization of the travel times, it can perform very well or very poorly. This is particularly troublesome in the situation where one is not making repeated decisions. Thus, there is need for the development of alternative routing models that go beyond looking at just the expected travel time on each route.

The shortcoming of optimization models with expected value are of course more general and do not occur only in transportation problems. Indeed, considerable effort has been made in recent years in the research community to develop models that incorporate ways of taking risk into account when making a decision. One such class of models is that of optimization problems with *stochastic dominance constraints*, first introduced by Dentcheva and Ruszczyński (2003). In such models, a constraint is imposed to ensure that the performance metric corresponding to a feasible solution stochastically dominates a given (random) benchmark. Stochastic dominance is a well-established tool to compare random outcomes (see, for example, Müller and Stoyan 2002). One form of such a relationship establishes that a random variable X stochastically dominates a random variable Y if and only if any risk-averse decision maker will prefer the outcomes of X to the outcomes of Y. Risk-aversion describes the situation in which the decision maker prefers the expected return of a random variable over the random return itself, which could be good or bad. Thus, when used as a constraint in an optimization model, stochastic dominance allows for incorporation of risk by restricting the feasible set to those solutions that are no riskier (in the above sense) than a given benchmark. Since its introduction by Dentcheva and Ruszczyński (2003), this type of models has gained attention in the literature, with applications in finance, energy, health care, homeland security and capital budget planning, among others; see, for instance, Karoui and Meziou (2006), Roman et al. (2006), Dentcheva and Ruszczyński (2006), Drapkin et al. (2011), Luedtke (2008), Hu et al. (2011b, 2014).

A few risk models have been developed in the specific context of optimal path problems in transportation. Examples include models based on the mean-variance rule (Loui 1983), expected utilities (Eiger et al. 1985, Murthy and Sarkar 1996, Sivakumar and Batta 1994, Sen et al. 2001), worst-case travel times (Yu and Yang 1998, Bell and Cassir 2002), robust optimization (Bertsimas and Sim 2003), prospect theory (Xiao and Lo 2013) and first-order stochastic dominance (Frank 1969, Mirchandani 1976, Miller-Hooks and Mahmassani 2003, Nie and Wu 2009). In Fan et al. (2005) and Nie and Fan (2006), the authors study models whereby the goal is to maximize the probability of on-time arrival. Jaillet et al. (2013) introduce the concept of a lateness index associated with the arrival times corresponding to a given routing policy. The lateness index

evaluates how well such arrival times meet pre-specified deadlines and involves the maximization of an expectation over a set of distributions — which is a well-known dual representation of coherent risk measures (see, e.g., Artzner et al. 1999). Less related to our work is the work of Ordóñez and Stier-Moses (2010) and Nikolova and Stier-Moses (2014), who discuss equilibrium properties of stochastic traffic assignment models under risk-aversion. Cominetti and Torrico (2013) provide a comprehensive study of the use of risk measures for modeling risk-averse route choice and traffic network equilibrium problems with random travel times. They find some key conditions that a risk measure must satisfy in order to fulfill a desirable property called *additive consistency*, which means that the risk preference between two random variables does not change by adding another independent variable to both variables (see Section 2.2 for further discussion on this issue).

Our work is closest to Nie et al. (2012b), who propose an optimization model with second-order stochastic dominance constraints to incorporate risk control into the choice of routes in a network. The model imposes the constraint that the total travel time on a feasible route must be no riskier than the travel time of a given benchmark route. In order to solve the problem, Nie et al. (2012b) propose a heuristic algorithm based on dynamic programming. While such an algorithm can handle well problems that are relatively small, it is not appropriate for larger problems as the number of possible scenarios may increase quickly with the total number of links in the network. For example, if the travel times on each arc are mutually independent then the number of scenarios grows exponentially with the total number of links.

In this paper, we study the same model as Nie et al. (2012b) but use a completely different strategy to solve it, which allows us to solve large instances of the problem. Although the theory of stochastic dominance-constrained problems allows to write the problem a binary (mixed-) integer linear program, the size of the resulting problems makes such a direct approach impractical in our context except for very small networks. Our approach is based on constructing a sequence of sampling-based approximations to the model. These approximations are rooted in the so-called sample average approximation (SAA) approach for stochastic optimization problems. SAA (Kleywegt et al. 2001) is a well-established technique to solve expected-value stochastic optimization problems; see Shapiro (2003) and Homem-de-Mello and Bayraksan (2014) for surveys of that area, and Verweij et al. (2003) and Kenyon and Morton (2003) for applications of this technique to stochastic routing problems in the risk-neutral case. More recently, the SAA was extended by Hu et al. (2011a) to solve optimization problems with stochastic dominance constraints. There are two aspects of such development: one is theoretical, i.e., showing asymptotic properties of the approximations as the sample size goes to infinity; the other is algorithmic, where statistical lower and upper bounds are derived for the optimal value of the problem based on a finite sample size. While the theoretical results of Hu et al. (2011a) apply in our context, the bounds do not, since the results in Hu et al. (2011a) assume a convex and compact feasible set — and in our case the feasible set is discrete as it corresponds to the set of feasible routes. Moreover, the results in Hu et al. (2011a) assume that the SAA problem (for a given sample size) can be solved but do not discuss how to solve it, as the objective of that paper is to present general theory; in our case, solving the approximating problem (which is a binary mixed integer program) requires special techniques due to the size of the problem.

The contributions of the present paper are the following: (i) We develop a new strategy to solve optimal path problems that incorporate risk control via stochastic dominance constraints, based on a sequence of sampling-based approximations with automatic increase of sample sizes. Because sampling is used, we can model large networks of realistic size; (ii) As each of the approximations is a mixed integer programming problem, we develop a specialized branch-and-cut algorithm to solve it; (iii) We propose new heuristic procedures to yield statistical lower and upper bounds for the optimal value of the problem. Bounds of this type are known for expected value problems (see, e.g., Bayraksan and Morton 2006) but are novel for discrete problems with stochastic dominance constraints; (iv) We present a new theoretical result showing that, when the travel time distributions are mutually independent, cycles can be prevented. Such cycles arise because the objective function in our problem include penalties for early arrival, but we show that such a problem may be avoided by adding one nonnegative continuous variable to the model; and (v) We demonstrate the effectiveness of our approach with numerical experiments for a real large-scale network with more than 1,500 arcs corresponding to part of the Chicago area.

The structure of the paper is as follows. In Section 2 we review concepts of stochastic dominance and present the optimal path model with stochastic dominance constraints. The issue of cycles that may appear in the solution of the optimization problem is dealt with in Section 3. In Section 4 we present the SAA formulation, a procedure to test feasibility of candidate solutions obtained from SAA, and a method to evaluate the solution quality via statistical lower and upper bounds. A strategy for variance reduction is also discussed. Section 5 presents a branch-and-cut algorithm to reduce computing time of the problem. Numerical results for the Chicago network are shown in Section 6. In Section 7 we present concluding remarks and discuss possible extensions of the model.

2 Basic Concepts and Model Formulation

2.1 Stochastic Dominance

Stochastic dominance is a concept widely used in many areas to compare random variables X and Y. Suppose momentarily that larger values are preferable to smaller ones (for example, X and Y represent revenue return). Assume that X and Y are integrable. Then, we say that X *dominates* Y *in first order* (denoted $X \succeq_1 Y$) if

$$F(X,\eta) \leq F(Y,\eta), \quad \forall \eta \in \mathbb{R},$$

where $F(X,\cdot)$ is the cumulative distribution function of X. We say that X *dominates* Y *in second order* (denoted $X \succeq_2 Y$) if

$$F_2(X,\eta) := \int_{-\infty}^{\eta} F(X,t)dt \leq \int_{-\infty}^{\eta} F(X,t)dt := F_2(Y,\eta) \quad \forall \eta \in \mathbb{R}.$$

It is well known that second-order dominance can also be written as $\mathbb{E}[(\eta - X)_+] \leq \mathbb{E}[(\eta - Y)_+] \quad \forall \eta \in \mathbb{R}$ (where $[\cdot]_+$ denotes the function $\max\{\cdot, 0\}$).

Stochastic dominance has a useful characterization in terms of risk. It is possible to show (see, e.g., Müller and Stoyan 2002) that $X \succeq_2 Y$ if and only if $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$

holds for all increasing concave utility functions $u(.)$ for which $\mathbb{E}[u(X)]$ and $\mathbb{E}[u(Y)]$ are finite. Since concave utility functions represent risk-averseness — in the sense that $u(\mathbb{E}[X]) \geq \mathbb{E}[u(X)]$, i.e., the certain value of $\mathbb{E}[X]$ is preferred to the expected utility $\mathbb{E}[u(X)]$ — it follows that $X \succeq_2 Y$ if and only if any risk-averse decision maker prefers the outcomes given by X to the outcomes given by Y. This notion is appropriate to model risk aversion when larger values are preferable to smaller ones (e.g. gains). When the opposite holds — as in our case, where the random variables correspond to travel times — then risk aversion corresponds to increasing convex functions. In that case X is preferable to Y (denoted $X \preceq Y$ in the sequel) if and only if $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ for all increasing convex u (this relationship is also called stochastic increasing convex ordering in the literature). It is easy to show that $X \preceq Y$ if and only if $-X \succeq_2 -Y$, which implies that we can characterize $X \preceq Y$ in the following way:

$$X \preceq Y \iff \mathbb{E}[(X-\eta)_+] \leq \mathbb{E}[(Y-\eta)_+] \quad \forall \eta \in \mathbb{R}. \tag{1}$$

Moreover, when the variable Y has a discrete distribution, say $y_i, i = 1, \ldots, m$, then we have that (see Dentcheva and Ruszczyński 2003)

$$X \preceq Y \iff \mathbb{E}\left[(X-y_i)_+\right] \leq \mathbb{E}\left[(Y-y_i)_+\right] \quad i = 1, \ldots, m. \tag{2}$$

2.2 Optimal Path Problem for Risk-averse Travelers

As mentioned earlier, there are multiple ways of incorporating the concept of risk-aversion into optimal path problems. Some of the existing approaches, however, have disadvantages. Cominetti and Torrico (2013) discuss this issue in detail. As shown in that paper, many risk measures fail to satisfy a property called *additive consistency*. In the context of routing, such a property postulates that, if a path with travel time X is less risky than a path with travel time Y, then the risk preference is not changed by adding an arc with travel time Z — with Z independent of X and Y — to both paths. While this is a very desirable property, Cominetti and Torrico (2013) show that many commonly used risk measures do not satisfy it; they also show that one class of risk measures that do satisfy the property are the so-called *entropic* risk measures of the form $\rho(X) = (1/\beta)\log \mathbb{E}[\exp(\beta X)]$ for $\beta \in \mathbb{R}$. Note however that this type of risk measures leads to highly nonlinear problems, although Jaillet et al. (2013) show that, in the context of their models, problems with entropic risk measures can be efficiently solved by a sequence of convex optimization problems.

Other ways to address risk aversion also have drawbacks. For example, utility-based models typically impose a specific utility function, which may not agree with the user's preferences (and may also fail to satisfy additive consistency). Models based on first-order stochastic dominance circumvent the issue of imposing a specific utility function but such a modeling approach in general leads to non-convex problems; see, for instance, Noyan et al. (2006). Of course, by exploiting the structure of the problem it is possible to develop better algorithms; for example, Nie and Wu (2009) develop a dynamic programming algorithm to solve optimal path problems with first-order dominance constraints.

Models based on second-order stochastic dominance address the above issues — they do not require the specification of a utility function, the resulting problem has nice properties such as convexity (see Müller and Stoyan 2002), and as Lemma 1 below shows, they do satisfy the additive consistency property.

