Submitted to *INFORMS Journal on Optimization*
manuscript

Satisficing Models under Uncertainty

Patrick Jaillet
Department of Electrical Engineering and Computer Science, Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, jaillet@mit.edu

Sanjay Dominik Jena
École des Sciences de la Gestion, Université du Québec à Montréal, Montréal, Canada, jena.sanjay-dominik@uqam.ca

Tsan Sheng Ng
Department of Industrial & Systems Engineering, National University of Singapore, Singapore 117576, isentsa@nus.edu.sg

Melvyn Sim
NUS Business School, National University of Singapore, Singapore 119245, dscsimm@nus.edu.sg

Satisficing, as an approach to decision-making under uncertainty, aims at achieving solutions that satisfy the problem's constraints as well as possible. Mathematical optimization problems that are related to this form of decision-making include the P-model of Charnes and Cooper (1963). In this paper, we propose a general framework of satisficing decision criteria, and show a representation termed the S-model, of which the P-model and robust optimization models are special cases. We then focus on the linear optimization case, and obtain a tractable probabilistic S-model, termed the T-model, whose objective is a lower bound of the P-model. We show that when probability densities of the uncertainties are log-concave, the T-model can admit a tractable concave objective function. In the case of discrete probability distributions, the T-model is a linear mixed integer optimization problem of moderate dimensions. Our computational experiments on a stochastic maximum coverage problem suggest that the T-model solutions can be highly competitive compared to standard sample average approximation models.

Key words: satisficing, optimization under uncertainty, mathematical optimization

1. Introduction

Uncertainty is ubiquitous in many real-world decision problems. A common approach to deal with uncertainty in an optimization problem is to incorporate attitudes of risk, when probability distributions are available or are ambiguous. Such models have been extensively studied in stochastic programming (see, e.g., Prékopa 1995, Birge and Louveaux 2011, Prékopa 2003), robust optimization (see, e.g., Ben-Tal and Nemirovski 1999, Bertsimas et al. 2011) and distributionally robust optimization (see, e.g., Delage and Ye 2010, Goh and Sim 2010, Wiesemann et al. 2014). In many practical situations, the goal of the decision maker may not necessarily be to maximize benefits or minimize costs, but rather to mitigate a collection of risks such as running over budgets, non-fulfillment of service and other execution failures for which the negative impacts are hard to quantify. For instance, it is reasonable in a project management problem with uncertain activity completion times to ensure that the project can be completed on schedule and within the allocated budget (see, e.g., Goh and Hall 2013).

The concept of *satisficing*, a portmanteau of the terms 'satisfy' and 'suffice' first introduced by Simon (1959), addresses uncertainty with the aims of achieving feasibility in an uncertain environment. Here, decision-makers may be more interested in obtaining solutions that can "satisfice" the constraints of the problem in some sense, as well as possible. Satisficing as an objective in decision making has mostly been explored from the economic perspective (see, e.g. Güth 2010, Stüttgen et al. 2012). Charnes and Cooper (1963) were the first to incorporate the principal idea of satisficing in the mathematical framework of success probability maximization, which has been termed the *P-model*. In simplified form, the *P-model* can be stated as:

$$\begin{array}{l} \max \ \ln \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{T}(\boldsymbol{x})\right] \\ \text{ s.t. } \boldsymbol{x} \in \mathcal{X}, \end{array} \tag{1}$$

where $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^N$ is a N dimensional vector of decision variables, $\tilde{\boldsymbol{z}}$ is a K dimensional random vector (under the probability measure $\mathbb{P}$ and with a distribution of support $\mathcal{W}$) representing uncertain perturbations and $\mathcal{T}(\boldsymbol{x}) \subseteq \mathcal{W}$ defines the tolerance set of uncertain perturbations for which the solution $\boldsymbol{x}$ would remain feasible. For a given $\boldsymbol{x} \in \mathcal{X}$, the tolerance set $\mathcal{T}(\boldsymbol{x})$ can, for example, represent the ubiquitous linear optimization format

$$\mathcal{T}(\boldsymbol{x}) = \{\boldsymbol{z} \in \mathcal{W} \mid \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \geq \boldsymbol{b}(\boldsymbol{z})\}, \tag{2}$$

for given affine maps $\boldsymbol{A} : \mathbb{R}^K \mapsto \mathbb{R}^{M \times N}$ and $\boldsymbol{b} : \mathbb{R}^K \mapsto \mathbb{R}^M$ (such 'affine dependence' assumptions are rather common in the literature for modelling influence of uncertainties on problem parameters, see, e.g., Ben-Tal et al. 2004, Chen et al. 2007). It can also represent a linear optimization with recourse such as:

$$\mathcal{T}(\boldsymbol{x}) = \left\{ \boldsymbol{z} \in \mathcal{W} \left| \begin{array}{l} \exists \boldsymbol{y}(\boldsymbol{\zeta}) \in \mathbb{R}^{N_2} \ \forall \boldsymbol{\zeta} \in \mathcal{W} \ : \\ \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}) \end{array} \right. \right\}, \tag{3}$$

for a given matrix $\boldsymbol{B} \in \mathbb{R}^{M \times N_2}$. Here, $\boldsymbol{x}$ can be referred to as the *here-and-now* decision variables, and a given realization of the uncertainty $\boldsymbol{z}$ is then acceptable for $\boldsymbol{x}$ (i.e., belongs to $\mathcal{T}(\boldsymbol{x})$), if there exists a *wait-and-see* recourse function $\boldsymbol{y}(\boldsymbol{z})$ that insures feasibility of the linear constraints in (3). We assume however, unless otherwise stated, that $\mathcal{T}(\boldsymbol{x})$ can be general in structure and is therefore not limited to the above cases.

Even in the case of the linear optimization format, the P-model (1) is in general known to be a difficult optimization problem. For instance, evaluating the log-probability objective function can itself be computationally challenging even when $M = 1$ and the random variables are i.i.d. uniformly distributed (Nemirovski and Shapiro 2006). For more general distributions, one may use sample average approximation approaches, such as Monte Carlo methods (see, e.g., Shapiro 2003) to approximate the objective function. Unfortunately, even small problems with relatively simple structure can require hundreds of samples to achieve a desired level of accuracy (Shapiro and Homem-de-Mello 2000, Pagnoncelli et al. 2009).

In view of the above, the goal of this work is to develop a new and generalized representation of *satisficing decision criteria* inspired by the P-model, which evaluates how well a solution $\boldsymbol{x}$ would remain feasible in the problem's constraints under uncertainty. In this regard, the P-model is an optimization problem that maximizes a log-probability satisficing decision criterion, $\nu_P : \mathcal{X} \mapsto \mathbb{R}$, given by

$$\nu_P(\boldsymbol{x}) = \ln \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{T}(\boldsymbol{x})\right]. \tag{4}$$

Note that the decision criterion is based on a log-probability function, instead of directly a probability function. Using the convention $\ln 0 = -\infty$, if $\boldsymbol{x}$ is always infeasible in the problem's constraints, then we can set $\nu_P(\boldsymbol{x}) = -\infty$ for such $\boldsymbol{x}$, thereby extending the definition of ν_P to $\mathbb{R}^N$ such that $\nu_P : \mathbb{R}^N \mapsto \mathbb{R} \cup \{-\infty\}$. The log-probabilty is also convenient in some situations where it leads to a concave objective function. Consider the case of $\mathcal{T}(\boldsymbol{x})$ in (2), and assume also $\boldsymbol{A}(\cdot)$ to be constant and $\boldsymbol{b}(\tilde{\boldsymbol{z}})$ to be affinely dependent on $\tilde{z}_1, \ldots, \tilde{z}_K$, which are independently distributed random variables with log-concave density functions. In this special case, the log-probability criterion ν_P is known to be concave (see, e.g., Prékopa 2003, Theorem 2.5). Furthermore, in the absence of recourse, if the probability function is log-concave in $\boldsymbol{x}$, then the objective function of Problem (1) would be concave. In general, however, as also mentioned above, (1) is computationally hard. Hence, our ambition is to develop alternative satisficing models that are computationally more attractive, that can be applied to optimization paradigms beyond the linear optimization context, and yet retain some salient satisficing characteristics of the P-model.

In relation to the above, in this paper we propose a functional representation of a class of satisficing decision criteria, and its corresponding *S-model*, where the objective is to maximize a satisficing decision criterion from that class. We show that the S-model framework encompasses as special cases the P-model and models based on robust optimization in the literature (see Section 2). In Section 3, we focus on tractable probabilistic S-models. We define a subclass, the *T-model*, which while providing a lower bound for the P-model, can be much more computationally appealing than the

P-model. Our numerical studies in Section 4 for a stochastic maximum coverage problem demonstrate that the T-model outperforms other benchmarks based on sample average approximations in terms of solution quality, computational efficiency and scalability.

Notation. Given $N \in \mathbb{N}$, we use $[N]$ to denote the set of running indices, $\{1,\ldots,N\}$. We generally use bold faced characters such as $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ to represent vectors and matrices. We use x_i to denote the i^{th} element of vector $\boldsymbol{x}$. We let $\boldsymbol{e}_i$ be the unit vector, i.e., the i^{th} column of the identity matrix. We use the tilde sign to denote an uncertain or random parameter such as $\tilde{z}$ without necessarily associating it with a particular probability distribution. For a set $\mathcal{U} \subseteq \mathbb{R}^K$, $\mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{U}]$ represents the probability of $\tilde{\boldsymbol{z}}$ being in the set $\mathcal{U}$ evaluated on the distribution $\mathbb{P}$. We use the convention, $\ln 0 = \max \emptyset = -\infty$ and $\min \emptyset = \infty$.

2. Satisficing Decision Criteria and Models

In this section, we formalize the notion of satisficing decision criteria that would encompass, as special cases, the log-probability decision criterion in (4), as well as satisficing decision criteria derived from robust optimization models. We then provide a general representation of this class of satisficing decision criteria, which enables us to construct different types of satisficing models for decision-making. We then establish links to existing satisficing decision criteria in the literature and provide some examples of tractable satisficing models.

2.1. Definition and representation of satisficing decision criteria

Based on the premise of a satisficing approach to decision-making under uncertainty as motivated in the introduction, we first define the key characteristics of a *satisficing decision criterion*. In the remainder of the paper, as also introduced in Section 1, $\mathcal{W}$ refers to the support of the K-dimensional primitive uncertain variables $\tilde{\boldsymbol{z}}$, $\boldsymbol{x} \in \mathcal{X}$ refers to values of decision variables from the set $\mathcal{X} \subseteq \mathbb{R}^N$, and $\mathcal{T}(\boldsymbol{x})$ refers to its associated tolerance set, comprised of all $\boldsymbol{z}$ for which $\boldsymbol{x}$ remains feasible. By convention we will set $\mathcal{T}(\boldsymbol{x}) = \emptyset$ for any $\boldsymbol{x} \in \mathbb{R}^N \setminus \mathcal{X}$.

Definition 1. Given a family of tolerance sets, $\mathcal{T}(\boldsymbol{x}) \subseteq \mathcal{W}$. A function $\nu : \mathbb{R}^N \mapsto \mathbb{R} \cup \{-\infty\}$ is a *satisficing decision criterion* if it has the following two properties.

1. (Satisficing dominance) If $\mathcal{T}(\boldsymbol{y}) \subseteq \mathcal{T}(\boldsymbol{x})$, then $\nu(\boldsymbol{x}) \geq \nu(\boldsymbol{y})$.
2. (Infeasibility) If $\mathcal{T}(\boldsymbol{x}) = \emptyset$ then $\nu(\boldsymbol{x}) = -\infty$.

The *satisficing dominance* property ensures that, if for all $\boldsymbol{z} \in \mathcal{W}$, $\boldsymbol{x}$ is feasible (i.e. $\boldsymbol{z} \in \mathcal{T}(\boldsymbol{x})$) whenever $\boldsymbol{y}$ is feasible (i.e., $\boldsymbol{z} \in \mathcal{T}(\boldsymbol{y})$), then $\boldsymbol{x}$ should be no less preferred than $\boldsymbol{y}$. As a consequence, if $\boldsymbol{x}$ and $\boldsymbol{y}$ are always feasible across all uncertain outcomes $\boldsymbol{z} \in \mathcal{W}$, then they should be most preferred (i.e., have highest ν value) and $\nu(\boldsymbol{x}) = \nu(\boldsymbol{y})$, because $\nu(\boldsymbol{x}) \geq \nu(\boldsymbol{y})$ and $\nu(\boldsymbol{y}) \geq \nu(\boldsymbol{x})$. In contrast, the *infeasibility* property requires that a solution that is not feasible in any $\boldsymbol{z} \in \mathcal{W}$ would not be an acceptable solution.

We now provide a general representation of any such satisficing decision criterion ν in the following result.

THEOREM 1. *Given a family of tolerance sets, $\mathcal{T}(\boldsymbol{x}) \subseteq \mathcal{W}$, consider a function $\nu : \mathbb{R}^N \mapsto \mathbb{R} \cup \{-\infty\}$ defined as*

$$\nu(\boldsymbol{x}) = \max_{\boldsymbol{\alpha} \in \mathcal{S}} \{\rho(\boldsymbol{\alpha}) \mid \mathcal{U}(\boldsymbol{\alpha}) \subseteq \mathcal{T}(\boldsymbol{x})\} \tag{5}$$

for some function $\rho : \mathcal{S} \to \mathbb{R} \cup \{-\infty\}$ on domain $\mathcal{S} \subseteq \mathbb{R}^Q$, and for some family of nonempty uncertainty sets $\mathcal{U}(\boldsymbol{\alpha}) \subseteq \mathcal{W}$ defined for all $\boldsymbol{\alpha} \in \mathcal{S}$. Then the function ν is a satisficing decision criterion. Moreover, any satisficing decision criterion can be represented in a form given by (5) with $\mathcal{S} \subseteq \mathbb{R}^N$.
Proof: See Appendix A.1.

Theorem 1 provides a simple yet elegant representation of the satisficing decision criterion, and as an outcome, we can now propose a general format of the *satisficing model*, named the *S-model*, where satisficing is the objective as follows:

$$\begin{array}{rl} \max & \rho(\boldsymbol{\alpha}) \\ \text{s.t.} & \mathcal{U}(\boldsymbol{\alpha}) \subseteq \mathcal{T}(\boldsymbol{x}) \\ & \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\alpha} \in \mathcal{S}. \end{array} \tag{6}$$

In the above, the maximization is taken with respect to $\boldsymbol{x}$ and $\boldsymbol{\alpha}$, and $\rho(\boldsymbol{\alpha}), \mathcal{U}(\boldsymbol{\alpha})$ and $\mathcal{S} \subseteq \mathbb{R}^Q$ are specific problem-dependent representation choices.

It can be verified that the P-model is indeed an S-model. By applying (6) with objective $\rho(\boldsymbol{\alpha}) = \ln \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{U}(\boldsymbol{\alpha})]$, uncertainty set parameters $\boldsymbol{\alpha} \in \mathbb{R}^N$ and domain $\mathcal{S} = \mathcal{X}$, and the adjustable uncertainty sets $\mathcal{U}(\boldsymbol{\alpha}) = \mathcal{T}(\boldsymbol{\alpha})$, we have:

$$\begin{array}{rl} & \max \{\ln \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{U}(\boldsymbol{\alpha})] \mid \mathcal{U}(\boldsymbol{\alpha}) \subseteq \mathcal{T}(\boldsymbol{x}),\ \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\alpha} \in \mathcal{S}\} \\ \Leftrightarrow & \max \{\ln \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{T}(\boldsymbol{\alpha})] \mid \mathcal{T}(\boldsymbol{\alpha}) \subseteq \mathcal{T}(\boldsymbol{x}),\ \boldsymbol{x}, \boldsymbol{\alpha} \in \mathcal{X}\} \\ \Leftrightarrow & \max \{\ln \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{T}(\boldsymbol{x})] \mid \boldsymbol{x} \in \mathcal{X}\}, \end{array} \tag{7}$$

where the second equivalence above follows from noting that for any given $\boldsymbol{x}$, choosing $\boldsymbol{\alpha} = \boldsymbol{x}$ always solves the maximization over $\boldsymbol{\alpha}$.

Incidentally, the S-model is quite broad and could encompass various kinds of optimization problems such as a robust linear optimization problem

$$\min_{\boldsymbol{x} \in \mathcal{X}} \{\boldsymbol{c}'\boldsymbol{x} \mid \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \le \boldsymbol{b}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in U\},$$

where U is some assumed uncertainty set, and $\boldsymbol{c}$ are cost coefficients in the objective function. By introducing an auxiliary uncertain parameter, z_0, the above can be re-written as

$$\begin{array}{rll} \max & \alpha & \\ \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \le \boldsymbol{b}(\boldsymbol{z}) & \forall \boldsymbol{z} \in U \\ & -\boldsymbol{c}'\boldsymbol{x} \le z_0 & \forall z_0 \ :\ z_0 \ge \alpha \\ & \boldsymbol{x} \in \mathcal{X} & \\ & \alpha \in \mathbb{R}, & \end{array} \tag{8}$$

noting that $-\boldsymbol{c}'\boldsymbol{x} \leq z_0 \ \forall z_0 \geq \alpha$ is equivalent to $-\boldsymbol{c}'\boldsymbol{x} \leq \alpha$. One can verify that (8) is indeed a S-model (6) by defining $\rho(\alpha) = \alpha$, $\mathcal{U}(\alpha) = \{(\boldsymbol{z}, z_0) \mid \boldsymbol{z} \in U, z_0 \geq \alpha\}$, and $\mathcal{T}(\boldsymbol{x}) = \{(\boldsymbol{z}, z_0) \mid \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \leq \boldsymbol{b}(\boldsymbol{z}), \ -\boldsymbol{c}'\boldsymbol{x} \leq z_0\}$. While the mathematical connection is interesting, it may not be intuitive for the decision maker to interpret (8).

We next introduce some satisficing decision criteria that are quasi-concave or concave based on maximal uncertainty sets, so that the corresponding S-models can be formulated as tractable convex optimization problems.

2.2. Satisficing decision criteria using maximal uncertainty sets

We consider satisficing decision criteria in which function $\rho(\boldsymbol{\alpha})$ in (5) is related to the size of the adjustable uncertainty sets $\mathcal{U}(\boldsymbol{\alpha})$. Suppose $\mathcal{X}$ and $\mathcal{W}$ are tractable convex sets, *i.e.*, optimizing a linear function over each of the set can be performed in polynomial time (see, for instance, Ben-Tal and Nemirovski 1999). Consider the following satisficing decision criterion,

$$\nu_Q(\boldsymbol{x}) = \max_{\alpha \geq 0} \{\alpha \mid \forall \boldsymbol{z} \in \mathcal{U}(\alpha) \ \ \boldsymbol{z} \in \mathcal{T}(\boldsymbol{x})\} \tag{9}$$

for a family of convex uncertainty sets satisfying $\mathcal{U}(\alpha_1) \subseteq \mathcal{U}(\alpha_2) \subseteq \mathcal{W}$ for all $0 \leq \alpha_1 \leq \alpha_2$, and a tolerance set of the form

$$\mathcal{T}(\boldsymbol{x}) = \{\boldsymbol{x} \in \mathcal{W} \ \mid \ f_i(\boldsymbol{x}, \boldsymbol{z}) \leq 0, \forall i \in [M]\},$$

where $f_i : \mathcal{X} \times \mathcal{W} \mapsto \mathbb{R}$, $i \in [M]$ are saddle functions, *i.e.*, convex over $\boldsymbol{x} \in \mathcal{X}$ for given $\boldsymbol{z} \in \mathcal{W}$ and concave over $\boldsymbol{z} \in \mathcal{W}$ for given $\boldsymbol{x} \in \mathcal{X}$. Observe that, because the family of uncertainty sets $\mathcal{U}(\alpha)$ is non-decreasing in $\alpha \in \mathbb{R}_+$, the level set $\mathcal{X}_\beta = \{\boldsymbol{x} \in \mathcal{X} \ \mid \ \nu_Q(\boldsymbol{x}) \geq \beta\}$ is a convex set,

$$\mathcal{X}_\beta = \{\boldsymbol{x} \in \mathcal{X} \mid f_i(\boldsymbol{x}, \boldsymbol{z}) \leq 0, \forall i \in [M] \ \ \forall \boldsymbol{z} \in \mathcal{U}(\beta)\}$$

and hence $\nu_Q(\boldsymbol{x})$ is a quasi-concave satisficing decision criterion. Moreover, by articulating the satisficing decision criterion as a constraint in the following optimization problem

$$\min_{\boldsymbol{x} \in \mathcal{X}} \{\boldsymbol{c}'\boldsymbol{x} \mid \nu_Q(\boldsymbol{x}) \geq \beta\},$$

where the specified parameter β is associated with the size of the uncertainty set, and $\boldsymbol{c}$ are some cost coefficients in the objective function, we obtain the tractable robust optimization models proposed in Ben-Tal et al. (2015) as follows,

$$\min_{\boldsymbol{x} \in \mathcal{X}} \{\boldsymbol{c}'\boldsymbol{x} \mid f_i(\boldsymbol{x}, \boldsymbol{z}) \leq 0, \forall i \in [M] \ \ \forall \boldsymbol{z} \in \mathcal{U}(\beta)\}. \tag{10}$$

In an S-model, the decision maker maximizes the satisficing decision criterion, subject to a constraint on the cost budget τ as follows,

$$\max_{\boldsymbol{x} \in \mathcal{X}} \{\nu_Q(\boldsymbol{x}) \mid \boldsymbol{c}'\boldsymbol{x} \leq \tau\},$$

or equivalently

$$\max_{\boldsymbol{x}\in\mathcal{X},\alpha\geq 0} \{\alpha \mid f_i(\boldsymbol{x},\boldsymbol{z}) \leq 0, \forall i \in [M], \quad \forall \boldsymbol{z} \in \mathcal{U}(\alpha), \boldsymbol{c}'\boldsymbol{x} \leq \tau\}. \tag{11}$$

Observe that the decision maker in the S-model has to specify the acceptable cost target, τ, which is often easier to interpret compared to the size of the uncertainty set, β, required in the robust optimization problem. Moreover, solving (11) is almost as efficient as solving the robust optimization problem based on Ben-Tal et al. (2015). The problem can be solved via a binary search in α, where in each step, the remaining problem with a fixed α is a robust optimization problem with a structure similar to that of (10).

In the more general case where $\boldsymbol{\alpha} \in \mathcal{S} \subseteq \mathbb{R}^N$ with $N > 1$, the following satisficing decision criterion, inspired by Zhang et al. (2016), can lead to computationally amenable formulations. Given tractable convex sets $\mathcal{X}$ and polyhedral support $\mathcal{W} = \{\boldsymbol{z} \in \mathbb{R}^K \mid \boldsymbol{D}\boldsymbol{z} \leq \boldsymbol{d}\}$, we consider the following satisficing decision criterion,

$$\nu_C(\boldsymbol{x}) = \max_{\boldsymbol{\Sigma}\succeq\boldsymbol{0},\boldsymbol{\mu}\in\mathbb{R}^N} \{\log\det\boldsymbol{\Sigma} \mid \forall \boldsymbol{z} \in \mathcal{U}(\boldsymbol{\Sigma},\boldsymbol{\mu}) \quad \boldsymbol{z} \in \mathcal{T}(\boldsymbol{x})\} \tag{12}$$

for a family of convex uncertainty sets

$$\mathcal{U}(\boldsymbol{\Sigma},\boldsymbol{\mu}) \triangleq \{\boldsymbol{z} \in \mathcal{W} \mid \exists \boldsymbol{\zeta} \in \mathcal{B}, \boldsymbol{z} = \boldsymbol{\Sigma}\boldsymbol{\zeta} + \boldsymbol{\mu}\}, \tag{13}$$

for some tractable full-dimensional base uncertainty set $\mathcal{B} \subseteq \mathcal{W}$, and a tolerance set of the form

$$\mathcal{T}(\boldsymbol{x}) = \{\boldsymbol{x} \in \mathcal{W} \mid f_i(\boldsymbol{x}) \leq b_i(\boldsymbol{z}), \forall i \in [M]\},$$

where $f_i : \mathcal{X} \mapsto \mathbb{R}$ and $b_i : \mathcal{W} \mapsto \mathbb{R}$, $i \in [M]$ are respectively convex and affine functions. Observe that the epi-graph of $-\nu_C(\boldsymbol{x})$, $\{(\boldsymbol{x}, y) \in \mathcal{X} \times \mathbb{R} \mid -\nu_C(\boldsymbol{x}) \leq y\}$ is a convex set, and hence $\nu_C(\boldsymbol{x})$ is a concave satisficing decision criterion. Here, $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are decision parameters that provide an affine transformation of the base uncertainty set, $\mathcal{B}$ to the uncertainty set, $\mathcal{U}(\boldsymbol{\Sigma},\boldsymbol{\mu})$. Note that the objective function of the satisficing decision criterion, is related to the volume of the uncertainty set, $\mathcal{U}(\boldsymbol{\Sigma},\boldsymbol{\mu})$, since

$$\text{vol}(\mathcal{U}(\Sigma,\boldsymbol{\mu})) = \det(\Sigma)\text{vol}(\mathcal{B}).$$

Hence, we can interpret the satisficing decision criterion of (12) as one that evaluates the volume of the largest uncertainty set that can be contained within the tolerance set.

The corresponding S-model can be formulated as a convex log-determinant semidefinite optimization problem with robust linear optimization type constraints as follows,

$$\begin{array}{ll} \max & \log\det\boldsymbol{\Sigma} \\ \text{s.t.} & f_i(\boldsymbol{x}) \leq b_i(\boldsymbol{\Sigma}\boldsymbol{\zeta} + \boldsymbol{\mu}) \ \forall \boldsymbol{\zeta} \in \mathcal{B}, i \in [M] \\ & \boldsymbol{D}(\boldsymbol{\Sigma}\boldsymbol{\zeta} + \boldsymbol{\mu}) \leq \boldsymbol{d} \qquad \forall \boldsymbol{\zeta} \in \mathcal{B} \\ & \boldsymbol{x} \in \mathcal{X} \\ & \boldsymbol{\Sigma} \succeq \boldsymbol{0}, \boldsymbol{\mu} \in \mathbb{R}^K. \end{array} \tag{14}$$

Note that the semi-infinite constraints in (14) involve uncertainty sets that are not dependent on the decision variables, which keeps the problem computationally tractable. We also remark that Problem (14) generalizes some earlier works, such as Hendrix et al. (1996), who consider robustness in product design problems, where uncertainties are associated with implementation errors of the design solutions. Their setting is a simplified instance of the above with $\mathcal{B}$ assumed to be a p-norm ball (with $p=1,\ 2$, or ∞), $\mathbf{\Sigma}=\tau\boldsymbol{diag}(\mathbf{1})$, where $\boldsymbol{diag}(\mathbf{1})$ is the diagonal matrix of ones and τ a scalar variable, and $\boldsymbol{\mu}=\mathbf{0}$. Also, in the model of Zhang et al. (2016), $\mathcal{W}=\mathbb{R}^K$ and $f_i(\boldsymbol{x})$, $i\in[m]$ are linear functions.