Lemma 1. *Let U, V and W be random variables such that $U \preceq V$, and W is independent of U and V. Then, $U + W \preceq V + W$.*

Proof: The result follows from the characterization (1). The assumption that $U \preceq V$ can be written as

$$\mathbb{E}[(U - \eta)_+] \;\leq\; \mathbb{E}[(V - \eta)_+] \quad \text{for all } \eta \in \mathbb{R}.$$

Thus, given any $w \in \mathbb{R}$ and any $\eta \in \mathbb{R}$ we have $\mathbb{E}[(U + w - \eta)_+] \leq \mathbb{E}[(V + w - \eta)_+]$ and hence, since W is independent of both U and V, it follows that

$$\mathbb{E}[(U + W - \eta)_+ \,|\, W] \;\leq\; \mathbb{E}[(V + W - \eta)_+ \,|\, W] \quad w.p.1.$$

By taking expectations on both sides we obtain

$$\mathbb{E}[(U + W - \eta)_+] \;\leq\; \mathbb{E}[(V + W - \eta)_+] \quad \text{for all } \eta \in \mathbb{R},$$

i.e., $U + W \preceq V + W$. □

The Model

In addition to risk-aversion, the routing models we consider incorporate penalties for early and late arrivals at the destination. Models with such penalties are well-documented in the literature, as they represent well the situation where the traveller has a desired time of arrival — to work, or to some event — and does not want to be late or waste time by arriving too early; see, for instance, De Palma et al. (1990) for a discussion. Nikolova et al. (2006) assign nonlinear (e..g, quadratic) penalties for early and late arrivals, and discuss the complexity of the resulting problem. Beaudry et al. (2010) study the application of penalty-based models in the context of transportation of patients in a hospital. The penalization of early and late arrivals can also be viewed as a way to implement a measure of the reliability of the network from the traveller's perspective; using such an approach, Casello et al. (2009) show, via a large empirical study on a rapid transit service in Canada, that increasing the reliability of arrivals of passengers can decrease the costs of transit users.

In that context, our models represent the perspective of a risk-averse traveller who desires a reliable service. Following Nie et al. (2012b), we consider an optimal path problem between an origin node r and a destination node s in which the goal is to minimize path-based penalty costs — more specifically, earliness and tardiness with respect to a fixed target arrival time. The arc-based transportation costs of a chosen path are considered in the constraints, in the form of comparison of transportation costs with a benchmark path. A benchmark path is a predefined path from origin to destination. This path could be obtained from experience or obtained by optimal path problem without dominance constraints. Our goal is then to find a path with minimum expected total penalty cost which is no riskier than the given benchmark path. It is important to emphasize that in our setting the path is chosen *a priori* — i.e., we are not dealing with the situation of adaptive algorithms whereby the route can be modified after the trip has started.

To illustrate the ideas, we present an example of a very simple network where there are three paths linking the source to destination. Path A has total travel time $T_A \sim N(13, 2)$, whereas paths B and C have travel times $T_B \sim N(10, 3)$ and $T_C \sim N(14, 2)$, respectively, where $N(a, b)$ indicate a normal distribution with mean a and standard

deviation b. Suppose that the target arrival time is $\tau_0 = 10$, and that the penalties for early (late) arrival are equal to 1 for every unit of time below (above) τ_0. By using well-known formulas for the truncated normal distribution, we can calculate the exact values of the penalty costs. It follows that path B is the one with smallest expected total penalty cost (2.39 versus 3.12 for path A and 4.03 for path C). However, suppose we use path C as a benchmark, i.e., we impose the constraint that the chosen path be no riskier than path C. It is well known that, when $X \sim N(\mu_X, \sigma_X)$ and $Y \sim N(\mu_Y, \sigma_Y)$ we have that $X \preceq Y$ if and only if $\mu_X \leq \mu_Y$ and $\sigma_X \leq \sigma_Y$ (see, e.g., Müller and Stoyan 2002). Thus, we see that $T_A \preceq T_C$ but $T_B \not\preceq T_C$, so path A becomes the optimal choice.

The problem for a general network can be formulated as:

$$\begin{aligned} \min \ & \mathbb{E}\left[\beta\left(\tau_0 - Tx\right)_+ + \gamma\left(Tx - \tau_0\right)_+\right] \\ \text{s.t.} \ & \\ & \sum_{a \in \mathcal{I}(i)} x_a - \sum_{a \in \mathcal{O}(i)} x_a \ = \ d_i \quad \forall i \in \mathcal{N} \\ & Tx \ \preceq \ Ty \\ & x_a \ \in \ \{0,1\} \quad \forall a \in \mathcal{A}. \end{aligned} \tag{3}$$

In the above formulation, $\mathcal{N}$ and $\mathcal{A}$ denote respectively the set of nodes and set of arcs of the network; $T = (T_a)_{a \in \mathcal{A}}$ denotes a random (row) vector of known distribution corresponding to the travel times of each arc; β and γ are respectively the early and late arrival penalty with respect to the target arrival time τ_0; $x = (x_a)_{a \in \mathcal{A}}$ is the vector of decision variables, where each x_a is a binary decision variable indicating whether or not arc a is used in a path; $y = (y_a)_{a \in \mathcal{A}}$ is a parameter such that y_a has the same interpretation as x_a but y corresponds to a fixed benchmark path. As customary, $\mathcal{I}(i)$ and $\mathcal{O}(i)$ denote respectively the set of incoming and outgoing arcs from node i, and the parameters d_i are such that $d_r = -1$, $d_s = 1$ and $d_i = 0$ otherwise. Note that the model can be easily generalized to the case where there is a desired time window $[\underline{\tau}, \overline{\tau}]$ for the arrival time instead of a fixed target arrival time τ_0. This is useful, for example, in the context of delivery of packages. In that case the objective function becomes $\mathbb{E}\left[\beta\left(\overline{\tau} - Tx\right)_+ + \gamma\left(Tx - \underline{\tau}\right)_+\right]$. The algorithms discussed later remain the same for the more general model. Also, it is important to observe that the numerical values of β and γ are not so important — what matters is the ratio between those quantities, which defines the cost of arriving too early relatively to arriving too late.

It is easy to see that, by using characterization (1) of the $\preceq$ order and introducing additional variables to linearize the $[\cdot]_+$ functions, problem (3) can be rewritten as a stochastic binary linear program. However, two major difficulties remain, which preclude the implementation of a method to solve it [1]: first, the fact that the objective function and some of the constraints involve expectations with respect to the distribution of T as well as variables that depend on T. Since T represents the vector of travel times on each arc, it is clear that even if we assume that the travel time on each arc can take on only a few possibilities, the total number of outcomes of T will be very large except for small networks, so the size of the model will not be practical. Another issue is the fact that several constraints must hold for all $\eta \in \mathbb{R}$, i.e., we have an uncountable number of constraints. When T has finite support distribution we can use property (2)

[1] Indeed, this formulation appears in Nie et al. (2012b) but is not implemented, due to these difficulties.

to reduce it to finitely many inequalities, but again the number of such inequalities — which is the same as the number of outcomes of the vector T — will be prohibitively large. Fortunately, as we shall see in Section 4 these issues can be addressed through the use of a sampling approach.

3 Dealing with cycles

One issue that arises with model (3) — regardless of the stochastic dominance constraints — is the possibility that the optimal solution generated by the model contains *cycles*. This is easily seen by observing that the flow constraint $\sum_{a\in\mathcal{I}(i)} x_a - \sum_{a\in\mathcal{O}(i)} x_a = 0$ is satisfied for any cycle that does not contain the origin or destination nodes. Adding a cycle, of course, increases the total travel time of the path; however, since the objective function penalizes early arrivals, it is possible that an optimal solution uses cycles to make the arrival time closer to the target time τ_0.

The presence of cycles, while mathematically correct, is clearly undesirable from a practical perspective. No decision maker would agree with a solution that contains cycles that are disjoint from the actual path from origin to destination, or one in which the vehicle keeps running "around the block" to spend time. One method to eliminate cycles without enumerating all subtours is to add real-valued variables that keep track of the nodes visited prior and after each arc on a path (see Jaillet et al. 2013). Such an approach, however, requires adding $|\mathcal{N}| \times |\mathcal{A}|$ variables to the model. In our context, where we are concerned about reaching a target arrival time, it is more natural to simply allow the vehicle to start the trip later. In the model, this can be accomplished by introducing a *release time* variable z which denotes the time when the vehicle starts the trip. The variable z is added to the objective function with a negative coefficient to force it to be positive when needed; on the other hand, to prevent trips from starting and finishing arbitrarily late, the coefficient of z is taken to be smaller (in absolute value) than the penalty for tardiness. With the introduction of the release time variable and by choosing $0 < \kappa < \gamma$, problem (3) becomes

$$\begin{aligned}
\min\ & \mathbb{E}\left[\beta\left(\tau_0 - Tx - z\right)_+ + \gamma\left(z + Tx - \tau_0\right)_+\right] - \kappa z \\
\text{s.t.}\ & \\
& \sum_{a\in\mathcal{I}(i)} x_a - \sum_{a\in\mathcal{O}(i)} x_a = d_i \quad \forall i \in \mathcal{N} \\
& Tx \preceq Ty \\
& x_a \in \{0,1\} \quad \forall a \in \mathcal{A} \\
& z \geq 0
\end{aligned} \tag{4}$$

Note that the benchmark path y is a feasible solution to problem (4). Moreover, for any realization of T and any $z \geq 0$ we have

$$\begin{aligned}
\beta\left(\tau_0 - Tx - z\right)_+ + \gamma\left(z + Tx - \tau_0\right)_+ - \kappa z &\geq \gamma\left(z + Tx - \tau_0\right)_+ - \kappa z \\
&\geq \gamma\left(z + Tx - \tau_0\right) - \kappa z \\
&= (\gamma - \kappa) z - \gamma\tau_0.
\end{aligned} \tag{5}$$

Since we choose $\kappa < \gamma$, it follows that the objective function in (4) is bounded from below by an affine function of z with positive slope. Therefore, there exists an optimal solution to problem (4).

The idea of introducing a release time variable to avoid cycles appeared in Nie et al. (2012b), but as the authors acknowledge in that paper, it was proposed as a heuristic approach — in the authors' words, "A more careful study of a potential model that is guaranteed to avoid artificial delays falls out of the scope of the paper." Here we analyze the use of a release time variable in more detail. The first question that arises is, does the inclusion of a release time variable guarantee that cycles are prevented? In the case of deterministic models, the answer is a clear yes — indeed, suppose there is an optimal solution (x^0, z^0) which contains cycles, and let C be the set of arcs in the cycles. Note that by the choice of κ the optimal solution must satisfy $z^0 + Tx^0 = \tau_0$, since this makes the term inside the expectation equal to zero. Suppose all arcs have positive travel time. Then, we simply consider an alternative solution (x^*, z^*) such that $x^*_a = x^0_a$ for all $a \notin C$, $x^*_a = 0$ for all $a \in C$, and we set $z^* := \tau_0 - \sum_{a \notin C} T_a x^0_a$. Clearly, the arrival time with the new solution is equal to τ_0, so no penalties for earliness or tardiness incur. Moreover, since the arcs have positive travel time we have that $z^* > z^0$ so $-\kappa z^* < -\kappa z^0$. It follows that the solution (x^*, z^*) is strictly better than (x^0, z^0), which is a contradiction. Thus, the optimal solution does not contain cycles.

In the stochastic case, however, the situation is more delicate. To see that, consider for example the model in Figure 1, where we set node 1 as the origin and node 2 as the destination, with no stochastic dominance constraints. Clearly, the only two feasible solutions are $x^1 = (1, 0, 0)$ and $x^2 = (1, 1, 1)$.

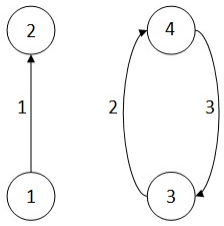


Figure 1: 3-ARC graph

The travel time distribution is given as follows:

Scenario	Probability	Arc 1	Arc 2	Arc 3
1	0.2	7	1	2.5
2	0.8	8	1	1

The objective function coefficients are $\beta = \gamma = 1$ and $\kappa = 0.01$, and the target arrival time is $\tau_0 = 10$. It is easy to see that without the release time variable z the optimal path is x^2 with objective value $0.2|10 - 10.5| + 0.8|10 - 10| = 0.1$. For the model with the release time variable z, the objective value for the solution x^1 is $0.2|10 - 7 - z| + 0.8|10 - 8 - z| - 0.01z$ which is minimized at $z = 2$ (with value 0.2-0.02=0.18), whereas the objective value for the solution x^2 is $0.2|10 - 10.5 - z| + 0.8|10 - 10 - z| - 0.01z$ which is minimized at $z = 0$ (with value 0.1-0=0.1). We see that even though we add the release time variable z the optimal route is still x^2, which contains a cycle.