In Section 3, we show that satisficing decision criteria with maximal uncertainty sets can also conservatively approximate probability measures, which are typically hard to evaluate exactly. Consequently, for a class of linear optimization problems, we will propose probabilistic S-models that approximate the P-models and illustrate their benefits of obtaining deterministic solutions that perform well in out-of-sample studies and require less computational effort compared to sample average approximation methods.

2.3. Decision-independent worst-case uncertainty

In this section, we consider a special case which can lead to further simplification of the S-model. This is stated formally as follows.

Definition 2. Suppose there exists a mapping $\xi:\mathcal{S}\mapsto\mathcal{W}$, such that $\xi(\boldsymbol{\alpha})\in\mathcal{U}(\boldsymbol{\alpha})$, and for any $\boldsymbol{x}\in\mathcal{X}$,

$$\xi(\boldsymbol{\alpha})\in\mathcal{T}(\boldsymbol{x})\Rightarrow\mathcal{U}(\boldsymbol{\alpha})\subseteq\mathcal{T}(\boldsymbol{x}). \tag{15}$$

Then, $\xi(\boldsymbol{\alpha})$ is a *decision-independent worst-case outcome* over the set $\mathcal{U}(\boldsymbol{\alpha})$. Furthermore, we say that a problem has decision-independent worst-case uncertainty if for every $\boldsymbol{\alpha}\in\mathcal{S}$ there is some decision-independent worst-case outcome $\xi(\boldsymbol{\alpha})$.

The following result shows that, whenever there is decision-independent worst-case uncertainty, the corresponding S-model can be simplified.

Proposition 1. *Let* $\xi(\boldsymbol{\alpha})$ *be a* decision-independent worst-case outcome *for a given* $\boldsymbol{\alpha}$ *as in Definition 2. The S-model* (6) *is then equivalent to:*

$$\max_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\alpha}\in\mathcal{S}}\left\{\rho(\boldsymbol{\alpha})\mid\xi(\boldsymbol{\alpha})\in\mathcal{T}(\boldsymbol{x})\right\}. \tag{16}$$

Proof: See Appendix A.2.

Proposition 1 implies that the S-model problem reduces to an equivalent problem without uncertainty. We provide a specific example of decision-independent worst-case uncertainty in Section 3. In practice, (16) can likely be solved more efficiently than the general case of (6), particularly in the context of problems with recourse under uncertainty, for example when the tolerance set is described as in (3). We refer readers to Appendix B for a discussion on the S-model with recourse, and the implication of decision-independent worst case uncertainty.

2.4. Relation with shortfall-based satisficing criterion

A *shortfall-based* satisficing decision criterion has been proposed in Brown and Sim (2009) and extended in Lam et al. (2013). These criteria characterize how well the solution $\boldsymbol{x} \in \mathcal{X}$ associated with a *shortfall function*, $s : \mathcal{X} \times \mathcal{W} \mapsto [-\infty, \infty]$ would avoid shortfalls, *i.e.,* $s(\boldsymbol{x}, \boldsymbol{z}) \leq 0$ for all outcomes $\boldsymbol{z} \in \mathcal{W}$. Hence, the corresponding tolerance set associated with the shortfall function is

$$\mathcal{T}(\boldsymbol{x}) = \{\boldsymbol{z} \in \mathcal{W} \mid s(\boldsymbol{x}, \boldsymbol{z}) \leq 0\}. \tag{17}$$

Definition 3. Given a shortfall function $s(\boldsymbol{x}, \boldsymbol{z})$, for $\boldsymbol{x} \in \mathcal{X}$, a shortfall-based satisficing criterion is a function $\rho : \mathcal{X} \mapsto \mathbb{R} \cup \{-\infty\}$ that is endowed with the following properties:
For all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$,

1. (Monotonicity) If $s(\boldsymbol{x}, \boldsymbol{z}) \leq s(\boldsymbol{y}, \boldsymbol{z})$ $\forall \boldsymbol{z} \in \mathcal{W}$, then $\rho(\boldsymbol{x}) \geq \rho(\boldsymbol{y})$.
2. (Satisficing) If $s(\boldsymbol{x}, \boldsymbol{z}) \leq 0$ $\forall \boldsymbol{z} \in \mathcal{W}$, then $\rho(\boldsymbol{x}) \geq \rho(\boldsymbol{y})$.
3. (Infeasibility) If $s(\boldsymbol{x}, \boldsymbol{z}) > 0$ $\forall \boldsymbol{z} \in \mathcal{W}$, then $\rho(\boldsymbol{x}) = -\infty$.

The shortfall-based satisficing criteria can be extended to encompass diversification preference as in Brown and Sim (2009) and Lam et al. (2013), and they can be characterized as optimizing over a family of monetary risk measures. As a distinction, we call the satisficing decision criterion in Definition 1 a *feasibility-based* satisficing decision criterion.

Observe that

$$\begin{array}{c} s(\boldsymbol{x}, \boldsymbol{z}) \leq s(\boldsymbol{y}, \boldsymbol{z}) \ \forall \boldsymbol{z} \in \mathcal{W} \Rightarrow \mathcal{T}(\boldsymbol{y}) \subseteq \mathcal{T}(\boldsymbol{x}), \\ s(\boldsymbol{x}, \boldsymbol{z}) > 0 \ \forall \boldsymbol{z} \in \mathcal{W} \quad \Rightarrow \quad \mathcal{T}(\boldsymbol{x}) = \emptyset. \end{array}$$

Hence, any feasibility-based satisficing decision criterion associated with a given shortfall function is also a shortfall-based satisficing decision criterion.

More generally, for a given tolerance set $\mathcal{T}(\boldsymbol{x})$, we can also define the shortfall function

$$s(\boldsymbol{x}, \boldsymbol{z}) \triangleq \min\{0 \mid \boldsymbol{z} \in \mathcal{T}(\boldsymbol{x})\}.$$

Consequently, the properties of monotonicity and satisficing are collectively synonymous with the satisficing dominance property:

$$\begin{array}{c} \mathcal{T}(\boldsymbol{y}) \subseteq \mathcal{T}(\boldsymbol{x}) \Leftrightarrow s(\boldsymbol{x}, \boldsymbol{z}) \leq s(\boldsymbol{y}, \boldsymbol{z}) \ \forall \boldsymbol{z} \in \mathcal{W}, \\ \mathcal{T}(\boldsymbol{x}) = \mathcal{W} \quad \Leftrightarrow \quad s(\boldsymbol{x}, \boldsymbol{z}) \leq 0 \ \forall \boldsymbol{z} \in \mathcal{W}. \end{array}$$

Likewise for the property of infeasibility, we have the following relation:

$$\mathcal{T}(\boldsymbol{x}) = \emptyset \Leftrightarrow s(\boldsymbol{x}, \boldsymbol{z}) > 0 \ \forall \boldsymbol{z} \in \mathcal{W}.$$

The feasibility-based satisficing decision criterion can therefore be viewed as a restricted form of a shortfall-based satisficing criterion. Indeed, a shortfall-based satisficing decision criterion may violate the property of satisficing dominance. Under this property, if two tolerance sets are identical, $\mathcal{T}(\boldsymbol{x}) = \mathcal{T}(\boldsymbol{y})$, then all feasibility-based satisficing decision criterion would yield $\nu(\boldsymbol{x}) = \nu(\boldsymbol{y})$, regardless of the level of infeasibility in the problems' constraints. In contrast, the shortfall-based satisficing decision criteria proposed in Brown and Sim (2009) and Lam et al. (2013), especially those that favors diversification, are designed to be sensitive to the level of shortfalls. Therefore, as in the case of the probability measure, the feasibility-based satisficing decision criterion is better suited in applications where quantifying the level of infeasibility in the problems' constraints are inconsequential to the decision maker.

3. Probabilistic S-model

We now consider the S-model (6) with $\rho(\boldsymbol{\alpha}) = \ln \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{U}(\boldsymbol{\alpha})]$, which relates to the probability that the adjustable uncertainty set $\mathcal{U}(\boldsymbol{\alpha})$ allows feasibility of any random outcomes of $\tilde{\boldsymbol{z}}$. Correspondingly, we propose the *probabilistic S-model*,

$$\begin{array}{rl} \max & \ln \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{U}(\boldsymbol{\alpha})] \\ \text{s.t.} & \mathcal{U}(\boldsymbol{\alpha}) \subseteq \mathcal{T}(\boldsymbol{x}) \\ & \boldsymbol{x} \in \mathcal{X} \\ & \boldsymbol{\alpha} \in \mathcal{S}. \end{array} \tag{18}$$

In general, the above can be viewed as a conservative approximation of the P-model (1). In other words, (18) gives a lower bound to the optimal value of the P-model by providing an internal approximation of $\mathcal{T}(\boldsymbol{x})$ with $\mathcal{U}(\boldsymbol{\alpha})$. In the remainder of this section, we focus on a class of tractable probabilistic S-models in the linear optimization case. To provide explicit formulations, the following key assumptions are made for the rest of the paper.

Assumptions

A1. We assume the linear optimization case as in (2), where $\mathcal{T}(\boldsymbol{x}) = \{\boldsymbol{z} \mid \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \geq \boldsymbol{b}(\boldsymbol{z})\}$, and $\boldsymbol{A}(\boldsymbol{z})$ and $\boldsymbol{b}(\boldsymbol{z})$ are affine mappings in $\boldsymbol{z}$ as follows:

$$a_{ij}(\boldsymbol{z}) = a_{ij}^0 + \sum_{k\in[K]} a_{ij}^k z_k \;\; \forall i \in [M], \forall j \in [N], \quad b_i(\boldsymbol{z}) = b_i^0 + \sum_{k\in[K]} b_i^k z_k \;\; \forall i \in [M]. \tag{19}$$

A2. We assume that the uncertain parameters $\tilde{z}_k$, $k \in [K]$ are independent, but not necessarily identically distributed random variables with support $\mathcal{W}_k$ such that $\mathcal{W} = \times_{k=1}^K \mathcal{W}_k$.

In view of (19), we can also refer to the uncertain parameters $\tilde{z}_k$, $k \in [K]$, as *random factors* in our model of uncertainty. Again, as noted earlier, this model of uncertainty does allow the consideration of dependencies among the constraint coefficients.

In the following, define $\underline{\alpha}_k$, $\overline{\alpha}_k$, with $\underline{\alpha}_k \leq \overline{\alpha}_k$, $\underline{\alpha}_k, \overline{\alpha}_k \in \mathcal{W}_k$ for all $k \in [K]$. Define also $\underline{\boldsymbol{\alpha}} = (\underline{\alpha}_1, \dots, \underline{\alpha}_K)$, $\overline{\boldsymbol{\alpha}} = (\overline{\alpha}_1, \dots, \overline{\alpha}_K)$, and $\boldsymbol{\alpha} = (\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}})$, where $\boldsymbol{\alpha}$ are the adjustable uncertainty set parameters in Problem (6). We then have $\boldsymbol{\alpha} \in \mathcal{S} \subseteq \mathbb{R}^Q$ where $Q = 2K$ and $\mathcal{S} = \{(\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}) \in \mathbb{R}^{2K} \;:\; \underline{\boldsymbol{\alpha}} \leq \overline{\boldsymbol{\alpha}},\; \underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}} \in \mathcal{W}\}$. The family of adjustable uncertainty sets $\mathcal{U}(\boldsymbol{\alpha})$ in (18) is here defined as:

$$\mathcal{U}(\boldsymbol{\alpha}) = \mathcal{U}(\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}) = \left\{ \boldsymbol{z} \in \mathbb{R}^K : \boldsymbol{z} \in [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}] \right\}.$$

Under the assumption of stochastic independence and the "box" typed sets $\mathcal{U}(\boldsymbol{\alpha})$, the objective function of the probabilistic S-model can then be evaluated as:

$$\ln \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{U}(\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}})\right] = \ln \prod_{k \in [K]} \mathbb{P}\left[\underline{\alpha}_k \leq \tilde{z}_k \leq \overline{\alpha}_k\right] = \sum_{k \in [K]} \ln \mathbb{P}[\underline{\alpha}_k \leq \tilde{z}_k \leq \overline{\alpha}_k].$$

Note that a computational attractiveness of the above is that it does not require high dimensional integration to be performed. The probabilistic S-model in (18), which we now call the *T-model*, then takes the following form:

$$\begin{array}{ll} \max & \displaystyle\sum_{k \in [K]} \ln \mathbb{P}\left[\underline{\alpha}_k \leq \tilde{z}_k \leq \overline{\alpha}_k\right] \\ \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}] \\ & \boldsymbol{x} \in \mathcal{X} \\ & \underline{\boldsymbol{\alpha}} \leq \overline{\boldsymbol{\alpha}},\; \underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}} \in \mathcal{W}. \end{array} \tag{20}$$

The next result shows a special case where the T-model (20) and P-model (1) turn out to be equivalent.

THEOREM 2. *The P-model (1) is equivalent to the T-model (20) if each constraint* $\boldsymbol{a}_i(\tilde{\boldsymbol{z}})\boldsymbol{x} \geq b_i(\tilde{\boldsymbol{z}})$, $i \in [M]$, *is affected by at most one random factor* $\tilde{z}_{\kappa(i)}$, *where* $\kappa : [M] \mapsto [K]$ *is a function that identifies the random factor* $\tilde{z}_k$ *for the* i^{th} *constraint.*

Proof: See Appendix A.3.

We now turn our attention to reformulations, or *robust counterparts*, of the T-model (20) that are more amenable to computation.

THEOREM 3. *The T-model (20), based on Assumptions A1 and A2, is equivalent to the following explicit nonlinear optimization problem:*

$$\begin{array}{ll} \max & \displaystyle\sum_{k \in [K]} \ln \mathbb{P}\left[\underline{\alpha}_k \leq \tilde{z}_k \leq \overline{\alpha}_k\right] \\ \text{s.t.} & \displaystyle\sum_{j \in [N]} a_{ij}^0 x_j + \sum_{k \in [K]} v_{ik} \geq b_i^0 \;\; \forall i \in [M] \\ & v_{ik} \leq \displaystyle\sum_{j \in [N]} a_{ij}^k x_j \overline{\alpha}_k - b_i^k \overline{\alpha}_k \;\; \forall i \in [M],\; k \in [K] \\ & v_{ik} \leq \displaystyle\sum_{j \in [N]} a_{ij}^k x_j \underline{\alpha}_k - b_i^k \underline{\alpha}_k \;\; \forall i \in [M],\; k \in [K] \\ & \boldsymbol{x} \in \mathcal{X}, \boldsymbol{v} \in \mathbb{R}^{M \times K}, \\ & \underline{\boldsymbol{\alpha}} \leq \overline{\boldsymbol{\alpha}}, \quad \underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}} \in \mathcal{W}. \end{array} \tag{21}$$

Proof: See Appendix A.4.

The next result shows that under some additional assumptions on the uncertainty model, we can also obtain a decision-independent worst case (see Definition 2) for the T-model.

Proposition 2. *Suppose that in the T-model* (20)*, there exists a partition* $\overline{[K]}, \underline{[K]} \subseteq [K]$*, i.e.,* $\overline{[K]} \cap \underline{[K]} = \emptyset$*,* $\overline{[K]} \cup \underline{[K]} = [K]$ *such that for all* $k \in \overline{[K]}$

$$\sum_{j \in [N]} a_{ij}^k x_j \leq b_i^k \quad \forall i \in [M], \boldsymbol{x} \in \mathcal{X} \tag{22}$$

and for all $k \in \underline{[K]}$

$$\sum_{j \in [N]} a_{ij}^k x_j > b_i^k \quad \forall i \in [M], \boldsymbol{x} \in \mathcal{X}. \tag{23}$$

For a given $\boldsymbol{\alpha} = (\overline{\boldsymbol{\alpha}}, \underline{\boldsymbol{\alpha}})$*, define the mapping* $\xi : \mathcal{S} \mapsto \mathcal{W}$ *as* $\xi(\boldsymbol{\alpha}) = (\xi_1, \ldots, \xi_{[K]})$*, with*

$$\xi_k = \overline{\alpha}_k \text{ if } \mathrm{k} \in \overline{[\mathrm{K}]}, \; \xi_\mathrm{k} = \underline{\alpha}_\mathrm{k} \text{ if } \mathrm{k} \in \underline{[\mathrm{K}]}. \tag{24}$$

Then, $\xi(\boldsymbol{\alpha})$ *is a decision-independent worst case outcome (see Definition 2).*
Proof: See Appendix A.5.

Speaking intuitively, the conditions in (22) and (23) arise if each uncertain parameter, $\tilde{z}_k$, $k \in [K]$ always has the same direction of influence in a manner that reduces the feasibility of the problem's constraints, regardless of the solution $\boldsymbol{x} \in \mathcal{X}$. For instance, a constraint requiring customer demands to be fulfilled always becomes more stringent with increasing levels of customer demands.

Applying the decision-independent worst-case outcome, the T-model formulation (21) can then be straightforwardly reduced to the following.

$$\begin{aligned} \max \quad & \sum_{k \in \overline{[K]}} \ln \mathbb{P}\left[\tilde{z}_k \leq \overline{\alpha}_k\right] + \sum_{k \in \underline{[K]}} \ln \mathbb{P}\left[\tilde{z}_k \geq \underline{\alpha}_k\right] \\ \text{s.t.} \quad & \sum_{j \in [N]} \left(a_{ij}^0 + \sum_{k \in \underline{[K]}} a_{ij}^k \underline{\alpha}_k + \sum_{k \in \overline{[K]}} a_{ij}^k \overline{\alpha}_k \right) x_j \geq b_i^0 + \sum_{k \in \underline{[K]}} b_i^k \underline{\alpha}_k + \sum_{k \in \overline{[K]}} b_i^k \overline{\alpha}_k \;\; \forall i \in [M], \\ & \boldsymbol{x} \in \mathcal{X}, \;\; \underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}} \in \mathcal{W}. \end{aligned} \tag{25}$$

Nevertheless, there are still some computational challenges of the T-model. Specifically,

1. the objective function is not necessarily concave;
2. the terms $x_j \underline{\alpha}_k$ and $x_j \overline{\alpha}_k \;\; j \in [N], k \in [K]$ are bilinear.

We next look at some useful cases that enable the T-model to be solved via general purpose MIP solvers such as CPLEX or Gurobi.

3.1. Case of log concave densities

For certain classes of continuously distributed random variables the T-model objective function in (20) remains concave in $(\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}})$. This is the case, for instance, for the class of random variables with *log-concave* densities, which include commonly-used distributions such as the exponential, uniform and normal distributions. Indeed, this is a consequence of the well known results of Prékopa (1980), from which we obtain the following corollary.

Corollary 1. *(See Prékopa 1980, Theorem 9) Suppose $\tilde{z}_k$ is a continuously distributed random variable with log-concave density function $f_k(z): \mathcal{W}_k \mapsto \mathbb{R}_+$. Then the function $F_k(\underline{\delta}, \overline{\delta}): \mathcal{D}_k \mapsto \mathbb{R}$,*

$$F_k(\underline{\delta}, \overline{\delta}) = \ln \mathbb{P}[\underline{\delta} \le \tilde{z}_k \le \overline{\delta}]$$

is a concave function of $(\underline{\delta}, \overline{\delta})$ on domain $\mathcal{D}_k = \left\{(\underline{\delta}, \overline{\delta}) \in \mathcal{W}_k^2 \ : \ \underline{\delta} < \overline{\delta}\right\}$.

Notwithstanding the fact that the constraint functions are bilinear, there are practical situations where Problem (21) will become tractable. Most notably, this is the case when the uncertainty occurs only at the right-hand side, so that $a_{ij}^k = 0, \ \forall i \in [M], j \in [N], k \in [K]$. Another tractable situation arises when x_j is discrete, in which case the bilinear terms $x_j \underline{\alpha}_k$ and $x_j \overline{\alpha}_k$ can be linearized using standard mixed integer programming techniques. For instance, if $x_j \in \{0, 1\}$, the bilinear term $x_j \underline{\alpha}_k$ can be replaced by a new decision variable r_{jk} that satisfies the following linear inequalities:

$$-\Theta(1 - x_j) \le r_{jk} - \underline{\alpha}_k \le \Theta(1 - x_j), \text{ and } \ -\Theta x_j \le r_{jk} \le \Theta x_j,$$

for a sufficiently large constant Θ.

Although the objective function in (20) is nonlinear, an important advantage of maximizing a concave objective function is that it can be solved efficiently in practice via piecewise linear approximations of arbitrary accuracy. We refer the readers to Appendix C for more details on formulating such approximations using cuts in a Branch-and-Cut fashion to facilitate the solution of the corresponding T-models.

3.2. Case of discrete distributions

We now focus on T-models based on random variables with discrete distributions. In particular, we model $\tilde{z}_k$ on the discrete support $\mathcal{W}_k = \left\{\zeta_k^1, \zeta_k^2, \ldots, \zeta_k^{L(k)}\right\}$, with strictly positive probability mass functions $\mathbb{P}[\tilde{z}_k = \zeta_k^\ell] = p_k^\ell, \ \forall \ell \in [L(k)]$. We also define $\lambda_k^0 = 0$ and $\lambda_k^\ell = \sum_{\ell' \in [\ell]} p_k^{\ell'}, \ \forall k \in [K], \ell \in [L(k)]$. Without loss of generality, we assume that the outcomes ζ_k^ℓ are ranked in nondecreasing values in $\ell = 1, \ldots, L(k)$. A specification of the T-model is developed as follows.

First, we define the adjustable uncertainty set parameters $\boldsymbol{\alpha}$ in (6). Let $\overline{\boldsymbol{\alpha}}_k = (\overline{\alpha}_k^1, \ldots, \overline{\alpha}_k^{L(k)})$, $\underline{\boldsymbol{\alpha}}_k = (\underline{\alpha}_k^1, \ldots, \underline{\alpha}_k^{L(k)})$, $\overline{\boldsymbol{\alpha}} = (\overline{\boldsymbol{\alpha}}_1, \ldots, \overline{\boldsymbol{\alpha}}_K)$, $\underline{\boldsymbol{\alpha}} = (\underline{\boldsymbol{\alpha}}_1, \ldots, \underline{\boldsymbol{\alpha}}_K)$, and $\boldsymbol{\alpha} = (\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}})$. Hence, we have $Q =$

$2\times\sum_{k\in[K]}L(k)$. In the following, $\underline{\alpha}_k^\ell$ and $\overline{\alpha}_k^\ell$ are modeled as binary variables that take value 1 if ζ_k^ℓ are the selected lower and upper bounds, respectively, of the interval that defines the discrete outcomes in the adjustable uncertainty set for random variable $\tilde{z}_k$, and 0 otherwise. We define:

$$\mathcal{S}=\left\{\boldsymbol{\alpha}\in\{0,1\}^Q \;\middle|\; \sum_{\ell\in[L(k)]}\overline{\alpha}_k^\ell=1,\;\; \sum_{\ell\in[L(k)]}\underline{\alpha}_k^\ell=1,\;\; \sum_{\ell\in[L(k)]}\ell(\overline{\alpha}_k^\ell-\underline{\alpha}_k^\ell)\geq 0\;\forall k\in[K]\right\}, \tag{26}$$

where $\mathcal{S}$ refers to the domain of the $\boldsymbol{\alpha}$ variables in the S-model (6). The intuition for the choice of $\boldsymbol{\alpha}$ above is to enable a partial ordering on the adjustable uncertainty sets $\mathcal{U}(\boldsymbol{\alpha})$ based on 'counting' the total number of outcomes ζ_k^ℓ (and weighted by respective probabilities p_k^ℓ) contained in intervals indicated by $\underline{\boldsymbol{\alpha}}_k$ and $\overline{\boldsymbol{\alpha}}_k$. The inequality $\sum_{\ell\in[L(k)]}\ell(\overline{\alpha}_k^\ell-\underline{\alpha}_k^\ell)\geq 0$, together with the fact that ζ_k^ℓ is a nondecreasing sequence in ℓ, ensures that the outcome values corresponding to $\overline{\boldsymbol{\alpha}}_k$ are at least as large as those of $\underline{\boldsymbol{\alpha}}_k$. A simple example is as follows. Suppose $L(k)=3$ for some k, so that $\overline{\boldsymbol{\alpha}}_k=(\overline{\alpha}_k^1,\overline{\alpha}_k^2,\overline{\alpha}_k^3)$, and $\underline{\boldsymbol{\alpha}}_k=(\underline{\alpha}_k^1,\underline{\alpha}_k^2,\underline{\alpha}_k^3)$. Assume an instance with $\overline{\alpha}_k^1=1$, $\overline{\alpha}_k^2=0$, $\overline{\alpha}_k^3=0$, and $\underline{\alpha}_k^1=0$, $\underline{\alpha}_k^2=1$, $\underline{\alpha}_k^3=0$. This satisfies the first two constraints in (26), but does not satisfy $\sum_{\ell\in[L(k)]}\ell(\overline{\alpha}_k^\ell-\underline{\alpha}_k^\ell)\geq 0$, since $1(1-0)+2(0-1)+3(0-0)=-1<0$. In this case, assuming that we fix $\underline{\alpha}_k^2=1$, the only feasible $\boldsymbol{\alpha}$ implies $\overline{\alpha}_k^2=1$ or $\overline{\alpha}_k^3=1$.

Next, we define the adjustable uncertainty sets:

$$\mathcal{U}(\boldsymbol{\alpha})=\left\{\boldsymbol{z}\in\mathcal{W}\;\middle|\;\sum_{\ell\in[L(k)]}\zeta_k^\ell\underline{\alpha}_k^\ell\leq z_k\leq\sum_{\ell\in[L(k)]}\zeta_k^\ell\overline{\alpha}_k^\ell,\;\forall k\in[K]\right\}. \tag{27}$$

It can be observed that the sets $\mathcal{U}(\boldsymbol{\alpha})$ as defined above are entirely analogous to the adjustable uncertainty sets of the previous case in Section 3.1 using continuous intervals, with the exception that here $\mathcal{U}(\boldsymbol{\alpha})$ is finite and countable.