The problem with the above example is that, since there are multiple realizations of the travel times, we cannot simply set z=slack time as in the deterministic case —indeed, in the example the slack time is 3 under one scenario and 2 under the other one. However, in the special case where the arc travel times are *mutually independent*,

the inclusion of the variable z does have the desired effect of preventing cycles. This is stated precisely in Theorem 1 below. The proof will use the following lemma:

Lemma 2. *Let U, V and W be random variables such that $U + W \preceq V$, $W \geq 0$ w.p.1 and U and W are independent. Then, $U \preceq V$.*

Proof: The result follows from the characterization (1), by noting that for any $w \geq 0$ and any $\eta \in \mathbb{R}$ we have that $(U-\eta)_+ \leq (U+w-\eta)_+$ w.p.1, thus $\mathbb{E}[(U-\eta)_+] \leq \mathbb{E}[(U+w-\eta)_+]$ and hence, since U and W are independent,

$$\mathbb{E}[(U-\eta)_+] \;\leq\; \mathbb{E}[(U+W-\eta)_+ \,|\, W] \quad w.p.1.$$

By taking expectations on both sides we obtain

$$\mathbb{E}[(U-\eta)_+] \;\leq\; \mathbb{E}[(U+W-\eta)_+] \;\leq\; \mathbb{E}[(V-\eta)_+],$$

where the second inequality corresponds to the assumption $U + W \preceq V$. □

Theorem 1. *Suppose the random variables $\{T_a : a \in \mathcal{A}\}$ are mutually independent and such that $P(T_a > 0) > 0$ for all a. Then, no optimal solution to (4) contains cycles.*

Proof: Suppose we solve problem (4) and obtain an optimal solution (x^0, z^0) such that x^0 contains cycles. Let the corresponding objective function value be v^0. Path x^0 can be separated into $x^0 = (x^B, x^C)$, where x^B and x^C denote respectively the simple path and all the cycle components. Accordingly, we write $T = (T^B, T^C)$.

We claim that we can construct a simple path solution which is strictly better than (x^0, z^0). To do that, note initially that we can write the optimal value v^0 as

$$\begin{aligned} v^0 =\; & \mathbb{E}\left[\beta\left(\tau_0 - Tx^0 - z^0\right)_+ + \gamma\left(z^0 + Tx^0 - \tau_0\right)_+ - \kappa z^0\right] \qquad (6)\\ =\; & \mathbb{E}_{T^C}\left[\mathbb{E}_{T^B}\left[\beta\left(\tau_0 - T^Bx^B - T^Cx^C - z^0\right)_+ + \gamma\left(z^0 + T^Bx^B + T^Cx^C - \tau_0\right)_+ - \kappa z^0 \,\middle|\, T^C\right]\right]. \end{aligned}$$

Next, define the quantity $z(T^C)$ as an optimal solution of the problem

$$\min_{z \geq 0}\; \mathbb{E}_{T^B}\left[\beta\left(\tau_0 - T^Bx^B - z\right)_+ + \gamma\left(z + T^Bx^B - \tau_0\right)_+ - \kappa z \,\middle|\, T^C\right]. \qquad (7)$$

Note that existence of $z(T^C)$ is ensured because for each realization of T^C the objective function in (7) is bounded from below by an affine function of z with positive slope, which is shown by using the same argument leading to (5). Optimality of $z(T^C)$ implies that, given any value of T^C, $z(T^C)$ is no worse than the particular solution $z^0 + T^Cx^C$, i.e., we have

$$\begin{aligned} &\mathbb{E}_{T^B}\left[\beta\left(\tau_0 - T^Bx^B - z(T^C)\right)_+ + \gamma\left(z(T^C) + T^Bx^B - \tau_0\right)_+ - \kappa z(T^C) \,\middle|\, T^C\right]\\ &\qquad \leq\; \mathbb{E}_{T^B}\left[\beta\left(\tau_0 - T^Bx^B - T^Cx^C - z^0\right)_+ + \gamma\left(z^0 + T^Cx^C + T^Bx^B - \tau_0\right)_+ - \kappa(z^0 + T^Cx^C) \,\middle|\, T^C\right] \qquad (8) \end{aligned}$$

with probability one. Now, the assumption of independence of $\{T_a : a \in \mathcal{A}\}$ implies that the conditional expectation in (7) is the same for all values of T^C. Let $z^B = z(T^C)$ denote a common optimal solution of (7).

Define $x^* := (x^B, 0)$ and $z^* := z^B$. We now show that (x^*, z^*) — which is a simple path solution by construction — is a better solution to (4) than (x^0, z^0). First we check that (x^*, z^*) is a feasible solution to that problem. It is clear that x^* satisfies the flow constraints since x^* coincides with x^0 except on cycles, which have net flow equal to zero. As for the dominance constraint $Tx \preceq Ty$, we have that $T^B x^B + T^C x^C \preceq Ty$ (since x^0 is feasible) and hence by Lemma 2 and the independence assumption of travel times we have that $T^B x^B \preceq Ty$, i.e., $Tx^* \preceq Ty$.

We now check that the objective value corresponding to (x^*, z^*) is smaller than the optimal value v^0. Indeed, it follows from (8) that

$$\begin{aligned}
&\mathbb{E}_{T^B}\left[\beta\left(\tau_0 - T^B x^B - z^B\right)_+ + \gamma\left(z^B + T^B x^B - \tau_0\right)_+ - \kappa z^B \,|\, T^C\right] \\
&\leq\ \mathbb{E}_{T^B}\left[\beta\left(\tau_0 - Tx^0 - z^0\right)_+ + \gamma\left(z^0 + Tx^0 - \tau_0\right)_+ - \kappa z^0 \,|\, T^C\right] - \kappa\mathbb{E}[T^C x^C \,|\, T^C] \quad \text{w.p.1,}
\end{aligned} \tag{9}$$

where we used the fact that $Tx^0 = T^B x^B + T^C x^C$. Since inequality (9) holds w.p.1, by applying the expectation with respect to T^C on both sides we obtain

$$\begin{aligned}
&\mathbb{E}\left[\beta\left(\tau_0 - Tx^* - z^*\right)_+ + \gamma\left(z^* + Tx^* - \tau_0\right)_+ - \kappa z^*\right] \\
&\leq\ \mathbb{E}\left[\beta\left(\tau_0 - Tx^0 - z^0\right)_+ + \gamma\left(z^0 + Tx^0 - \tau_0\right)_+ - \kappa z^0\right] - \kappa\mathbb{E}[T^C x^C] \\
&=\ v^0 - \kappa\mathbb{E}[T^C x^C],
\end{aligned}$$

where we used the identities $Tx^* = T^B x^B$ and $z^* = z^B$, as well as (6). Finally, the assumption that $P(T_a > 0) > 0$ for all a implies that $\mathbb{E}[T^C x^C] > 0$. It follows that (x^*, z^*) is a strictly better solution than (x^0, z^0), so we conclude that no optimal solution to (4) can contain cycles.

□

4 Sampling

4.1 Sample Average Approximation

As mentioned earlier, it is difficult to solve problem (4) in realistic settings due to the large number of scenarios and constraints. The Sample Average Approximation (SAA) approach provides with an efficient way to address that issue. The main idea of SAA is to replace the expectations in the problem with sample averages. Such an approach has sound theoretical basis, see for instance Shapiro (2003) and Homem-de-Mello and Bayraksan (2014) for overviews.

To use SAA to approximate problem (4), let $\{T^1, \ldots, T^N\}$ be a random sample of the vector of travel times T and define the following quantities:

$$F(T, x, z) \ :=\ \beta(\tau_0 - Tx - z)_+ + \gamma(z + Tx - \tau_0)_+ - \kappa z \tag{10}$$

$$\widehat{f}_N(x, z) \ :=\ \frac{1}{N}\sum_{j=1}^{N} F(T^j, x, z). \tag{11}$$

Note that $f(x, z) := \mathbb{E}[F(T, x, z)]$ is the objective function in (4), and $\widehat{f}_N(x, z)$ is a sample average approximation of that function. Next, let

$$G(T, x, \eta) \ :=\ (Tx - \eta)_+ - (Ty - \eta)_+ \tag{12}$$

and

$$\widehat{g}_N(x,\eta) \;:=\; \frac{1}{N}\sum_{j=1}^{N} G(T^j,x,\eta) \tag{13}$$

be a sample average approximation of the function $g(x,\eta) := \mathbb{E}[G(T,x,\eta)]$.

Following Hu et al. (2011a), the SAA formulation for optimization problem (4) is then written as follows:

$$\begin{aligned} & \min \widehat{f}_N(x,z) \\ \text{s.t.} \quad & \\ & \sum_{a\in\mathcal{I}(i)} x_a - \sum_{a\in\mathcal{O}(i)} x_a \;=\; d_i \quad \forall i\in\mathcal{N} \\ & \widehat{g}_N(x,T^\ell y) \;\le\; 0 \quad \ell = 1,\ldots,N \\ & x_a \;\in\; \{0,1\} \quad \forall a\in\mathcal{A} \\ & z \;\ge\; 0. \end{aligned} \tag{14}$$

A few remarks about problem (14) are in order. First, note that the stochastic dominance constraint $Tx \preceq Ty$ can be written (in light of (1)) as $g(x,\eta) \le 0$ for all $\eta \in \mathbb{R}$. This results in an uncountable number of constraints. However, by using property (2), Hu et al. (2011a) show that this set of constraints can be approximated by $\widehat{g}_N(x,T^\ell y) \le 0$ for all $\ell = 1,\ldots,N$. Note that sampling plays a double role in such an approximation: $\widehat{g}_N$ approximates the expectation in g, and the infinitely many constraints defined for each $\eta \in \mathbb{R}$ are replaced with finitely many constraints calculated at $\eta = T^\ell y$ for $\ell = 1,\ldots,N$. To understand the quality of such an approximation, given $\varepsilon \in \mathbb{R}$ define the sets

$$\begin{aligned} S^\varepsilon \;&:=\; \{x \,:\, g(x,\eta) \le \varepsilon \;\forall \eta\in\mathbb{R}\} \\ S_N \;&:=\; \{x \,:\, \widehat{g}_N(x,T^\ell) \le 0,\; \ell = 1,\ldots,N\}. \end{aligned}$$

Note that S^ε is a relaxation of the set $S := \{x \,:\, Tx \preceq Ty\}$ when $\varepsilon > 0$, and is a restriction of that set when $\varepsilon < 0$. Then, Hu et al. (2011a) show that there exist constants $M,\ \beta > 0$ such that for any $\varepsilon > 0$ we have

$$P\left(S^{-\varepsilon} \subseteq S_N \subseteq S^\varepsilon\right) \;\ge\; 1 - Me^{-\beta\epsilon^2 N}. \tag{15}$$

Essentially, (15) ensures that the probability that the feasibility set of the approximating problem is "sandwiched" between $S^{-\varepsilon}$ and S^ε goes to one exponentially fast with N. The rate of convergence, of course, depends on the value of the tolerance ε.