Combining the definitions of $\mathcal{S}$ in (26) and of $\mathcal{U}(\boldsymbol{\alpha})$ in (27), the resulting T-model can then be written as:

$$\begin{array}{ll}\max & \ln\mathbb{P}[\tilde{\boldsymbol{z}}\in\mathcal{U}(\boldsymbol{\alpha})]\\ \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}\geq\boldsymbol{b}(\boldsymbol{z})\quad\forall\boldsymbol{z}\in\mathcal{U}(\boldsymbol{\alpha})\\ & \displaystyle\sum_{\ell\in[L(k)]}\underline{\alpha}_k^\ell=1,\quad\sum_{\ell\in[L(k)]}\overline{\alpha}_k^\ell=1,\qquad\qquad \forall k\in[K]\\ & \displaystyle\sum_{\ell\in[L(k)]}\ell(\overline{\alpha}_k^\ell-\underline{\alpha}_k^\ell)\geq 0\qquad\qquad \forall k\in[K]\\ & \underline{\boldsymbol{\alpha}}_k,\;\overline{\boldsymbol{\alpha}}_k\in\{0,1\}^{L(k)}\;\;\forall k\in[K],\quad\boldsymbol{x}\in\mathcal{X}.\end{array} \tag{28}$$

The following result provides the robust counterpart model to (28), which can then be further reformulated as a linear MIP model.

Theorem 4. *Under discrete distributions, the T-model (28), based on assumptions A1 and A2, is equivalent to the following reformulation:*

$$
\begin{aligned}
\max\ & \sum_{k\in[K]} s_k \\
\text{s.t.}\ & s_k \le \ln(\gamma) - 1 + \sum_{\ell\in[L(k)]} \frac{1}{\gamma}\left(\lambda_k^{\ell}\overline{\alpha}_k^{\ell} - \lambda_k^{\ell-1}\underline{\alpha}_k^{\ell}\right) && \forall \gamma\in\mathcal{C}_k, k\in[K] \\
& \sum_{j\in[N]} a_{ij}^0 x_j + \sum_{k\in[K]} v_{ik} \ge b_i^0 && \forall i\in[M] \\
& v_{ik} \le \sum_{\ell\in[L(k)]}\sum_{j\in[N]} \left(a_{ij}^k x_j - b_i^k\right)\zeta_k^{\ell}\overline{\alpha}_k^{\ell} && \forall i\in[M],\ k\in[K] \\
& v_{ik} \le \sum_{\ell\in[L(k)]}\sum_{j\in[N]} \left(a_{ij}^k x_j - b_i^k\right)\zeta_k^{\ell}\underline{\alpha}_k^{\ell} && \forall i\in[M],\ k\in[K] \\
& \sum_{\ell\in[L(k)]} \overline{\alpha}_k^{\ell} = 1,\quad \sum_{\ell\in[L(k)]} \underline{\alpha}_k^{\ell} = 1 && \forall k\in[K] \\
& \sum_{\ell\in[L(k)]} \ell(\overline{\alpha}_k^{\ell} - \underline{\alpha}_k^{\ell}) \ge 0 && \forall k\in[K] \\
& \boldsymbol{x}\in\mathcal{X},\ \boldsymbol{v}\in\mathbb{R}^{M\times K},\ \boldsymbol{s}\in\mathbb{R}^K \\
& \underline{\boldsymbol{\alpha}}_k,\ \overline{\boldsymbol{\alpha}}_k \in\{0,1\}^{L(k)}\ \ \forall k\in[K],
\end{aligned}
\tag{29}
$$

where

$$
\mathcal{C}_k = \left\{\lambda_k^{\overline{\ell}} - \lambda_k^{\underline{\ell}-1}\ \middle|\ \underline{\ell},\overline{\ell}\in[L(k)], \overline{\ell}\ge\underline{\ell}\right\}\quad \forall k\in[K].
$$

Proof: See Appendix A.6.

Formulation (29) contains bilinear terms $x_j\underline{\alpha}_k^{\ell}$ and $x_j\overline{\alpha}_k^{\ell}$, which can further be linearized as in earlier cases. The resulting problem is then a linear MIP and can be solved by general purpose solvers. Although $|\mathcal{C}_k|$ is at most $\frac{1}{2}L(k)(L(k)+1)$, it may still be impractical to introduce the entire first set of constraints of Problem (29). Nevertheless, as discussed earlier, we can also introduce these constraints as cuts and solve the MIP in a Branch-and-Cut fashion.

Finally, if the T-model also has decision-independent worst-case uncertainty, it can further be simplified as follows.

Corollary 2. *Under discrete distributions, if the T-model (28) has decision-independent worst case uncertainty (see Definition 2), then it has the following formulation:*

$$\begin{aligned}
\max \; & \sum_{k\in\overline{[K]}}\sum_{\ell\in[L(k)]} \ln\left(\lambda_k^\ell\right)\overline{\alpha}_k^\ell + \sum_{k\in\underline{[K]}}\sum_{\ell\in[L(k)]} \ln\left(1-\lambda_k^{\ell-1}\right)\underline{\alpha}_k^\ell \\
\text{s.t.} \; & \sum_{j\in[N]}\left(a_{ij}^0 + \sum_{k\in\overline{[K]}}\sum_{\ell\in[L(k)]} a_{ij}^k\zeta_k^\ell\overline{\alpha}_k^\ell + \sum_{k\in\underline{[K]}}\sum_{\ell\in[L(k)]} a_{ij}^k\zeta_k^\ell\underline{\alpha}_k^\ell\right)x_j \\
& \qquad \geq b_i^0 + \sum_{k\in\overline{[K]}}\sum_{\ell\in[L(k)]} b_i^k\zeta_k^\ell\overline{\alpha}_k^\ell + \sum_{k\in\underline{[K]}}\sum_{\ell\in[L(k)]} b_i^k\zeta_k^\ell\underline{\alpha}_k^\ell \;\; \forall i\in[M] \\
& \sum_{\ell\in[L(k)]}\overline{\alpha}_k^\ell = 1, \quad \sum_{\ell\in[L(k)]}\underline{\alpha}_k^\ell = 1 \qquad \forall k\in[K] \\
& \underline{\boldsymbol{\alpha}}_k,\; \overline{\boldsymbol{\alpha}}_k \in \{0,1\}^{L(k)} \;\; \forall k\in[K], \quad \boldsymbol{x}\in\mathcal{X}.
\end{aligned} \tag{30}$$

Proof: The above follows from a straightforward application of formulation (25) for the T-model with decision-independent worst-case uncertainty. □

3.3. On the T-model and SAA approximations to P-models

The T-model in (20) is reproduced here in a more generic format:

$$\begin{aligned}
\max \; & \sum_{k\in[K]} \ln \mathbb{P}\left[\underline{\alpha}_k \leq \tilde{z}_k \leq \bar{\alpha}_k\right] \\
\text{s.t.} \; & [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}] \subseteq \mathcal{T}(\boldsymbol{x}) \\
& \boldsymbol{x}\in\mathcal{X} \\
& \underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}} \in \mathbb{R}^K,
\end{aligned}$$

which maximizes the volume of the hyper-rectangle set $[\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}]$ that can be contained within the tolerance set, $\mathcal{T}(\boldsymbol{x})$. Owing to the assumptions of independent factors, $\tilde{z}_k$, $k\in[K]$, the objective function, that is related to the volume of the hyper-rectangle, can be easily computed. It is clear that the T-model can be viewed as an approximation of the P-model (1) which avoids intractable multidimensional integration operations of the tolerance set. Obviously, if $\mathcal{T}(\boldsymbol{x})$ are also hyper-rectangular, for all $\boldsymbol{x}\in\mathcal{X}$, then the T-model is exactly the P-model (e.g. see Theorem 2 for the case of linear feasibility constraints).

However, it can well be the case that the probability of feasibility $\mathbb{P}[\tilde{\boldsymbol{z}}\in\mathcal{T}(\boldsymbol{x})]$, as used in the P-model, is significantly higher than $\mathbb{P}[\tilde{\boldsymbol{z}}\in\mathcal{U}(\boldsymbol{\alpha})]$, as used in the probabilistic S-model objective (see, e.g., Ben-Tal et al. 2009), even when $\boldsymbol{\alpha}$ is optimal in (18) for the given $\boldsymbol{x}$. Consider an example of the feasibility probability of a single constraint without decision variables as follows:

$$\mathbb{P}\left[3\sqrt{K} \geq \sum_{k\in[K]} \tilde{z}_k\right], \tag{31}$$

Assume for simplicity that $\tilde{z}_k$'s are independent and standard normal-distributed so that the probability would be $\Phi(3) \approx 0.9987$, where $\Phi(\cdot)$ denotes the standard normal cdf. Applying the T-model then produces the following:

$$\mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{U}(\boldsymbol{\alpha}^*)] = \mathbb{P}\left[\tilde{\boldsymbol{z}} \le \boldsymbol{\alpha}^*\right],$$

where

$$\boldsymbol{\alpha}^* = \arg\max_{\boldsymbol{\alpha}} \left\{ \sum_{k\in[K]} \ln \mathbb{P}\left[\tilde{z}_k \le \alpha_k\right] \;\middle|\; \sum_{k\in[K]} \alpha_k \le 3\sqrt{K} \right\}$$

and $\ln \mathbb{P}\left[\tilde{z}_k \le \alpha_k\right] = \ln \Phi(\alpha_k)$. Because of the log concavity of the normal distribution, we have $\alpha_k^* = \frac{3}{\sqrt{K}}$ for all $k \in K$, and consequently,

$$\mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{U}(\boldsymbol{\alpha}^*)] = \Phi\left(\frac{3}{\sqrt{K}}\right)^K,$$

which diminishes quickly to zero as K increases. For $K = 15$, the probability evaluated in the T-model would be 0.0045, although the actual feasibility of the constraint remains at 0.9987 regardless of K.

Nevertheless, despite the weakness in the bound, the quality of the approximation depends on the geometry of the tolerance set, and on how well it can be stretched maximally from within by a hyper-rectangular set. Take for example the case with right side uncertainty on a single constraint,

$$\mathbb{P}\left[\boldsymbol{a}'\boldsymbol{x} \ge \sum_{k\in[K]} \tilde{z}_k\right],$$

for which the optimal solution for the T-model coincides with the P-model, when $\boldsymbol{a}'\boldsymbol{x}$ is maximized over $\boldsymbol{x} \in \mathcal{X}$.

Another popular approach to approximate the P-model is to use sample average approximation (SAA). For the SAA approach, we assume that a set of L independent samples $\boldsymbol{z}_1, \ldots \boldsymbol{z}_L, \in \mathbb{R}^K$ is available. The SAA model can be written as the following optimization problem:

$$\begin{array}{ll} \max & \ln \dfrac{|\mathcal{S}|}{L} \\ \text{s.t.} & \mathrm{CH}\left(\{\boldsymbol{z}_l\}_{l\in\mathcal{S}}\right) \subseteq \mathcal{T}(\boldsymbol{x}) \\ & \boldsymbol{x} \in \mathcal{X} \\ & \mathcal{S} \subseteq [L], \end{array}$$

where $\mathrm{CH}(\mathcal{W})$ denotes the convex hull of $\mathcal{W}$, assuming that the tolerance set is convex for any given $\boldsymbol{x} \in \mathcal{X}$. Hence, the SAA in the above format can also be viewed as an S-model, which essentially maximizes the number of samples that can be packed within the tolerance set. The SAA has the benefits of being more generic and not being confined to independent random factors as assumed in our T-model. However, since SAA is not a deterministic model, the quality of the approximation depends on the sample size L. While SAA solutions are expected to improve with increasing L, in practice, they may not scale computationally well, in particular when the optimization problems are of combinatorial nature. Even if all the samples are feasible in the SAA model, the probability of $\tilde{\boldsymbol{z}}$ being in the convex hull of the SAA samples may not scale well with K. For instance, if $\tilde{z}_k$, $k \in [K]$ are i.i.d. discrete random variables, each taking values in $\{0, 1\}$ with equal probability, then we have:

$$\mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathrm{CH}\left(\{\boldsymbol{z}_l\}_{l\in\mathcal{S}}\right)\right] \leq L/2^K.$$

We numerically evaluate the probability value for the case of i.i.d. uniformly distributed random variables, each taking values in $[0, 1]$. Table 1 shows the probability $\mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathrm{CH}\left(\{\tilde{\boldsymbol{z}}_l\}_{l\in[L]}\right)\right]$ evaluated based on 10,000 randomly generated instances. For each instance, we solve a linear optimization problem to determine its feasibility and we take the average number of feasible instances to determine each entry of the table. Similar to the T-model, these probabilities may grossly under estimate

K	$L = 100$	$L = 200$	$L = 400$	$L = 800$
1	0.981	0.989	0.995	0.997
3	0.689	0.791	0.869	0.917
5	0.269	0.405	0.535	0.644
7	0.056	0.124	0.216	0.326
9	0.010	0.026	0.065	0.114
15	<0.0001	<0.0001	0.0003	0.0009

Table 1 Feasibility probability $\mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathrm{CH}\left(\{\tilde{\boldsymbol{z}}_l\}_{l\in[L]}\right)\right]$, for $\tilde{\boldsymbol{z}}$ and each $\tilde{\boldsymbol{z}}_l$ being a vector of K i.i.d. uniformly distributed random variables, based on 10,000 randomly generated samples.

the actual feasibility probability of the tolerance set, which is more complex to analyze precisely. To simplify the analysis, we consider an uncertain linear optimization problem with only right hand side uncertainty. Suppose all the L samples are feasible, then

$$[\boldsymbol{A}\tilde{\boldsymbol{x}}]_i \geq \max_{l\in[L]} \left\{[\boldsymbol{b}(\tilde{\boldsymbol{z}}_l)]_i\right\} \qquad \forall i \in [M], \tag{32}$$

noting that the optimal SAA solution $\tilde{\boldsymbol{x}}$ is random because of the dependency on the L samples. Hence, by Bonferroni inequality, we have

$$\begin{aligned}
\mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{T}(\tilde{\boldsymbol{x}})\right] &= \mathbb{P}\left[\boldsymbol{A}\tilde{\boldsymbol{x}} \geq \boldsymbol{b}(\tilde{\boldsymbol{z}})\right] \\
&\geq \mathbb{P}\left[\max_{l\in[L]} \left\{[\boldsymbol{b}(\tilde{\boldsymbol{z}}_l)]_i\right\} \geq [\boldsymbol{b}(\tilde{\boldsymbol{z}})]_i \quad \forall i \in [M]\right] \\
&\geq 1 - \sum_{i\in M} \mathbb{P}\left[\max_{l\in[L]} \left\{[\boldsymbol{b}(\tilde{\boldsymbol{z}}_l)]_i\right\} < [\boldsymbol{b}(\tilde{\boldsymbol{z}})]_i\right] \\
&\geq 1 - \frac{M}{L+1} = \frac{L+1-M}{L+1}.
\end{aligned} \tag{33}$$

If the right hand side uncertainties among different constraints are independent, then the bound would be

$$\mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{T}(\tilde{\boldsymbol{x}})\right] \geq \left(\frac{L}{L+1}\right)^M,$$

which can be tight if the constraints (32) are tight and $\tilde{\boldsymbol{z}}$ are continuously distributed. For instance, if $M = 500$ and the computational limit permits at most $L = 1000$, then using the SAA approach, the maximum assurance probability could be at most 50%, or 61% if independence among constraints can be assumed. The above suggests that for an optimization problem with a large number of uncertain parameters and constraints, one may require the number of samples to be substantially larger as well. Indeed, this is consistent with the observations from our numerical studies (see Section 4), in which for a given L, the performance of the SAA degrades significantly as M increases, while it does not seem to have the same impact on the T-model.

4. Computational Studies on Stochastic Maximum Coverage Problems

In this section, we illustrate the advantages of the T-model by means of computational experiments on facility location problems under uncertainty. Specifically, we consider the *stochastic maximum coverage problem* (SMCP), in which a set of facilities has to be selected such that the probability that all uncertain customer demands can be met is maximized.

4.1. Definition, formulations, and experimental setup

We consider a single-stage SMCP, in which selection of facilities and demand allocation are performed before the customer demands are observed. The notation for the SMCP is as follows. There are I number of candidate facility locations and J customer locations. The subset $\mathcal{I}_j \subseteq [I]$ are facilities that can serve customer j, and $\mathcal{J}_i \subseteq [J]$ is the subset of customers that can be served by facility i. Let c_i and f_i be the capacity and construction costs, respectively, for facility $i \in [I]$ and β the total budget available for facility constructions. The demand of each customer $j \in [J]$ is modelled as an affine function of K independently distributed random factors $\tilde{z}_1, \ldots, \tilde{z}_K$ given by

$$b_j(\tilde{\boldsymbol{z}}) = b_j^0 + \sum_{k \in [K]} b_j^k \tilde{z}_k. \tag{34}$$

We compare the performance of three models: T-model, P-model and an expected demand shortfall model. To be fair in the comparison, the solutions to these models are obtained using the same empirical distribution of the random factors associated with the demands. Specifically, the empirical distribution of the random factors are given by the samples $\hat{\boldsymbol{\zeta}}^\ell \in \mathbb{R}^K$, $\ell \in [L]$, each with equal probability of realization. Subsequently, the quality of the solutions are evaluated using out-of-sample tests based on the true underlying distribution of the random factors.

Details for modelling the customer demands are found in Appendix E. The test cases used in this section do not assume decision-independent worst-case uncertainty (see Definition 2 in Section 2.3). We refer readers to Appendix F for numerical studies where we consider a two-stage version of the problem for which decision-independent worst-case uncertainty has to be assumed.

P-model for the SMCP. Let binary variables x_i be 1 if facility i has been selected, and 0 otherwise. In the single-stage problem SMCP, the demand allocation from facility i to customer j denoted by y_{ij} has to be decided before $\tilde{\boldsymbol{z}}$ is known. The formulation for the corresponding P-model is as follows.

$$\begin{aligned}
\max\ & \ln \mathbb{P}\left[\sum_{i\in\mathcal{I}_j} y_{ij} \geq b_j(\tilde{\boldsymbol{z}}) \quad \forall j \in [J]\right] & \\
\text{s.t.}\ & \sum_{j\in\mathcal{J}_i} y_{ij} \leq c_i x_i & \forall \boldsymbol{z}\in\mathcal{W}, i\in[I] \\
& \sum_{i\in[I]} f_i x_i \leq \beta & \\
& y_{ij} \geq 0 & \forall i\in[I], j\in\mathcal{J}_i \\
& x_i \in \{0,1\} & \forall i \in [I],
\end{aligned} \tag{35}$$

where the objective function maximizes the log-probability that the uncertain demands can be satisfied. The first set of constraints are the facility capacity constraints, while the second constraint restricts the costs of opening facilities to the available budget β.

In the computational exercises, we solve an SAA model of (35), defined as P-1, maximizing the probability of satisfying all demands using the empirical probability distribution. We further consider a second SAA model defined as E-1 that aims at minimizing the expected level of demand shortfall using the empirical distribution. A computational advantage of model E-1 over P-1 is that it avoids the use of additional binary variables for modeling the probability function in the objective. The explicit formulations for the SAA models are provided in Appendix D.2.

T-model for the SMCP. In the numerical studies we implement the T-model based on formulation (21) using the empirical distribution of the random factors, which is for the case of discrete distributions of $\tilde{\boldsymbol{z}}$ (see § 3.2). Note that in general, the random demand factors $\tilde{z}_k$, $\forall k \in [K]$ may impact customer demands in either direction.

The T-model for the single-stage SMCP based on Theorem 4, which is termed here as Model T-1, can then be stated as:

$$
\begin{aligned}
\max \quad & \sum_{k\in[K]} s_k \\
\text{s.t.} \quad & s_k \le \ln\left(\frac{\gamma}{L}\right) - 1 + \frac{1}{\gamma}\sum_{\ell\in[L]}\left(\ell\overline{\alpha}_k^\ell - (\ell-1)\underline{\alpha}_k^\ell\right) && \forall \gamma\in[L], k\in[K] \\
& \sum_{i\in\mathcal{I}_j} y_{ij} \ge b_j^0 + \sum_{k\in[K]:b_j^k<0}\sum_{\ell\in[L]} b_j^k\zeta_k^\ell\underline{\alpha}_k^\ell + \sum_{k\in[K]:b_j^k>0}\sum_{\ell\in[L]} b_j^k\zeta_k^\ell\overline{\alpha}_k^\ell && \forall j\in[J] \\
& \sum_{j\in\mathcal{J}_i} y_{ij} \le c_i x_i && \forall i\in[I] \\
& \sum_{i\in[I]} f_i x_i \le \beta \\
& \sum_{\ell\in[L]} \underline{\alpha}_k^\ell = 1 && \forall k\in[K] \\
& \sum_{\ell\in[L]} \overline{\alpha}_k^\ell = 1 && \forall k\in[K] \\
& \sum_{\ell\in[L]} \ell(\overline{\alpha}_k^\ell - \underline{\alpha}_k^\ell) \ge 0 && \forall k\in[K] \\
& \boldsymbol{s}\in\mathbb{R}^K \\
& y_{ij}\ge 0 && \forall i\in[I],\ j\in\mathcal{J}_i \\
& x_i\in\{0,1\} && \forall i\in[I] \\
& \underline{\alpha}_k^\ell, \overline{\alpha}_k^\ell \in\{0,1\} && \forall \ell\in[L],\ k\in[K],
\end{aligned}
\tag{Model T-1}
$$

where $(\zeta_k^1,\ldots,\zeta_k^L)$ denotes the ordered statistics of the empirical samples $(\hat{\zeta}_k^1,\ldots,\hat{\zeta}_k^L)$ for each factor $k\in[K]$.

In the numerical studies, all mathematical models have been implemented in C/C++ using the IBM CPLEX 12.8.0.0 Callable Library. All optimization problems have been solved to an optimality of 1%, since we found that proving optimality took unnecessarily long computing times without improving the solution quality. The code has been compiled and executed on openSUSE 11.3. Each problem instance has been run on a single Intel E5-2683 v4 processor (2.1GHz), limited to 24GB of RAM and a maximum of 12 hours computing time. Out-of-sample tests were performed based on simulating 100,000 independent random demand samples. We refer readers to Appendix E for a detailed description of the settings for the computational experiments.

4.2. Computational and solution performance

Computational efficiency. Table 2 compares the computational efficiency of the three models for the single-stage SMCP. Each row corresponds to the problem instances with J customers and presents average results over 12 instances (3 different numbers of candidate facility locations, 2 different levels of customer-facility connectivity and 2 levels of demand correlation). The results are based on solving the models using an empirical distribution with different sample sizes L. The average computing times are based on all instances with the same number of customers J for which the respective model has found a feasible solution. The column "# no" reports the number of instances which have not been solved to optimality using the respective model, whereas column "# ns" refers to the number of instances for which the solver has not found any feasible solution.

		Model T-1			Model P-1			Model E-1		
J	L	time (min.)	# no	# ns	time (min.)	# no	# ns	time (min.)	# no	# ns
100	50	0.0	0	0	0.0	0	0	0.0	0	0
	250	0.1	0	0	0.0	0	0	0.1	0	0
	500	0.2	0	0	0.7	0	0	0.1	0	0
	1000	1.3	0	0	121.1	0	0	0.6	0	0
250	50	0.1	0	0	0.1	0	0	0.1	0	0
	250	0.3	0	0	0.4	0	0	0.3	0	0
	500	1.4	0	0	6.8	0	0	2.0	0	0
	1000	2.8	0	0	361.8	0	0	2.6	0	0
500	50	1.4	0	0	0.3	0	0	0.3	0	0
	250	0.9	0	0	1.5	0	0	1.2	0	0
	500	20.3	0	0	362.7	0	0	9.6	0	0
	1000	13.0	0	0	187.8	4	4	13.0	0	0

Table 2 Computational efficiency for different instance sizes J and sample sizes L for the single-stage SMCP. Average computing time (in minutes) are for the instances where the solver has found a feasible solution. Column '# no' refers to the number of instances for which the solver has not proven optimality, whereas column '# ns' represents the number of instances for which the solver did not find any feasible solution.

The results suggest that for smaller or moderate sized problems, the solution times of all the models appear reasonably on-par (note that these average times do not include instances which have not been solved). When sample sizes used are large (i.e., $L = 1000$), Model P-1 required significantly longer solution times. Furthermore, some instances of Model P-1 could not be solved within the time limit of 12 hours (or even find a feasible solution). In contrast, all instances of model T-1 and E-1 are solved.

Solution performance. We next investigate how well the models perform in terms of success rate and demand shortfall as shown in Table 3. Here, only test instances that have been solved by all models for empirical sample sizes L (50, 250 and 500) are considered. In other words, all models have been evaluated on the same instances for a fair comparison. Also, two versions of the T-model are tested. The first version (Model T-1), is assumed to observe the random factors $\tilde{\boldsymbol{z}}$ used in the affine factor demand model (34), and exploits the fact that $\tilde{\boldsymbol{z}}$ are independently distributed. The second version (Model T†-1), observes only the customer demands $b_j(\tilde{\boldsymbol{z}})$ for all $j \in [J]$, and wrongly assumes that the demands are independently distributed.

First, from the results, all the models benefit from larger sample sizes L, leading to solutions with higher success rates and lower shortfalls in general. However, it can be observed that the solution performance achieved by Model T-1 is consistently better than Model T†-1, P-1 and E-1 across almost all problem sizes in J and sample sizes L used.

			Model T-1		Model T†-1		Model P-1		Model E-1	
		#	succ.	short	succ.	short	succ.	short	succ.	short
J	L	inst	rate %	fall	rate %	fall	rate %	fall	rate %	fall
100	50	12	54.88	1.0	22.81	1.4	22.93	1.5	23.49	1.4
	250	12	85.31	0.1	73.80	0.2	73.61	0.2	73.63	0.2
	500	12	89.11	0.1	83.47	0.1	83.37	0.1	83.14	0.1
250	50	12	49.58	2.2	6.76	3.0	6.79	3.0	6.83	3.0
	250	12	69.93	0.3	46.29	0.5	46.09	0.5	46.38	0.5
	500	12	79.02	0.2	63.43	0.3	62.95	0.3	62.19	0.3
500	50	6	97.55	0.0	1.08	4.0	1.04	4.0	1.07	4.0
	250	6	97.90	0.0	29.48	0.7	29.33	0.7	29.43	0.7
	500	6	98.00	0.0	51.24	0.3	51.15	0.3	50.97	0.3

Table 3 Performance comparison of the models (including Model T†-1 that observes only final demand) for the single-stage SMCP for instances that have been solved by all models and all reported numbers of scenarios.