Another important remark about problem (14) is that although the problem is nonlinear, it can be easily linearized by the introduction of auxiliary variables e_+^j, e_-^j

and $s^{j\ell}$, $j, \ell = 1, \ldots, N$. The resulting linear problem is formulated as follows:

$$\min \frac{1}{N} \sum_{j=1}^{N} (e_+^j + e_-^j) - \kappa z$$

s.t.

$$\begin{aligned}
\sum_{a \in \mathcal{I}(i)} x_a - \sum_{a \in \mathcal{O}(i)} x_a &= d_i \quad \forall i \in \mathcal{N} \\
\frac{1}{N} \sum_{j=1}^{N} s^{j\ell} &\leq \frac{1}{N} \sum_{j=1}^{N} (T^j y - T^\ell y)_+ \quad \ell = 1, \ldots, N && \text{(16a)} \\
e_+^j &\geq \beta(\tau_0 - T^j x - z) \quad j = 1, \ldots, N && \text{(16b)} \\
e_-^j &\geq \gamma(z + T^j x - \tau_0) \quad j = 1, \ldots, N && \text{(16c)} \\
s^{j\ell} &\geq T^j x - T^\ell y \quad j, \ell = 1, \ldots, N && \text{(16d)} \\
x_a &\in \{0, 1\} \quad \forall a \in \mathcal{A} \\
z,\ e_+^j,\ e_-^j,\ s^{j\ell} &\geq 0 \quad j, \ell = 1, \ldots, N.
\end{aligned}$$

Note that since the benchmark y is fixed, the quantities $\frac{1}{N} \sum_{j=1}^{N} (T^j y - T^\ell y)_+$ can be computed in advance. The size of problem (16), of course, grows quickly with N, as it has order of N^2 variables and constraints. However, property (15) suggests we may not need too many samples to approximate the feasibility set. Moreover, classical results and numerical experiments for SAA for stochastic optimization problems also ensure that good approximations of the optimal values and optimal solutions of the original problem can be obtained without using too many samples; we refer again to Shapiro (2003) and Homem-de-Mello and Bayraksan (2014) for more detailed discussions. Finally, note that Theorem 1 also applies to problem (14) — indeed, the proof of that theorem is based on realizations of the travel times T, not on their distribution; therefore, the proof remains the same for the SAA version of problem (4).

4.2 Evaluating Solution Quality

By solving problem (14) we obtain an optimal solution corresponding to the particular sample $\{T^1, \ldots, T^N\}$ we have generated. However, due to the randomness of sampling there is no guarantee that such a solution is optimal for the original problem — while the theory ensures this is the case asymptotically (seeHu et al. 2011a), i.e., as N goes to infinity, there is no such guarantee for finite N. Therefore it is necessary to test the quality of the solution.

Tests for assessment of quality of solutions of stochastic optimization problems via sampling have been proposed in the literature; see, for instance, Bayraksan and Morton (2009) and references therein. Such tests typically consist of evaluating statistical lower and upper bounds for the optimal value of the problem — if the gap is sufficiently narrow, the candidate solution is considered to be a good one. The tests described in those works, however, assume that randomness is present only in the objective function. Such an assumption precludes the direct use of such tests in our context where randomness appears in the constraints as well. Wang and Ahmed (2008) proposes a modification of such tests for the case where the original problem has a finite number

of expected value constraints, and Hu et al. (2011a) extend it to the case of stochastic dominance constraints which, as we have seen, is equivalent to having infinitely many expected value constraints. The approach in Wang and Ahmed (2008) and Hu et al. (2011a), however, is based on computing a Lagrangian function, which guarantees that the optimality gap will be small for a good solution provided the problem is convex. Since our problem has integer variables, we cannot expect that the duality gap will be zero, so the Lagrangian-based test may yield a large gap even for a good solution.

To circumvent this issue, we propose an extension of the procedure from Mak et al. (1999) that accounts for the stochastic dominance constraints. The main issue with having stochastic constraints is that, since such constraints are approximated by sampling, there is a possibility that the solution obtained with the approximation is infeasible for the original problem.

For completeness, we briefly review the procedure in Mak et al. (1999). Consider problem (3) without the stochastic dominance constraints, and let S denote the feasibility set of such problem (note that S is deterministic). Consider the corresponding SAA problem with an arbitrary sample size N. Let v^* and v_N denote the respective optimal values of these problems, and let (x_N, z_N) be an optimal solution to the SAA problem. Then, given a solution $(\widehat{x}, \widehat{z})$ which is feasible for the problem being approximated, i.e., $(\widehat{x}, \widehat{z}) \in S$, we have

$$\begin{aligned} \mathbb{E}[v_N] \;=\; \mathbb{E}\left[\min_{(x,z)\in S} \widehat{f}_N(x,z)\right] \;&\leq\; \min_{(x,z)\in S} \mathbb{E}\left[\widehat{f}_N(x,z)\right] \;=\; \min_{(x,z)\in S} f(x,z) \;=\; v^* \\ &\leq\; f(\widehat{x},\widehat{z}) \;=\; \mathbb{E}\left[\widehat{f}_N(\widehat{x},\widehat{z})\right], \end{aligned} \tag{17}$$

so we see that the optimality gap $f(\widehat{x}, \widehat{z}) - v^*$ can be bounded as follows,

$$\begin{aligned} f(\widehat{x},\widehat{z}) - v^* \;&\leq\; \mathbb{E}\left[\widehat{f}_N(\widehat{x},\widehat{z})\right] - \mathbb{E}[v_N] \;=\; \mathbb{E}\left[\widehat{f}_N(\widehat{x},\widehat{z})\right] - \mathbb{E}\left[\widehat{f}_N(x_N,z_N)\right] \\ &=\; \mathbb{E}\left[\widehat{f}_N(\widehat{x},\widehat{z}) - \widehat{f}_N(x_N,z_N)\right]. \end{aligned} \tag{18}$$

Let $\Psi(\widehat{x}, \widehat{z})$ denote the quantity on the right-hand side. To estimate $\Psi(\widehat{x}, \widehat{z})$, the basic idea is to solve the SAA problem M times, generating a new sample for each problem independently of the previous ones. Let $\widehat{f}_N^k$ denote the function used in the kth replication and let (x_N^k, z_N^k) denote an optimal solution in the kth replication. Then, $\Psi(\widehat{x}, \widehat{z})$ can be estimated by

$$\Psi_M(\widehat{x},\widehat{z}) \;:=\; \frac{1}{M}\sum_{k=1}^{M} \widehat{f}_N^k(\widehat{x},\widehat{z}) - \widehat{f}_N^k(x_N^k, z_N^k). \tag{19}$$

Note that each term on the right-hand side of (19) is nonnegative. Since the replications are independent, by taking M sufficiently large (say, $M = 30$) we have by the Central Limit Theorem that the distribution of $\Psi_M(\widehat{x}, \widehat{z})$ is approximately Normal. A $100(1-\alpha)\%$ confidence interval for $\Psi(\widehat{x}, \widehat{z})$ can be computed as

$$\Psi_M(\widehat{x},\widehat{z}) \pm \frac{\nu_{\alpha/2} s_M}{\sqrt{M}}, \tag{20}$$

where s_M denotes the sample standard deviation calculated from the terms in (19) and ν_α is the inverse value of the cumulative standard normal distribution function at $1-\alpha$.

If the right extreme of the interval in (20) is small, the solution $(\widehat{x}, \widehat{z})$ is considered good.

In the procedure described above there are two important assumptions: (i) the set S must be deterministic for the inequality $\mathbb{E}\left[\min_{(x,z)\in S} \widehat{f}_N(x,z)\right] \leq \min_{(x,z)\in S} \mathbb{E}\left[\widehat{f}_N(x,z)\right]$ to be applicable, and (ii) the candidate solution $(\widehat{x}, \widehat{z})$ must be feasible. When the stochastic dominance constraint is included is S, the SAA problem (14) has a random feasibility set $\widehat{S}_N$ which approximates S. Moreover, we need to ensure that the candidate solution $(\widehat{x}, \widehat{z})$ is feasible for both the original problem as well as the SAA problem, in order to ensure that the gap estimate in (18) is nonnegative.

To circumvent these obstacles, we propose the following heuristic approach. To deal with issue (i) above, we observe that in light of (15) we expect that the sets S_N and S be close even for moderate values of N. Recent stability results for problems with stochastic dominance constraints developed in Dentcheva and Römisch (2013), combined with classical stability results for variations in the objective function (see, e.g., Römisch 2003) ensure that when S_N and S are close we have

$$\min_{(x,z)\in S} \widehat{f}_N(x,z) \;\approx\; \min_{(x,z)\in S_N} \widehat{f}_N(x,z)$$

and then we can write

$$\begin{aligned} \mathbb{E}[v_N] \;=\; \mathbb{E}\left[\min_{(x,z)\in S_N} \widehat{f}_N(x,z)\right] \;&\approx\; \mathbb{E}\left[\min_{(x,z)\in S} \widehat{f}_N(x,z)\right] \\ &\leq\; \min_{(x,z)\in S} \mathbb{E}\left[\widehat{f}_N(x,z)\right] \;\leq\; f(\widehat{x},\widehat{z}) \end{aligned} \tag{21}$$

provided that $(\widehat{x}, \widehat{z}) \in S$. In addition, in order to use the multiple-replication gap estimator (19), we need to ensure that $(\widehat{x}, \widehat{z}) \in S_N^k$ for all k, where S_N^k is the feasibility set for the problem in the kth replication.

We proceed now with a more detailed description of our feasibility and optimality tests. Given a candidate solution $(\widehat{x}, \widehat{z})$, the first task is to check the feasibility of that solution for the original problem. Assuming that the deterministic constraints in (4) are satisfied, checking feasibility of $(\widehat{x}, \widehat{z})$ entails checking whether $T\widehat{x} \preceq Ty$. Note initially that a *necessary* condition for $T\widehat{x} \preceq Ty$ to hold is that $\mathbb{E}[T\widehat{x}] \leq \mathbb{E}[Ty]$, so a simple check is to verify whether

$$\mathbb{E}[T]\widehat{x} > \mathbb{E}[T]y. \tag{22}$$

If (22) holds, we know that $\widehat{x}$ is not feasible, so we need to generate a new candidate solution. Otherwise, we proceed with a second test, described next.

Observe that, as seen earlier, the condition $T\widehat{x} \preceq Ty$ is equivalent to $\mathbb{E}[G(T,\widehat{x},\eta)] \leq 0$ for all $\eta \in \mathbb{R}$, where $G(\cdot)$ is defined in (12). That is, we want to check whether

$$\sup_{\eta\in\mathbb{R}} \mathbb{E}[G(T,\widehat{x},\eta)] \;\leq\; 0. \tag{23}$$

The difficulty with such test, of course, is that it requires testing infinitely many inequalities. One way around that issue is to note that for any fixed x and any N we have

$$\sup_{\eta\in\mathbb{R}} \mathbb{E}[G(T,x,\eta)] \;\leq\; \mathbb{E}[\widetilde{g}_N(x)], \tag{24}$$

where

$$\widetilde{g}_N(x) \;:=\; \sup_{\eta\in\mathbb{R}} \widehat{g}_N(x,\eta), \tag{25}$$

and $\widehat{g}_N$ is defined in (13). Inequality (24) follows the same principle as (17), but with η as a variable and the inequalities reversed since we maximize over η. It follows that the inequality

$$\mathbb{E}\left[\widetilde{g}_N(\widehat{x})\right] \;\leq\; 0 \tag{26}$$

provides a sufficient condition for (23) to hold.

Our approach is to view (26) as a one-sided hypothesis test of the form $H_0 : \mu \leq 0$ against $H_1 : \mu > 0$, where $\mu = \mathbb{E}\left[\widetilde{g}_N(\widehat{x})\right]$. To test that hypothesis, we compute M independent replications of $\widetilde{g}_N(\widehat{x})$ and use

$$\mu_M \;:=\; \frac{1}{M}\sum_{k=1}^{M} \widetilde{g}_N^k(\widehat{x}) \tag{27}$$

as an estimator of μ, where $\widetilde{g}_N^k$ is the value from the kth replication. Note that each replication in principle requires finding the supremum in (25); however, as shown earlier, it suffices to check finitely many values of η given by $T_i^k y$, $i = 1, \ldots, N$.

Once μ_M in (27) is computed, we proceed as in a standard hypothesis test. First, the *p-value* corresponding to such a test is calculated. When M is sufficiently large, the p-value can be approximated by the quantity

$$1 - \Phi\left(\frac{\mu_M}{\sqrt{s_M^2/M}}\right), \tag{28}$$

where as customary the function Φ denotes the cumulative standard normal distribution function, and s_M is the sample standard deviation of $\{\widetilde{g}_N^k\}$, $k = 1, \ldots, M$. If the p-value is smaller than a pre-specified amount (say, α_1), then we reject H_0, in which case we cannot say whether $\widehat{x}$ is feasible. In that case, we discard $\widehat{x}$. Otherwise, it is necessary to check the possibility of type II error. Let α_2 be a pre-specified upper bound on the probability of a type II error. We can then compute a tolerance δ such that the probability a type II error is ensured to be smaller than α_2 whenever $\mu > \delta$. Such a value is given by (see, e.g., Homem-de-Mello et al. (2011))

$$\delta \;:=\; (\nu_{\alpha_1} + \nu_{\alpha_2})\,\frac{s_M}{\sqrt{M}}. \tag{29}$$

Ideally, of course, we would like δ to be small, so that the interval of values of μ for which there is no guarantee on the probability of a type II error (i.e., $0 < \mu \leq \delta$) is small. In order to make δ unit-free, we will divide the above quantity by $\mathbb{E}[T]y$, since for any x that is feasible for the original problem and any η we have

$$-\mathbb{E}[T]y \;\leq\; \mathbb{E}[(Tx-\eta)_+ - (Ty-\eta)_+] \;\leq\; \mathbb{E}[(Tx-\eta)_+] \;\leq\; \mathbb{E}[(Ty-\eta)_+] \;\leq\; \mathbb{E}[T]y.$$

To summarize, the general procedure to get candidate solution is listed below:

Procedure GenCand

1. Generate N_1 samples of the vector T for SAA problem (16) , solve it and get candidate solution $(\widehat{x}, \widehat{z})$. Repeat the procedure M_1 times and choose the solution with highest frequency (if there is a tie, choose the solution with lower average objective value). The purpose of repeating the procedure M_1 times is just to obtain a "stable" solution, since different samples may lead to different solutions. We choose as a candidate solution the one with highest frequency.
2. Given a sample size N_2, generate M_2 replications of $\widetilde{g}_{N_2}(\widehat{x})$. Let μ_{M_2} and s_{M_2} be respectively the mean and sample standard deviation of $\{\widetilde{g}_N^k(\widehat{x})\}$, $k = 1, \ldots, M_2$.
3. Calculate the p-value corresponding to the hypothesis test $H_0 : \mathbb{E}\left[\widetilde{g}_N(\widehat{x})\right] \leq 0$, which is given by (28);
4. If the p-value is less than α_1 then reject H_0, discard $\widehat{x}$, increase the sample size N_1 and go back to STEP 1 to generate a new candidate solution.
5. If the p-value is no less than α_1, calculate tolerance $\delta' := \delta/(\mathbb{E}[T]y)$, where δ is computed as in (29). If δ' is bigger than some pre-specified tolerance, say 10^{-3}, then increase N_2 and go back to STEP 2 to repeat the hypothesis test.
6. If both the p-value and the tolerance δ are within acceptable ranges, then we retain H_0, declare $\widehat{x}$ feasible for the original problem and keep that solution for later evaluation.

The output of procedure **GenCand** is a solution $(\widehat{x}, \widehat{z})$ which is considered feasible for the original problem. As mentioned before, in order to use the interval estimate given in (19) and (20) we need to verify that $\widehat{x}$ is feasible for each replication. Due to the randomness of samples in each replication, it is possible that the candidate solution $(\widehat{x}, \widehat{z})$ is infeasible to the SAA problem in some replications. The procedure below summarizes the calculations. In the procedure, an optimality tolerance error $\theta > 0$ is introduced to determine whether the confidence interval in (20) is too big—this occurs when the right extreme of the interval is bigger than a fraction θ of the average optimal value of the replications.

Procedure GapEst

1. Input: a candidate solution $(\widehat{x}, \widehat{z})$ obtained from procedure **GenCand**.
2. For $m = 1, \ldots, M_3$:
   (a) Generate N_3 samples of the vector T and solve SAA problem (16) to get optimal solution $(x_{N_3}^m, z_{N_3}^m)$ and objective value $v_{N_3}^m$.
   (b) Test feasibility of $\widehat{x}$ for this SAA problem:
      i. If $\widehat{x}$ is infeasible, then go back to STEP 2a.
      ii. Otherwise, get objective value $\widehat{v}_{N_3}^m$, and let $G_{N_3}^m := \widehat{v}_{N_3}^m - v_{N_3}^m$.
3. Calculate the mean and variance of $G_{N_3}^m, m = 1, \ldots, M_3$:

$$\overline{G}_{M_3} := \frac{1}{M_3} \sum_{m=1}^{M_3} G_{N_3}^m$$

and

$$S_G^2 = \frac{1}{M_3 - 1} \sum_{m=1}^{M_3} (G_{N_3}^m - \overline{G}_{N_3})^2.$$

Also, compute

$$\overline{v}_{N_3} := \frac{1}{M_3} \sum_{m=1}^{M_3} \widehat{v}_{N_3}^m.$$

4. Calculate $\epsilon_G = \frac{\nu_\alpha S_G}{\sqrt{M_3}}$, and obtain the one-sided confidence interval

$$[\max(0, \overline{G}_{M_3} - \epsilon_G),\ \overline{G}_{M_3} + \epsilon_G].$$

5. If $(\overline{G}_{M_3} + \epsilon_G) > \theta\, \overline{v}_{N_3}$, then evaluate $(\overline{G}_{M_3} - \epsilon_G)$:
   (a) If $(\overline{G}_{M_3} - \epsilon_G) > \theta\, \overline{v}_{N_3}$ then the confidence interval for the optimality gap does not cover zero, that is, the candidate solution $(\widehat{x}, \widehat{z})$ is not optimal. In that case, we increase the sample size N_1, run the procedure **GenCand** and go back to STEP 1.
   (b) Otherwise, the confidence interval for the optimality gap is too large. In that case, we need to increase N_3, and go back to STEP 2 to reevaluate candidate solution again.
6. If $(\overline{G}_{M_3} + \epsilon_G) \leq \theta\, \overline{v}_{N_3}$, then the confidence interval for the optimality gap is small enough and covers zero. STOP and output the confidence interval $[0, \overline{G}_{M_3} + \epsilon_G]$.

5 Branch and Cut

In Section 4 we saw how to apply an SAA approach to solve problem (4). As seen there, such an approach involves solving the SAA problem (16) — which is a mixed integer program — multiple times, which is clearly computationally intensive. Thus, to reduce the computational burden it is crucial to use strategies that can speed-up the optimization algorithm. For that purpose, we describe now a branch-and-cut algorithm to reduce the number of nodes visited and the number of constraints in computation.

The branch-and-cut algorithm requires solving the linear relaxation problem below.

Note that $\mathcal{J}$ and $\mathcal{L}$ are index sets that change as the algorithm progresses.

$$\begin{aligned}
&\min \frac{1}{N}\sum_{j\in 1,\cdots,N}(e_+^j + e_-^j) - \kappa z \\
&\text{s.t.} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (30)\\
&\sum_{a\in\mathcal{I}(i)} x_a - \sum_{a\in\mathcal{O}(i)} x_a \;=\; d_i \quad \forall i\in\mathcal{N}\\
&\frac{1}{N}\sum_{j=1}^{N} s^{j\ell} \;\le\; \frac{1}{N}\sum_{j=1}^{N}(T^j y - T^\ell y)_+ \quad \ell\in\mathcal{L}\\
&e_+^j \;\ge\; \beta(\tau_0 - T^j x - z) \quad j\in\mathcal{J}\\
&e_-^j \;\ge\; \gamma(z + T^j x - \tau_0) \quad j\in\mathcal{J}\\
&s^{j\ell} \;\ge\; T^j x - T^\ell y \quad j=1,\ldots,N,\ \ell\in\mathcal{L}\\
&x_a \;\in\; [0,1] \quad \forall a\in\mathcal{A}\\
&z,\ s^{j\ell} \;\ge\; 0 \quad j=1,\ldots,N, \ell\in\mathcal{L}.\\
&e_+^j,\ e_-^j \;\ge\; 0 \quad j\in 1,\cdots N
\end{aligned}$$

The branching of a node consists in splitting a binary variable x_a into $x_a = 0$ and $x_a = 1$. If the solution is integer, say x^k for (30), we could proceed to test if this solution is feasible for the SAA problem (16). If yes, then we have a candidate solution for optimality. Otherwise, more cuts need to be added to problem (30).

The test of feasibility for problem (16) consists of checking whether or not the stochastic dominance constraint (16a) is satisfied. This is accomplished by solving the problem

$$\max_{\ell\notin\mathcal{L}} \frac{1}{N}\sum_{j=1}^{N}(T^j x^k - T^\ell y)_+ - \frac{1}{N}\sum_{j=1}^{N}(T^j y - T^\ell y)_+. \tag{31}$$

If the optimal value of the above problem is less than or equal to zero, then x^k is feasible for problem (16). Otherwise, let ℓ_0 be the maximizer in (31). We then set $\mathcal{L} := \mathcal{L}\cup\{\ell_0\}$ and solve problem (30) again.

Now consider the linearization constraints (16b) and (16c). To check whether such constraints are satisfied for a candidate solution (x^k, z^k) (with x^k integer), we compute

$$\max_{j\notin\mathcal{J}} \tau_0 - T^j x^k - z^k \tag{32}$$

$$\min_{j\notin\mathcal{J}} \tau_0 - T^j x^k - z^k. \tag{33}$$

If both (32) and (33) have optimal value equal to zero, then (x^k, z^k) is optimal for problem (16). Otherwise, let j_1 and j_2 be respectively the optimizer in (32) and (33). We then set $\mathcal{J} := \mathcal{J}\cup\{j_1,\ j_2\}$ and solve problem (30) again.

We summarize the algorithm below.

5.1 Algorithm Display

1. Set $v^{\text{best}} := \infty$, Tree:= {the first node need to be branched}, and $\mathcal{J} := \mathcal{L} := \emptyset$.

2. For each *active node* in Tree:
   (a) Solve problem (30), call (x^0, z^0) its optimal solution and v^0 its optimal value.
   (b) If the problem is infeasible or $v^0 > v^{\text{best}}$, then prune the node from Tree and go to next active node;
   (c) If x^0 is integral, then go to STEP 2d to test linearization and feasibility constraints; Otherwise, branch current node to create two new active nodes, and go to next active node in Tree;
   (d) If the feasibility constraints are satisfied and $v^0 \leq v^{\text{best}}$, then update best solution found so far, i.e., set $x^{\text{best}} := x^0$, $z^{\text{best}} := z^0$, and $v^{\text{best}} := v^0$.
   (e) If the linearization constraints are satisfied, then prune the node from Tree and go to next active node;
   (f) If some feasibility or linearization constraints are violated, then update $\mathcal{J}$ and $\mathcal{L}$ to add feasibility cuts and/or linearization cuts to the problem (30) until such constraints are satisfied, and repeat the test in STEP 2b.
3. Output $x^{\text{best}}, z^{\text{best}}$ and v^{best}.

6 Numerical Experiments

We now present numerical results to illustrate the application of the model and algorithm described in the previous sections. The data, which were obtained from Nie et al. (2012a), correspond to part of the transportation network in the city of Chicago, containing 1522 arcs and 611 nodes. The area covered in the data stretches from downtown to the southwest part of the city. In our experiment, we set the intersection of Roosevelt Ave. and Western Ave. as the origin node (identified as Node S in all Figures), and the intersection of Chicago Ave. and Michigan Ave. as the destination node (Node T in all figures). In the data set, travel times on each arc have discrete uniform distribution between given minimum and maximum values; the support of each distribution has 100 points. Travel times on different arcs are assumed to be independent. We emphasize that such an assumption is made only because there were not enough data to accurately estimate correlations; for the algorithm proposed in this paper, independence is immaterial so long as one can sample from the marginal distributions while preserving the correlations (see, e.g., Biller and Ghosh 2006 for a discussion on generation of multivariate random samples). Note that, under the independence assumption, the total number of outcomes of the vector of travel times is $1522^{100} \approx 10^{318}$, which is an astronomical number; here we see that sampling is crucial to solve the problem.