Note that the Model T†-1 assumes independent demands and can be expressed as the following factor-based model,

$$b_j(\tilde{\boldsymbol{z}}) = \tilde{z}_j \qquad \forall j \in [J],$$

albeit incorrectly compared to the Model T-1.

Observe that if all the scenarios are feasible, then the Model P-1 and E-1 would also be optimal, and in which case, the constraints with uncertain demands would also coincide with the T†-1 constraints as follows

$$\sum_{i \in \mathcal{I}_j} y_{ij} \geq \max_{\ell \in [L]} \left\{ \hat{\zeta}_j^\ell \right\} = \zeta_j^L \qquad \forall j \in [J].$$

Hence, as we would expect in Table 3, the performance of the Model T†-1 is similar to that of Models P-1 and E-1, since these models' constraints require solutions to be feasible in the largest demand sample for each customer $j \in [J]$. On the other hand, Model T-1 generally outperforms the other models by a significant margin. A possible reason is that the T-1 model is able to exploit the fact that the random factors $\tilde{\boldsymbol{z}}$ are independently distributed, while P-1 and E-1, by construction, cannot do so. Indeed, from the results for Model T†-1, it can be seen that when this information is not available, its performance advantage is significantly reduced. In the case for Model T-1, however, there is an additional opportunity to exploit the structure of the affine factor demand model to stretch the hyper-rectangle sets $[\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}]$ of the random factors as much as possible.

Furthermore, the results also suggest that the solution quality of Models P-1 and E-1 tend to be more sensitive to the changes in sample size L, compared to Model T-1. As discussed in Section 3.3, SAA approximations to the P-model can result in quite poor performance if sample sizes used are not sufficiently large. This could be a possible reason why the out-of-sample success rates achieved by Model P-1 (and also E-1) are quite low in Table 3 (particularly for the cases of $J = 500$). More specifically, we note that the sample size L used in Table 3 is rather small in comparison to the dimension of the problem solved (i.e. with respect to number of customers J, in which case the number of constraints $M = J$). For example, when $M = 250$ and $L = 500$, the Bonferroni bound on the probability of the feasibility in the demand constraints (i.e., applying (33)), is $\frac{500+1-250}{500+1} \approx 0.5$. This implies that the SAA model could yield solutions that achieve success probabilities as low as 50%, even if all L samples are feasible. The out-of-sample probability in Table 3 in this case is about 63%. When M increases to 500, the bound reduces to almost zero, and the out-of-sample probability observed in the results reduces to about 51%, while the T-model yields a solution that achieves more than 90% feasibility. To ensure this in the SAA model, the bound would suggest a sample size of $L \approx 5000$ or larger, which can become computationally prohibitive.

Scalability. To demonstrate the scalability of the models, Table 4 shows some further performance comparisons (including average, minimum and maximum of the average success rates over all 60 test instances, based on 10 replications) for the different models and increasing sample sizes L. Instances for which no feasible solution has been found were counted as zero success rates. For Models P-1 and E-1, only results for sample sizes $L = 250$ and $L = 500$ are available, since in our computational tests these models were unable to solve many instances of larger sample sizes. Model T-1 on the other hand, could solve instances up to $L = 5000$. For $L = 250$ and $L = 500$, the average results of Model T-1 outperform those of the SAA benchmarks at all instance sizes. Larger sample sizes used for Model T-1 also led to increasing success rates and lower sample-to-sample variation. In conclusion, the T-model, when implemented as a sample based approximation to the P-model, experiences much better scalability compared to the standard SAA models P-1 and E-1.

	Model P-1		Model E-1		Model T-1				
J	$L = 250$	$L = 500$	$L = 250$	$L = 500$	$L = 250$	$L = 500$	$L = 1000$	$L = 2500$	$L = 5000$
100	71.62	83.00	71.92	82.30	82.68	87.23	89.19	92.47	92.88
	[62.2, 79.3]	[77.2, 88.8]	[62.7, 79.9]	[75.6, 88.8]	[62.4, 99.0]	[0.0, 98.7]	[0.0, 98.5]	[86.5, 98.3]	[87.1, 98.4]
250	44.45	64.94	44.61	64.00	67.46	76.15	81.49	84.91	85.79
	[35.0, 56.5]	[54.5, 73.4]	[35.2, 56.7]	[51.4, 73.8]	[35.2, 98.1]	[0.0, 98.2]	[0.0, 98.3]	[71.4, 98.1]	[73.2, 98.2]
500	21.69	41.89	21.76	40.70	55.80	64.56	70.94	74.25	75.35
	[12.4, 33.9]	[30.9, 53.5]	[12.5, 34.2]	[27.6, 53.4]	[12.6, 98.2]	[0.0, 98.0]	[43.1, 98.2]	[50.2, 98.2]	[52.6, 98.0]

Table 4 Scalability study for single stage SMCP, comparing average [minimum, maximum] success rates (over all 60 instances) among 10 replications with different sample sizes L.

Impact of investment budget. We now repeat the computational studies by varying β, the available investment budget. This type of analysis can be useful for decision-makers in calibrating and justifying investment budget requests. Figure 1 shows the average success rates, as well as the intervals spanned by the minimum and maximum success rates across 10 optimization runs for a specific problem instance with 500 facilities, 250 customers and $A = 20$. In each run, a sample of size $L = 250$ demand scenarios was generated and used for Models T-1, P-1 and E-1. The investment

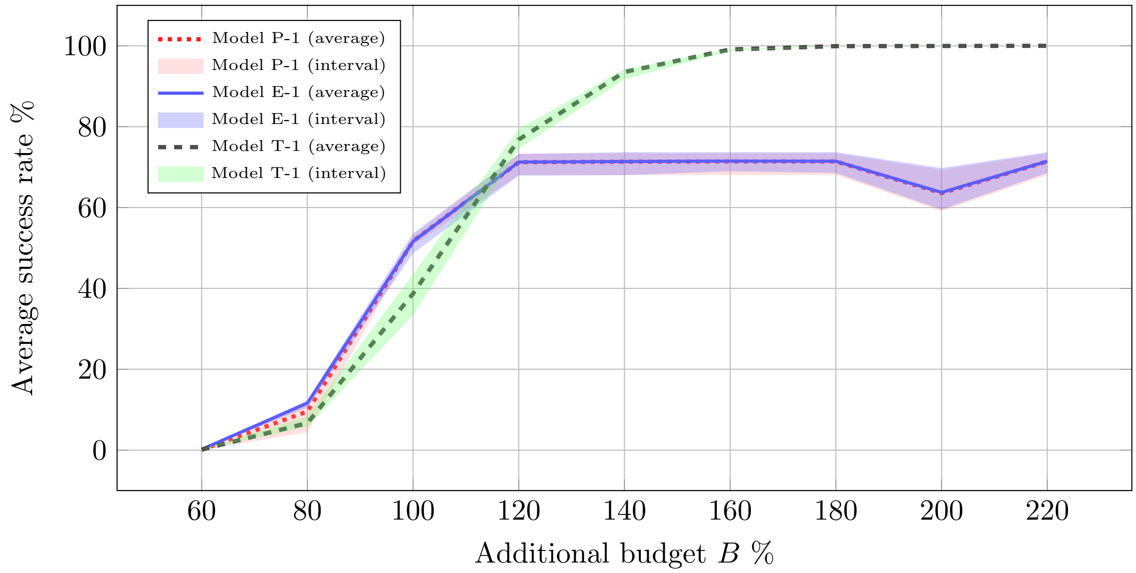

Figure 1 Success rates for Models T-1, P-1 and E-1. The graph illustrates the average success rates and the intervals between the minimum and maximum success rates across 10 replications (using sample size $L = 250$) for the single stage SMCP problem and an instance with 500 facilities, 250 customers and 10 random factors per customer.

budget β is varied by introducing percentage increments B of up to 220% of the minimum level assumed (that level of capacity budget required if all demands were to take on their average values).

In the solutions for all three models, the success rates continuously improve as more budget becomes available. Interestingly, both SAA benchmarks provide slightly higher success rates than the T-1 model with an additional budget of up to 100%. After that point, their success rates stagnate below 70% and do not improve even with high levels of β. Such performance of the P-1 and E-1 is likely due to the relatively small sample size L used in the numerical studies, in comparison to the dimension of the problem solved (e.g. in terms of the number of customer demands J, and consequently the number constraints $M = J$). For the case presented here with $M = L = 250$, the Bonferroni bound on the probability of the feasibility in the linear constraints (see (33)) is $\frac{250+1-250}{250+1} \approx 0.004$. This suggests that, even in the case when the budget is generous and the optimal SAA solution is feasible in all L samples, it is possible that the actual feasibility probability is very low. In contrast, the T-1 model makes effective use of the additional budget, converging to solutions that achieve nearly perfect feasibility in the random customer demand constraints. This suggests that the ability of the T-1 model to stretch its hyper-rectangle sets maximally is not significantly hindered by the sample size L used to calibrate the model.

The here proposed T-1 model, a tractable instance of the S-model, has shown to be an attractive alternative to classical SAA models. While the above experiments report on a single-stage stochastic problem, we note that experiments carried out on a two-stage stochastic variant of the problem have led to similar conclusions. For details, we refer the reader to Appendix F.

5. Conclusion

We have proposed a new satisficing decision criteria based on two key properties of satisficing decision making, and provided a functional representation of any such decision criterion. Based on this, we introduced a general formulation the S-model, in which the satisficing decision criterion is maximized. The P-model of Charnes and Cooper (1963) and other models based on robust optimization can be shown to be special cases of the S-model. We next introduced a class of tractable S-models that incorporates available probability distributions, termed the T-model, which, beyond its own merit, can also be regarded as a conservative and tractable approximation of the P-model. In the case when probability density functions of the uncertainties are log-concave, we propose a solution approach based on piecewise-linear concave approximations. In the case of discrete probability distributions, we provide a linear MIP formulation of the T-model. Finally, we presented computational studies on a single-stage and a two-stage stochastic maximum coverage problem, for which we have shown that the T-models outperform other benchmarks based on sampling average approximations both in terms of solution quality, and computational efficiency and scalability.

References

Ben-Tal, A., D. den Hertog and J.P. Vial. (2015). Deriving robust counterparts of nonlinear uncertain inequalities. Mathematical Programming 149(1-2): 265–299.

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski (2009). Robust Optimization. Princeton University Press, Princeton NJ, USA.

Ben-Tal, A., A. Goryashko, E. Guslitzer, and A. Nemirovski (2004). Adjustable robust solutions of uncertain linear programs. Mathematical Programming, 99(2):351–376.

Ben-Tal, A., and A. Nemirovski (1999). Robust Solutions of Uncertain Linear Programs. Operations Research Letters 25(1):1–13.

Bertsimas, D., D. B. Brown, and C. Caramanis (2011). Theory and Applications of Robust Optimization. SIAM Review 53(3):464–501.

Bertsimas, D., D. Pachamanova, and M. Sim (2004). Robust Linear Optimization under General Norms. Operations Research Letters 32(6):510–16.

Birge, J. R., and F. Louveaux (2011). Introduction to stochastic programming. Second Edition, Springer Verlag, New York.

Brown D. and Sim M. (2009) "Satisficing Measures for Risky Position" Management Science, 55(1):71–84.

Charnes, A., and W. Cooper (1963). Deterministic Equivalents for Optimizing and Satisficing under Chance Constraints. Operations Research 11(1):18–39.

Chen, X., M. Sim, and P. Sun (2007). A Robust Optimization Perspective on Stochastic Programming. Operations Research 55(6):1058–71.

Delage, E., and Y. Ye (2010). Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems. Operations Research 58(3):595–612.

Goh, J., and N. G. Hall (2013). Total Cost Control in Project Management via Satisficing. Management Science 59(6):1354–1372.

Goh, J., and M. Sim (2010). Distributionally robust optimization and its tractable approximations. Operations Research 58(4)902–917.

Güth, W. (2010). Satisficing and (un)bounded rationality – A formal definition and its experimental validity. Journal of Economic Behavior & Organization 73(3):308–316.

E.M.T. Hendrix, Carmen J. Mecking, Theo H.B. Hendriks (1996), Finding robust solutions for product design problems, European Journal of Operational Research, 92(1): 28-36.

Lam S.W., T.S. Ng, M. Sim, and J-H. Song (2013). "Multiple Objectives Satisficing under Uncertainty", Operations Research 61(1):pp.214–227.

Nemirovski, A., and A. Shapiro (2006). Convex Approximations of Chance Constrained Programs. SIAM Journal on Optimization 17(4):969–996.

Pagnoncelli, B. K., S. Ahmed, and A. Shapiro (2009). Sample Average Approximation Method for Chance Constrained Programming: Theory and Applications. Journal of Optimization Theory and Applications 142(2):399–416.

Prékopa, A. (1980). Logarithmic Concave Measures and Related Topics. Stochastic Programming, 63-82.

Prékopa, A. (1995). Stochastic Programming. Springer Netherlands.

Prékopa, A. (2003). Probabilistic programming, Handbooks in Operations Research and Management Science, 10:267–351. Elsevier.

Shapiro, A. (2003). Monte Carlo Sampling Methods. Handbooks in Operations Research and Management Science 10:353–425. Elsevier.