The number of possible paths between the source and destination in a network of this size is of course very large; Figure 2 displays a handful of those, for illustration purposes. The path BUS in Figure 2 indicates the bus route in Chicago's transit system. Two routes among the paths are suggested by Google Maps — the path that uses mostly expressways and the path with the least number of turns (those are indicated by EWY and DIR in Figure 2, respectively). The path indicated by PEWY1 travels partly on the expressway and partly on some arterial roads. The paths indicated by ALT1 and ALT2 are variations of a path that relies mostly on Ogden Ave. (the diagonal road seen on the map) to connect to downtown. The expected travel times for the different paths are shown in Table 1.

Paths	ALT1	ALT2	BUS	DIR	EWY	PEWY1
Exp travel time	804.95	804.96	1080.39	891.87	979.55	825.70

Table 1: Expected travel times for different paths (in seconds)

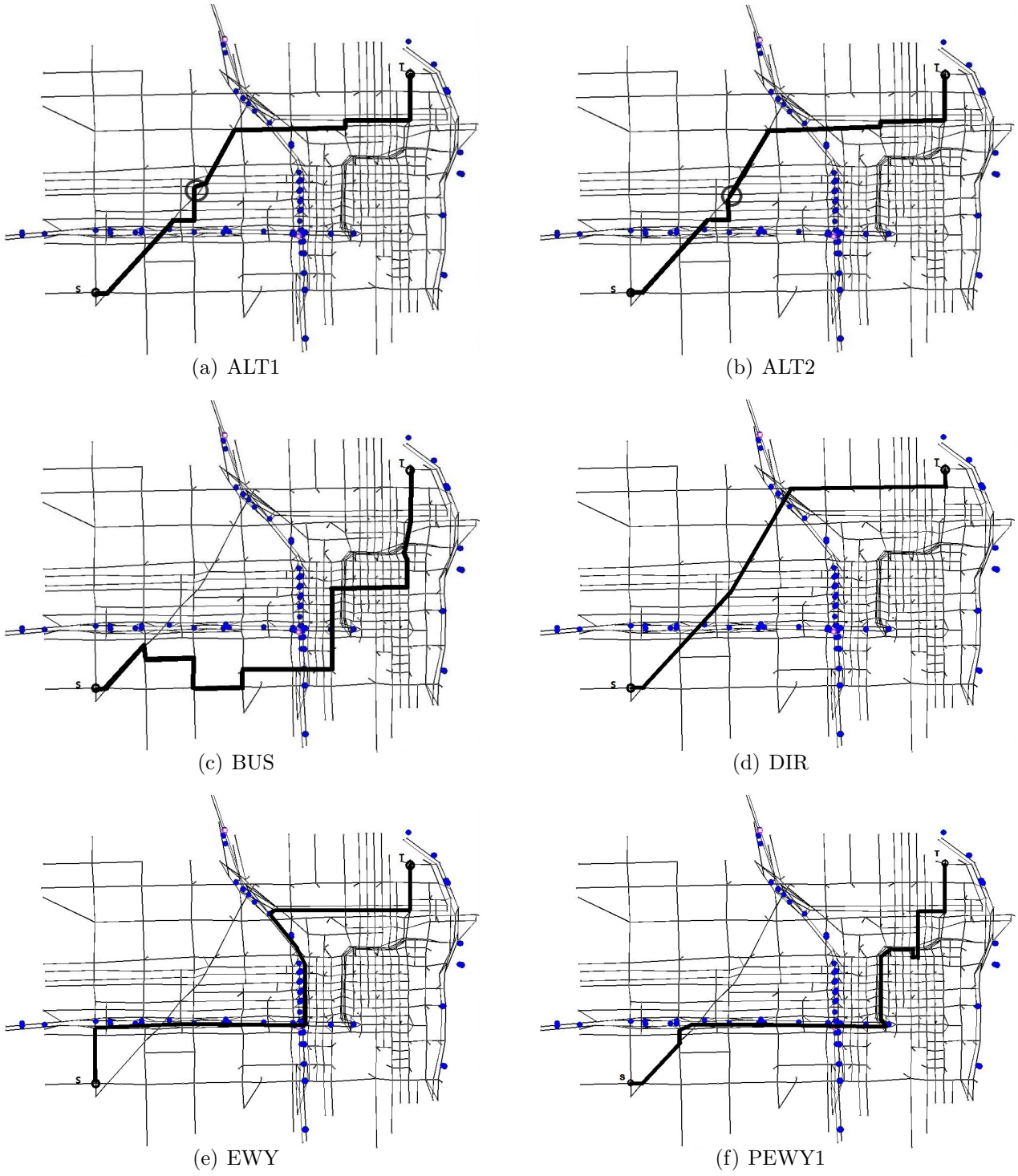
As discussed earlier, the goal of our model is to find an optimal path that is no riskier than the existing benchmark path given earliness and lateness penalties β, γ and the scheduled arrival time τ_0. We experimented with a few values for these constants, the results are reported below. In all experiments, we set the constant κ — which is the penalty for the release time variable z — equal to $\gamma/2$. Also, we chose different benchmarks to see how the optimal solution changes according to the benchmark path.

Note that the choice of the benchmark path requires a trade-off between having more feasible paths to choose from and protecting against the risk of long travel times. For example, choosing the path with least expected travel time as the benchmark will probably lead to a small feasible set (possibly even a singleton), so the resulting solution may not necessarily yield a low value for the objective function, which measures the penalties for deviating from the target arrival time τ_0. On the other hand, choosing a path with large expected travel time (and high variability) as the benchmark imposes a very weak risk requirement on the chosen solutions, i.a., almost any valid path between source and destination will satisfy the dominance constraint. The idea of the benchmark is to ensure that the path that is ultimately chosen is less risky than some standard path commonly used by the decision-maker. Some alternative ways of choosing the benchmark are (i) to use the path that is recommended by typical route-finding software, or (ii) to solve the problem without the dominance constraints — that is an easier problem — and then use use the resulting optimal path as a benchmark for the risk-averse problem.

For the scheduled arrival time τ_0, three cases are analyzed: (i) a small value $\tau_0 = 0$, which essentially means that the goal is to reach the destination as soon as possible; (ii) a large value that essentially ensures that no arrival will be late (the exact value for such a choice is immaterial, any very large number would be valid); and (iii) a moderate value which is set as the expected travel time of the benchmark path or other feasible path. The goal of the choice in (iii) is to forbid paths for which the total travel time is either too short or too long.

We now discuss the choice of the parameters M_i and N_i, $i = 1, 2, 3$ which are inputs for the procedures **GenCand** and **GapEst** described in Section 4.2. The value of M_1 is used just to generate a reasonable candidate solution, so we choose $M_1 = 10$. The value of M_2 should be large enough to allow the normality resulting from the central limit theorem to hold, so we take $M_2 = 30$. Finally, the value of M_3 for gap estimation is suggested to be 20-30 in Bayraksan and Morton (2009), so we use $M_3 = 20$. The initial sample sizes N_1, N_2 and N_3 are set to be 100; whenever the algorithm dictates that sample size N_i be increased, this is done in increments of 25.

We finish this preliminary discussion by mentioning that we used Latin Hypercube Sampling (LHS) for all sampling within a replication (samples are independent across replications). LHS is a well known induced-correlation technique introduced by McKay et al. (1979) to reduce variance, see also Homem-de-Mello and Bayraksan (2014) for a discussion of the use of that method in stochastic optimization.



(a) ALT1 (b) ALT2

(c) BUS (d) DIR

(e) EWY (f) PEWY1

Figure 2: Feasible paths (the dots represent loop detectors on the highways).

6.1 Results

6.1.1 Testing the multi-replication procedure

In our first experiment, we set the bus route as the benchmark — we purposely wanted to have a loose benchmark to avoid infeasibility issues, as the goal was to check the

quality of the optimality test. We also set $\beta = \gamma = 1$, and chose the moderate value $\tau_0 = 803$ as the target value. The tolerance parameters δ and θ were set respectively as 10^{-3} and 1%.

After running procedure **GenCand**, we obtained the solution indicated in Figure 2(a), which was then picked as the candidate path $\widehat{x}$ and submitted to procedure **GapEst**. For $N_1 = 100, N_2 = 100, N_3 = 100$, the procedure yielded [0.19, 0.57] as a 95% confidence interval for the optimality gap. Since the average optimal value over the M_3 replications was $\overline{v}_{N_3} = 34.59$, we see that the right extreme of the confidence interval is bigger than 1% of $\overline{v}_{N_3}$, but the left extreme is smaller than that amount. Therefore, it was necessary to increase sample size N_3 and repeat the procedure to reevaluate the candidate solution. After a few similar iterations, the confidence interval obtained with $N_3 = 225$ was $[0.18, 0.33]$, with $\overline{v}_{N_3} = 35.01$. Now the right extreme of the interval is smaller than 1% of the average optimal value over the M_3 replications, so we could say that candidate solution $(\widehat{x}, \widehat{z})$ is optimal for the original problem with 95% confidence.

For the purpose of illustration, Table 2 displays the optimality gap obtained in each of the M_3 replications with $N_3 = 225$. Note that $\widehat{x}$ was feasible in all of the replications.

$(\widehat{x}, \widehat{z})$	$\widehat{v}^m_{N_3}$	$(x^m_{N_3}, z^m_{N_3})$	$v^m_{N_3}$	$G^m_{N_3}$
	35.5036	(ALT2, 34.3393)	34.9743	0.5293
	37.7516	(ALT1, 37.4121)	37.3418	0.4098
	34.9301	(ALT2, 37.4723)	34.1863	0.7438
	36.9160	(ALT1, 34.2822)	36.7311	0.1850
	34.3311	(ALT1, 32.5483)	34.2887	0.0424
	34.2690	(ALT1, 30.8609)	34.2524	0.0165
	34.9839	(ALT1, 36.5507)	34.4966	0.4874
	34.2120	(ALT1, 34.3770)	33.9482	0.2638
	35.0448	(ALT1, 24.0720)	34.8437	0.2011
(ALT1, 29.661)	32.9433	(ALT1, 31.2097)	32.9224	0.0209
	36.1879	(ALT1, 35.0193)	36.0306	0.1573
	32.0926	(ALT1, 31.3528)	32.0518	0.0408
	33.8994	(ALT1, 38.2757)	33.5679	0.3316
	34.2178	(ALT1, 30.2999)	34.2074	0.0105
	35.6715	(ALT1, 37.0268)	35.3312	0.3403
	34.6549	(ALT1, 38.1137)	34.2986	0.3563
	37.0944	(ALT1, 39.8465)	36.6736	0.4208
	37.0181	(ALT1, 35.6200)	36.8150	0.2031
	32.3820	(ALT2, 35.2804)	32.1751	0.2070
	36.0973	(ALT2, 32.9037)	35.9206	0.1767

Table 2: 20 replications for evaluating candidate solution with $N_1 = 100, N_2 = 100, N_3 = 225$.

6.1.2 Testing the effectiveness of the release time variable

Our second experiment aimed at testing the effectiveness of including the release time variable z to avoid cycles. For that purpose, we again set the bus route as the benchmark route. We set $\beta = 3$, $\gamma = 1$, and $\tau_0 = 1400$, thus forcing late arrivals. For comparison purposes, we also ran the model without incorporating the variable z. The optimal solution in this case is the path marked in Figure 3. Note that the path contains five cycles of two arcs each. Since the cycles are really small, we could only mark the rough area of the cycles. Clearly, such a solution is not desirable.

In contrast, when the release time variable z is included in the model, we obtain — for the same parameters as above — the path ALT1 as the optimal path, with a release time $z = 656.6$.

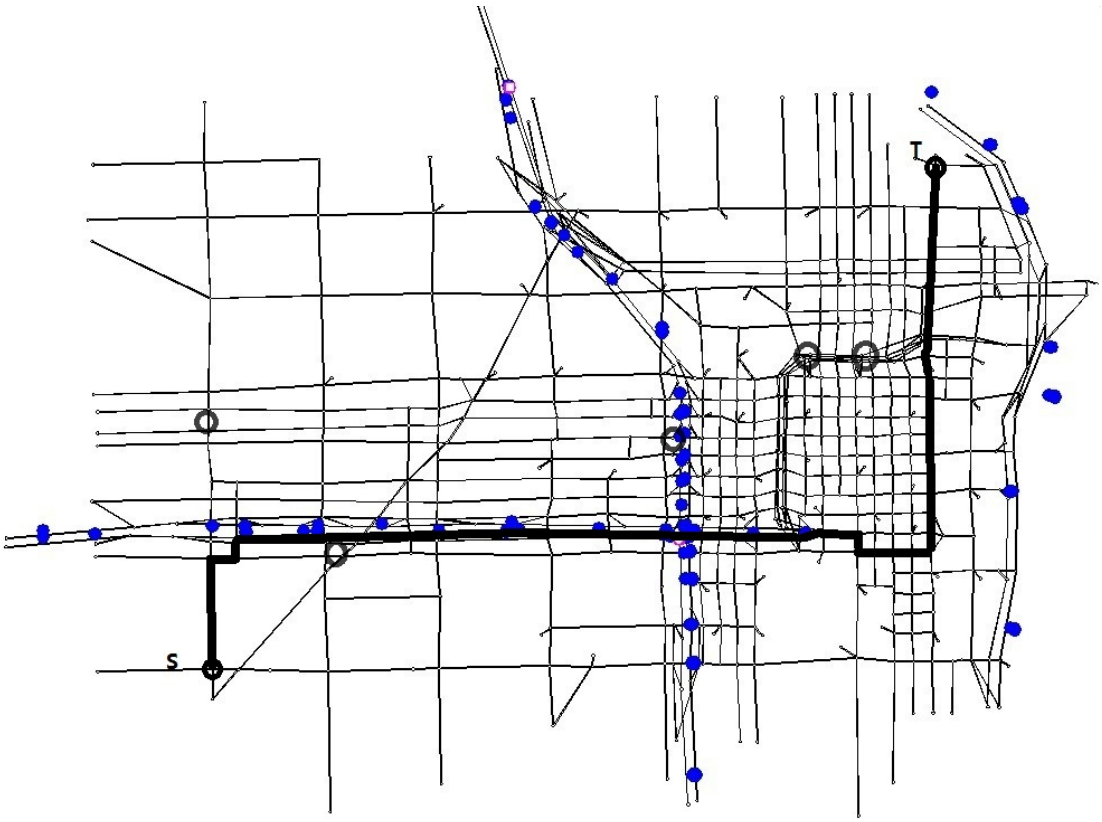


Figure 3: New path with cycle.

6.1.3 Testing different parameter values

In our third experiment, we chose different parameter combinations to test the sensitivity of the optimal path. The results are shown in Tables 3 and 4. In the tables, the column *Benchmark* indicates the benchmark path, chosen among the BUS, EWY, DIR, ALT1 and ALT2 paths from Figure 2 . The columns *Opt x* and *Opt z* show respectively the optimal path and optimal release time obtained with each configuration. The column *Obj. Value* displays the objective value for each problem (note that such value is not the total expected cost for earliness and lateness as it includes the

penalty for release time). *CI* is the one-sided 95% confidence interval for optimality gap. *Node* presents the average number of nodes branched in each replication. Finally, the column N_1, N_2, N_3 shows the final sample sizes used in the procedure (recall that sample sizes are increased in increments of 25).

$\beta = 1$, $\gamma = 1$, $\tau_0 = 0$						
Benchmark	Opt x	Opt z	Obj.Value	CI	Node	N_1, N_2, N_3
BUS	ALT1	0	804.9449	[0, 0]	1.6667	100, 100, 100
EWY	ALT1	0	804.9449	[0, 0]	1.8333	100, 100, 100
DIR	ALT1	0	804.9449	[0, 0]	1.1333	100, 100, 100
ALT1	ALT1	0	804.9449	[0, 0]	1.0667	100, 100, 100
ALT2	ALT2	0	804.9603	[0, 0]	24.7333	100, 100, 100
$\beta = 1$, $\gamma = 1$, $\tau_0 = 803$						
Benchmark	Opt x	Opt z	Obj.Value	CI	Node	N_1, N_2, N_3
BUS	ALT1	29.6610	35.0101	[0, 0.3341]	22.0870	100, 100, 225
EWY	ALT1	29.6610	34.0016	[0, 0.2120]	21.6974	100, 100, 300
DIR	ALT1	34.9254	34.0521	[0, 0.2905]	46.7500	150, 100, 150
ALT1	ALT1	28.3608	34.2108	[0, 0.3394]	3.9333	100, 100, 125
ALT2	ALT2	29.6746	35.2798	[0, 0.2496]	30.4000	100, 100, 125
$\beta = 1$, $\gamma = 1$, $\tau_0 = 1400$						
Benchmark	Opt x	Opt z	Obj.Value	CI	Node	N_1, N_2, N_3
BUS	ALT1	625.3610	-265.7663	[0, 0.6693]	14.7353	100, 100, 100
EWY	ALT1	625.3610	-265.5896	[0, 0.8682]	81.0526	100, 100, 100
DIR	ALT1	630.5240	-265.7726	[0, 0.7639]	22.2353	100, 100, 100
ALT1	ALT1	625.3610	-265.1846	[0, 0.4550]	2.8000	100, 100, 100
ALT2	ALT2	626.6750	-264.1333	[0, 0.3576]	31.6333	100, 100, 100
$\beta = 3$, $\gamma = 1$, $\tau_0 = 0$						
Benchmark	Opt x	Opt z	Obj.Value	CI	Node	N_1, N_2, N_3
BUS	ALT1	0	804.9449	[0, 0]	1.6667	100, 100, 100
EWY	ALT1	0	804.9449	[0, 0]	1.8333	100, 100, 100
DIR	ALT1	0	804.9449	[0, 0]	1.1333	100, 100, 100
ALT1	ALT1	0	804.9449	[0, 0]	1.0667	100, 100, 100
ALT2	ALT2	0	804.9603	[0, 0]	27.0667	100, 100, 100
$\beta = 3$, $\gamma = 1$, $\tau_0 = 803$						
Benchmark	Opt x	Opt z	Obj.Value	CI	Node	N_1, N_2, N_3
BUS	ALT2	55.9889	43.0253	[0, 0.3158]	25.3014	125, 100, 150
EWY	ALT1	56.3233	44.1235	[0, 0.3681]	27.6378	175, 100, 300
DIR	ALT2	58.1520	43.1138	[0, 0.4114]	43.3714	125, 100, 150
ALT1	ALT1	52.2191	43.4742	[0, 0.3944]	8.6500	125, 100, 175
ALT2	ALT2	49.1741	43.5080	[0, 0.3083]	38.2200	125, 100, 150

Table 3: Sensitivity Analysis.

$\beta = 3$, $\gamma = 1$, $\tau_0 = 1400$						
Benchmark	Opt x	Opt z	Obj.Value	CI	Node	N_1, N_2, N_3
BUS	ALT1	656.6360	-253.8734	[0, 0.7075]	12.6757	100, 100, 100
EWY	ALT1	646.1570	-253.1176	[0, 1.9605]	13.5400	100, 100, 100
DIR	ALT2	653.4030	-254.5511	[0, 1.5771]	23.7973	125, 100, 100
ALT1	ALT1	646.1570	-254.1837	[0, 1.7365]	7.3333	100, 100, 100
ALT2	ALT2	647.2690	-254.4200	[0, 1.1906]	21.7333	100, 100, 100
$\beta = 1$, $\gamma = 3$, $\tau_0 = 0$						
Benchmark	Opt x	Opt z	Obj.Value	CI	Node	N_1, N_2, N_3
BUS	ALT1	0	2414.8346	[0, 0]	1.6667	100, 100, 100
EWY	ALT1	0	2414.8346	[0, 0]	1.8333	100, 100, 100
DIR	ALT1	0	2414.8346	[0, 0]	1.1333	100, 100, 100
ALT1	ALT1	0	2414.8346	[0, 0]	1.0667	100, 100, 100
ALT2	ALT2	0	2414.8809	[0, 0]	38.7333	100, 100, 100
$\beta = 1$, $\gamma = 3$, $\tau_0 = 803$						
Benchmark	Opt x	Opt z	Obj.Value	CI	Node	N_1, N_2, N_3
BUS	ALT1	10.8126	82.9083	[0, 0.7303]	29.4580	100, 100, 175
EWY	ALT1	10.8126	82.9083	[0, 0.7303]	29.5725	100, 100, 175
DIR	ALT1	17.6302	82.0759	[0, 0.4887]	14.9600	100, 100, 150
ALT1	ALT1	10.8126	80.8962	[0, 0.8000]	4.3429	100, 100, 100
ALT2	ALT2	18.8600	83.0679	[0, 0.3239]	22.4667	100, 100, 100
$\beta = 1$, $\gamma = 3$, $\tau_0 = 1400$						
Benchmark	Opt x	Opt z	Obj.Value	CI	Node	N_1, N_2, N_3
BUS	ALT1	614.6300	-812.4292	[0, 0.6581]	9.3023	100, 100, 100
EWY	ALT1	614.6300	-813.3732	[0, 0.6392]	11.0930	100, 100, 100
DIR	ALT1	614.6300	-812.4628	[0, 0.5637]	12.3947	100, 100, 100
ALT1	ALT1	614.6300	-812.4274	[0, 0.2807]	7.6667	100, 100, 100
ALT2	ALT2	615.8600	-812.4321	[0, 0.3239]	19.8667	100, 100, 100

Table 4: Sensitivity Analysis(cont.)

A few comments are in order. First, notice that the optimal path is always either path ALT1 or path ALT2, which are indeed very similar. Also, we see that, for the cases when $\tau_0 = 0$, we obtain the same optimal solutions regardless of the values of β and γ. That makes sense because when $\tau_0 = 0$, z and β do not play any role. A change in the parameter γ only multiplies the optimal objective value accordingly. Note also that when $\tau_0 = 0$ the confidence interval for the optimality gap is [0, 0], which indicates that the same optimal solution was obtained in all of the replications.

A comparison of results for $\beta = 1,\ \gamma = 1$ and $\beta = 3,\ \gamma = 1$ for the same value of τ_0 shows that when earliness penalty β increases the traveler would rather be late, which means the vehicle departs later and therefore z is bigger. Similarly, when changing the lateness penalty $\gamma = 1$ to $\gamma = 3$, the traveler would rather be early, which means the vehicle departs earlier and therefore z is smaller. Note here that for the same set of

parameters β, γ, τ_0, the optimal release time for different benchmark paths are close but not the same due to the randomness of sampling in the SAA procedure.

The choice of benchmark path will also affect the optimal path. If the benchmark is chosen to be path ALT1 — which is optimal in most cases — then there will be fewer feasible solutions. Even though ALT2 is also an optimal path in other cases, it does not dominate ALT1 in second order. Therefore the optimal solution for the case with benchmark path ALT1 is easier to find; indeed, we see that the average number of nodes explored in such cases is much smaller than for the other benchmarks.

6.1.4 Testing the effect of the dominance constraints

It is clear from the above results that, paths ALT1 and ALT2 are both superior to other paths. In fact, when we solve the optimal path problem without dominance constraints we can again see that ALT1 and ALT2 are still the optimal paths. In order to test for the effect of dominance constraints, we artificially inflated the variance of some arcs in ALT1 and ALT2 while keeping their expected travel time the same. The idea of such a change is that, by inflating the variance of those arcs, paths ALT1 and ALT2 become riskier and therefore will no longer dominate the other benchmarks, so they will no longer be optimal in such cases (note however that since we do not change the expected travel time, the inflated paths ALT1 and ALT2 will not be dominated by the other benchmarks either, cf. equation (22)). To accomplish the variance inflation we proceeded as follows. Let X be the travel time on a given arc that we want to inflate. We then define a new travel time X' as $X' := kX - (k-1)\mathbb{E}[X]$, where k is a scalar. Clearly, $\mathbb{E}[X'] = \mathbb{E}[X]$, but the variance of X' is $\mathrm{Var}(X') = k^2\mathrm{Var}(X)$. We pick $1 < k < \frac{\mathbb{E}[X]}{\mathbb{E}[X]-\min X}$ to make sure that X' positive w.p.1.

We tested the above procedure for the configuration with $\beta = \gamma = 1$, $\tau_0 = 803$, and test five benchmark paths. The result is summarized in Table 5. We see that now the path PEWY1 is optimal for all but the ALT1 and ALT2 benchmarks, which shows that the dominance constraint fulfills its mission of removing risky paths.

$\beta = 1$, $\gamma = 1$, $\tau_0 = 803$						
Benchmark	Opt x	Opt z	Obj.Value	CI	Node	N_1, N_2, N_3
BUS	PEWY1	14.5229	45.1031	[0, 0.4070]	30.2069	125, 100, 250
EWY	PEWY1	12.7212	44.4898	[0, 0.3496]	35.5693	150, 100, 150
DIR	PEWY1	10.9922	45.0784	[0, 0.4197]	84.1539	100, 100, 250
ALT1	ALT1	50.9570	49.7994	[0, 0.3453]	3.6667	100, 100, 125
ALT2	ALT2	43.9097	49.2423	[0, 0.4454]	19.8000	125, 100, 125

Table 5: Test dominance constraints

7 Conclusion

While path optimization problems with stochastic travel times have long been studied in the literature, recent attention has been given to approaches that incorporate measures of risk into such problems. In this paper we have presented one such approach,

where risk is represented in constraints that ensure that a chosen path must stochastically dominate (in second order) a benchmark route. Previous work in the literature on optimization problems with stochastic dominance constraints has demonstrated the applicability of such techniques whenever there is a natural benchmark to compare against. In the context of the present paper, the benchmark can be easily interpreted as a standard route (e.g., one that uses expressways).

One difficulty that arises when using stochastic dominance constraints is that the explicit form of such problems — where the dominance constraint is replaced by a set of linear inequalities — leads to extremely large formulations when the number of possible scenarios is large, which is the case of large networks. To circumvent that problem, we have used a Sample Average Approximation approach, which has been well-established in the literature as a way to approximate large stochastic optimization problems; along the way, we have provided some new statistical bounds on the optimal value that are specific for problems with stochastic dominance constraints. Our numerical experiments for a large-scale network with more than 1,500 arcs corresponding to part of the Chicago area demonstrate the potential and the feasibility of using the proposed approach in real world problems.

The models presented in this paper could be extended in a number of ways; for example, to multi-objective problems obtained by adding fixed cost for transportation and other possible criteria one may want to evaluate. Multi-benchmark problems could also be considered. Such an extension is more involved since one is dealing with multivariate stochastic dominance constraints. Previous work in the literature has dealt with such problems (Dentcheva and Ruszczyński 2009, Homem-de-Mello and Mehrotra 2009, Hu et al. 2011b, 2014) but the adaptation to the path optimization setting requires careful study of the optimization algorithm to be used (perhaps a different form of the branch-and-cut developed in this paper) and the derivation of appropriate statistical bounds.

Acknowledgments

The authors thank Marco (Yu) Nie and Xing Wu for providing the data used in the numerical experiments. We also thank three anonymous referees for their comments, which helped us improve the presentation of our results. This work has been supported in part by the National Science Foundation under Grant 1033048, and by Conicyt-Chile under grant Fondecyt 1120244.

References

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Math Financ*, 9:203–227, 1999.

G. Bayraksan and D. P. Morton. Assessing solution quality in stochastic programs. *Math Program*, 108:495–514, 2006.

G. Bayraksan and D. P. Morton. Assessing solution quality in stochastic programs via sampling. In *Tutorials in Operations Research*, pages 102–122. INFORMS, 2009.

A. Beaudry, G. L. andTeresa Melo, and S. Nickel. Dynamic transportation of patients in hospitals. *OR Spectrum*, 32(1):77–107, 2010.

M. G. Bell and C. Cassir. Risk-averse user equilibrium traffic assignment: an application of game theory. *Transportation Research Part B*, 36(8):671–681, 2002.

D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Mathematical Programming Series B*, 98:49–71, 2003.

B. Biller and S. Ghosh. Multivariate input processes. In S. G. Henderson and B. L. Nelson, editors, *Simulation*, volume 13 of *Handbooks in Operations Research and Management Science*, chapter 5. Elsevier Science Publishers B.V., Amsterdam, Netherlands, 2006.

J. M. Casello, A. Nour, and B. Hellinga. Quantifying impacts of transit reliability on user costs. *Transportation Research Record: Journal of the Transportation Research Board*, 2112(1):136–141, 2009.

R. Cominetti and A. Torrico. Additive consistency of risk measures and its application to risk-averse routing in networks. Manuscript, University of Chile, available at `arXiv:1312.4193v1[math.OC]`, 2013.

A. De Palma, P. Hansen, and M. Labbé. Commuters' paths with penalties for early or late arrival time. *Transportation Science*, 24(4):276–286, 1990.

D. Dentcheva and W. Römisch. Stability and sensitivity of stochastic dominance constrained optimization models. *SIAM J Optimiz*, 23(3):1672–1688, 2013.

D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM J. Optim.*, 14(2):548–566, 2003.

D. Dentcheva and A. Ruszczyński. Portfolio optimization with stochastic dominance constraints. *Journal of Banking Finance*, 30:433–451, 2006.

D. Dentcheva and A. Ruszczyński. Optimization with multivariate stochastic dominance constraints. *Math. Programming*, 117:111–127, 2009.

D. Drapkin, R. Gollmer, U. Gotzes, F. Neise, and R. Schultz. Risk modeling via stochastic dominance in power systems with dispersed generation. In *Stochastic Optimization Methods in Finance and Energy*, volume 163 of *International Series in Operations Research & Management Science*, pages 253–271. Spinger, 2011.

A. Eiger, P. B. Mirchandani, and H. Soroush. Path preferences and optimal paths in probabilistic networks. *Transportation Science*, 19(1):75–84, 1985.

Y. Fan, R. Kalaba, and J. Moore. Arriving on time. *Journal of Optimization Theory and Applications*, 127(3):497–513, 2005.

H. Frank. Shortest paths in probabilistic graphs. *Operations Research*, 17(4):583–599, 1969.

T. Homem-de-Mello and G. Bayraksan. Monte Carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, 19:56–85, 2014.

T. Homem-de-Mello and S. Mehrotra. A cutting surface method for linear programs with polyhedral stochastic dominance constraints. *SIAM J Optimiz*, 20(3):1250–1273, 2009.

T. Homem-de-Mello, V. L. de Matos, and E. C. Finardi. Sampling strategies and stopping criteria for stochastic dual dynamic programming: a case study in long-term hydrothermal scheduling. *Energy Systems*, 2:1–31, 2011.

J. Hu, T. Homem-de-Mello, and S. Mehrotra. Sample average approximation of stochastic dominance constrained programs. *Math. Program., Ser. A*, 133:171–201, 2011a.

J. Hu, T. Homem-de-Mello, and S. Mehrotra. Risk adjusted budget allocation models with application in homeland security. *IIE Trans*, 43:819–839, 2011b.

J. Hu, T. Homem-de-Mello, and S. Mehrotra. Stochastically weighted stochastic dominance concepts with an application in capital budgeting. *Eur J Oper Res*, 232: 572–583, 2014.

P. Jaillet, J. Qi, and M. Sim. Routing optimization with deadlines under uncertainty. Manuscript, Laboratory for Information and Decision Systems, Operations Research Center, MIT, 2013.

N. E. Karoui and A. Meziou. Constrained optimization with respect to stochastic dominance: Application to portfolio insurance. *Math Financ*, 16(1):103–117, 2006.

A. Kenyon and D. P. Morton. Stochastic vehicle routing problems with random travel times. *Transport Sci*, 37(1):69–82, 2003.

A. Kleywegt, A. Shapiro, and T. Homem-de-Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12 (2):479–502, 2001.

R. P. Loui. Optimal paths in graphs with stochastic or multidimensional weights. *Communications of the ACM*, 26(9):670–676, 1983.

J. Luedtke. New formulations for optimization under stochastic dominance constraints. *SIAM J Optimiz*, 19(3):1433–1450, 2008.

W. K. Mak, D. P. Morton, and R. K. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Oper Res Lett*, 24:47–56, 1999.

M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245, 1979.

E. D. Miller-Hooks and H. S. Mahmassani. Path comparisons for a priori and time-adaptive decisions in stochastic, time-varying networks. *European Journal of Operational Research*, 146(2):67–82, 2003.

P. B. Mirchandani. Shortest distance and reliability of probabilistic networks. *Computers and Operations Research*, 3(4):347–355, 1976.

A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Chichester, 2002.

I. Murthy and S. Sarkar. A relaxation-based pruning technique for a class of stochastic shortest path problems. *Transportation Science*, 30(3):220–236, 1996.

Y. Nie and Y. Fan. Arriving-on-time problem: Discrete algorithm that ensures convergence. *Transporttion Research Record*, 1964:193–200, 2006.

Y. Nie and X. Wu. Shortest path problem considering on-time arrival probability. *Transportation Research Part B*, 43(6):597–613, 2009.

Y. Nie, X. Wu, J. F. Dillenburg, and P. C. Nelson. Reliable route guidance: A case study from Chicago. *Transportation Research Part A*, 46(2):403–419, 2012a.

Y. Nie, X. Wu, and T. Homem-de-Mello. Optimal path problems with second-order stochastic dominance constraints. *Networks and Spatial Economics*, 12:561–587, 2012b.

E. Nikolova and N. E. Stier-Moses. A mean-risk model for the stochastic traffic assignment problem with stochastic travel times. *Oper Res*, 62(2):366–382, 2014.

E. Nikolova, M. Brand, and D. R. Karger. Optimal route planning under uncertainty. In *ICAPS*, volume 6, pages 131–141, 2006.

N. Noyan, G. Rudolf, and A. RuszczyńSki. Relaxations of linear programming problems with first order stochastic dominance constraints. *Operations Research Letters*, 34(6):653–659, 2006.

F. Ordóñez and N. E. Stier-Moses. Wardrop equilibria with risk-averse users. *Transportation Science*, 44(1):63–86, 2010.

D. Roman, K. Darby-Dowman, and G. Mitra. Portfolio construction based on stochastic dominance and target return distributions. *Math. Programming*, 108:541–569, 2006.

W. Römisch. Stability of stochastic programming problems. In A. Ruszczyński and A. Shapiro, editors, *Handbook of Stochastic Optimization*. Elsevier Science Publishers B.V., Amsterdam, Netherlands, 2003.

S. Sen, R. Pillai, S. Joshi, and A. Rathi. A mean-variance model for route guidance in advanced traveler information systems. *Transportation Science*, 35(1):37–49, 2001.

A. Shapiro. Monte Carlo sampling methods. In A. Ruszczynski and A. Shapiro., editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier Science Publishers B.V., Amsterdam, Netherlands, 2003.

R. Sivakumar and R. Batta. The variance-constrained shortest path problem. *Transportation Science*, 28(4):309–316, 1994.

B. Verweij, S. Ahmed, A. Kleywegt, G. Nemhauser, and A. Shapiro. The sample average approximation method applied to stochastic routing problems: A computational study. *Computational Optim. Appl.*, 24:289–333, 2003.

W. Wang and S. Ahmed. Sample average approximation of expected value constrained stochastic programs. *Oper Res Lett*, 36:515–519, 2008.

L. Xiao and H. K. Lo. Adaptive vehicle routing for risk-averse travelers. *Procedia – Social and Behavioral Sciences*, 80:633–657, 2013.

G. Yu and J. Yang. On the robust shortest path problem. *Computers and Operations Research*, 25(6):457–468, 1998.