Shapiro, A., and T. Homem-de-Mello (2000). On the Rate of Convergence of Optimal Solutions of Monte Carlo Approximations of Stochastic Program. SIAM Journal of Control Optimization 11(1):70–86.

Simon, H. A. (1959). Theories of Decision-Making in Economics and Behavioral Science. The American Economic Review 49(3):253–283.

Stüttgen, P., P. Boatwright, and R. T. Monroe (2012). A Satisficing Choice Model. Marketing Science, 31(6):878–899.

Wiesemann, W., D. Kuhn, and M. Sim (2014). Distributionally Robust Convex Optimization. Operations Research 62(6):1358–1376.

Wolsey, L. A. (1998). Integer Programming. Wiley and Sons.

Zhang, X., M. Kamgarpour, A. Georghiou, P. Goulart, J. Lygeros (2016). "Robust optimal control with adjustable uncertainty sets," Automatica 75(C):249—259.

Appendix

A. Technical proofs

A.1. Proof of Theorem 1

Proof: We first establish that the function in (5) is indeed a satisficing decision criterion. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ such that $\mathcal{T}(\boldsymbol{y}) \subseteq \mathcal{T}(\boldsymbol{x})$. We then have $\mathcal{U}(\boldsymbol{\alpha}) \subseteq \mathcal{T}(\boldsymbol{x})$ whenever $\mathcal{U}(\boldsymbol{\alpha}) \subseteq \mathcal{T}(\boldsymbol{y})$. This implies that $\nu(\boldsymbol{y}) \leq \nu(\boldsymbol{x})$ by definition of (5), which establishes the satisficing dominance property. Assume now that $\mathcal{T}(\boldsymbol{x}) = \emptyset$, then Problem (5) will always be infeasible for all $\boldsymbol{\alpha} \in \mathcal{S}$. Hence, we have $\nu(\boldsymbol{x}) = -\infty$, which shows the infeasibility property.

We next show that any satisficing decision criterion can be represented in a form given by (5) in which $P = N$. Consider a satisficing decision criterion, $\bar{\nu} : \mathbb{R}^N \mapsto \mathbb{R} \cup \{-\infty\}$, as described in Definition 1. Define

$$\mathcal{S} = \left\{ \boldsymbol{\alpha} \in \mathbb{R}^N \mid \mathcal{T}(\boldsymbol{\alpha}) \neq \emptyset \right\},$$

$$\mathcal{U}(\boldsymbol{\alpha}) = \mathcal{T}(\boldsymbol{\alpha})$$

and $\rho(\boldsymbol{\alpha}) = \bar{\nu}(\boldsymbol{\alpha})$. Assume that $\mathcal{S}$ is non-empty (otherwise any satisficing decision criterion ν would be degenerate, i.e., $\nu(\boldsymbol{x}) = -\infty$ for all $\boldsymbol{x} \in \mathbb{R}^N$, a trivial case). It follows from the above definition that $\mathcal{S} \subseteq \mathbb{R}^N$ and $\mathcal{U}(\boldsymbol{\alpha}) \neq \emptyset$ for all $\boldsymbol{\alpha} \in \mathcal{S}$. For a given $\boldsymbol{x} \in \mathbb{R}^N$, define $\nu(\boldsymbol{x})$ as

$$\nu(\boldsymbol{x}) = \max_{\boldsymbol{\alpha} \in \mathcal{S}} \{\bar{\nu}(\boldsymbol{\alpha}) \mid \mathcal{U}(\boldsymbol{\alpha}) \subseteq \mathcal{T}(\boldsymbol{x})\} \tag{36}$$

Note that $\nu(\boldsymbol{x})$ is of the form represented in (5). We show that $\nu(\boldsymbol{x}) = \bar{\nu}(\boldsymbol{x})$. First, if $\mathcal{T}(\boldsymbol{x}) = \emptyset$, Problem (36) would be infeasible for all $\boldsymbol{\alpha} \in \mathcal{S}$. Hence, we would have $\nu(\boldsymbol{x}) = -\infty = \bar{\nu}(\boldsymbol{x})$. Next, suppose $\mathcal{T}(\boldsymbol{x}) \neq \emptyset$. It is clear that $\boldsymbol{\alpha} = \boldsymbol{x}$ must be a feasible solution in (5), since $\mathcal{U}(\boldsymbol{\alpha}) = \mathcal{T}(\boldsymbol{\alpha})$ by definition. Furthermore, since for all such feasible $\boldsymbol{\alpha}$, we have $\mathcal{T}(\boldsymbol{\alpha}) = \mathcal{U}(\boldsymbol{\alpha}) \subseteq \mathcal{T}(\boldsymbol{x})$, by satisficing dominance of $\bar{\nu}$ we have $\bar{\nu}(\boldsymbol{x}) \geq \bar{\nu}(\boldsymbol{\alpha})$, and hence indeed $\nu(\boldsymbol{x}) = \bar{\nu}(\boldsymbol{x})$ as claimed. □

A.2. Proof for Proposition 1

The S-model in (6) is re-written here as:

$$\max_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{\alpha} \in \mathcal{S}} \{\rho(\boldsymbol{\alpha}) \mid \mathcal{U}(\alpha) \subseteq \mathcal{T}(\boldsymbol{x})\} \tag{37}$$

Note that any feasible solution in (16) must be feasible in the S-model (37), based on the definition of $\xi(\boldsymbol{\alpha})$. On the other hand, (16) is a relaxation of (37) in the constraints, since $\xi(\alpha) \in \mathcal{U}(\boldsymbol{\alpha})$. Finally, the objective functions of both problems are the same. The equivalence thus follows. □

A.3. Proof for Theorem 2

Proof: We focus on the nontrivial case when the objective of the P-model (1) is finite. Since each constraint is affected by at most one random factor, we consider the following *restricted* affine factor model based on (19):

$$a_{ij}(\boldsymbol{z}) = a_{ij}^0 + a_{ij}^{\kappa(i)} z_{\kappa(i)} \quad \forall i \in [M], j \in [N], \qquad b_i(\boldsymbol{z}) = b_i^0 + b_i^{\kappa(i)} z_{\kappa(i)} \quad \forall i \in [M].$$

Consider any $\boldsymbol{x} \in \mathcal{X}$ such that $\mathbb{P}[\boldsymbol{A}(\tilde{\boldsymbol{z}})\boldsymbol{x} \geq \boldsymbol{b}(\tilde{\boldsymbol{z}})] > 0$. Let

$$\mathcal{M}_k^0(\boldsymbol{x}) = \left\{ i \in [M] \mid \sum_{j\in[N]} a_{ij}^k x_j = b_i^k, \kappa(i) = k \right\},$$

$$\mathcal{M}_k^+(\boldsymbol{x}) = \left\{ i \in [M] \mid \sum_{j\in[N]} a_{ij}^k x_j > b_i^k, \kappa(i) = k \right\}, \ \mathcal{M}_k^-(\boldsymbol{x}) = \left\{ i \in [M] \mid \sum_{j\in[N]} a^k x_j < b_i^k, \kappa(i) = k \right\},$$

for $k \in [K]$. Observe that under the restricted affine uncertainty, the robust counterpart for the i^{th} constraint $\boldsymbol{a}_i(\boldsymbol{z})'\boldsymbol{x} \geq b_i(\boldsymbol{z}),\ \forall \boldsymbol{z} \in [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}]$, $i \in [M]$, of Problem (20) is equivalent to

$$\sum_{j\in[N]} a_{ij}^0 x_j + \min_{\underline{\alpha}_{\kappa(i)} \leq z_{\kappa(i)} \leq \overline{\alpha}_{\kappa(i)}} \left(\sum_{j\in[N]} a_{ij}^{\kappa(i)} x_j - b_i^{\kappa(i)} \right) z_{\kappa(i)} \geq b_i^0,$$

or equivalently, we have for all $k \in [K]$,

$$\sum_{j\in[N]} a_{ij}^0 x_j \geq b_i^0 \quad \forall i \in \mathcal{M}_k^0(\boldsymbol{x}), \qquad z_k = \underline{\alpha}_k \geq \frac{b_i^0 - \sum_{j\in[N]} a_{ij}^0 x_j}{\sum_{j\in[N]} a_{ij}^k x_j - b_i^k} \quad \forall i \in \mathcal{M}_k^+(\boldsymbol{x}), \qquad z_k = \overline{\alpha}_k \leq \frac{b_i^0 - \sum_{j\in[N]} a_{ij}^0 x_j}{\sum_{j\in[N]} a_{ij}^k x_j - b_i^k} \quad \forall i \in \mathcal{M}_k^-(\boldsymbol{x}).$$

Note that since $\mathbb{P}[\boldsymbol{A}(\tilde{\boldsymbol{z}})\boldsymbol{x} \geq \boldsymbol{b}(\tilde{\boldsymbol{z}})] > 0$, it follows that the constraints from the set $\bigcup_{k\in[K]} \mathcal{M}_k^0$ are all satisfied. Based on the convention that maximization and minimization objective values are '$-\infty$' and '$+\infty$' respectively for the case of infeasibility, we have that

$$\begin{aligned}
&\ln \mathbb{P}[\boldsymbol{A}(\tilde{\boldsymbol{z}})\boldsymbol{x} \geq \boldsymbol{b}(\tilde{\boldsymbol{z}})] \\
&= \ln \mathbb{P}\left[\sum_{j\in[N]} \left(a_{ij}^0 + a_{ij}^{\kappa(i)} \tilde{z}_{\kappa(i)} \right) x_j \geq b_i^0 + b_i^{\kappa(i)} \tilde{z}_{\kappa(i)},\ \forall i \in \mathcal{M}_k^+(\boldsymbol{x}) \cup \mathcal{M}_k^-(\boldsymbol{x}) \cup \mathcal{M}_k^0(\boldsymbol{x}),\ k \in [K] \right] \\
&= \sum_{k\in[K]} \ln \mathbb{P}\left[\max_{i\in\mathcal{M}_k^+(\boldsymbol{x})} \left(\frac{b_i^0 - \sum_{j\in[N]} a_{ij}^0 x_j}{\sum_{j\in[N]} a_{ij}^k x_j - b_i^k} \right) \leq \tilde{z}_k \leq \min_{i\in\mathcal{M}_k^-(\boldsymbol{x})} \left(\frac{b_i^0 - \sum_{j\in[N]} a_{ij}^0 x_j}{\sum_{j\in[N]} a_{ij}^k x_j - b_i^k} \right) \right] \\
&= \max_{\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}} \in \mathcal{W}} \left\{ \sum_{k\in[K]} \ln \mathbb{P}[\underline{\alpha}_k \leq \tilde{z}_k \leq \overline{\alpha}_k] \mid \underline{\alpha}_k \geq \frac{b_i^0 - \sum_{j\in[N]} a_{ij}^0 x_j}{\sum_{j\in[N]} a_{ij}^k x_j - b_i^k}\ \forall i \in \mathcal{M}_k^+(\boldsymbol{x}),\ \overline{\alpha}_k \leq \frac{b_i^0 - \sum_{j\in[N]} a_{ij}^0 x_j}{\sum_{j\in[N]} a_{ij}^k x_j - b_i^k}\ \forall i \in \mathcal{M}_k^-(\boldsymbol{x}), k \in [K] \right\} \\
&= \max_{\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}} \in \mathcal{W}} \left\{ \sum_{k\in[K]} \ln \mathbb{P}[\underline{\alpha}_k \leq \tilde{z}_k \leq \overline{\alpha}_k] \mid \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \geq \boldsymbol{b}(\boldsymbol{z})\ \forall \boldsymbol{z} \in [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}] \right\}.
\end{aligned}$$

□

A.4. Proof for Theorem 3

Proof: Under the general affine factor model in (19), the robust counterpart for each of the i^{th} constraint $\boldsymbol{a}_i(\boldsymbol{z})'\boldsymbol{x} \geq b_i(\boldsymbol{z}),\ \forall \boldsymbol{z} \in [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}]$, $i \in [M]$, of Problem (20) can be written as:

$$\sum_{j\in[N]} a_{ij}^0 x_j + \min_{\underline{\boldsymbol{\alpha}} \leq \boldsymbol{z} \leq \overline{\boldsymbol{\alpha}}} \left\{ \sum_{k\in[K]} \left(\sum_{j\in[N]} a_{ij}^k x_j - b_i^k \right) z_k \right\} \geq b_i^0.$$

Since the minimization in the above operates on a linear function in $\boldsymbol{z}$, with $\boldsymbol{z} \in [\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}]$, the above is equivalent to:

$$\sum_{j\in[N]} a_{ij}^0 x_j + \sum_{k\in[K]} \min_{z_k \in \{\underline{\alpha}_k, \overline{\alpha}_k\}} \left\{ \left(\sum_{j\in[N]} a_{ij}^k x_j - b_i^k \right) z_k \right\} \geq b_i^0, \tag{38}$$

which results in the desired formulation. □

A.5. Proof for Proposition 2

Based on the affine factor model assumption in (19), the constraints in (20) can be re-written as:

$$\sum_{k\in[K]}\sum_{j\in[N]}(a_{ij}^k x_j - b_i^k)z_k + \sum_{j\in[N]} a_{ij}^0 x_j - b_i^0 \geq 0 \quad \forall i\in[M]\ \forall \boldsymbol{z}\in[\underline{\boldsymbol{\alpha}},\overline{\boldsymbol{\alpha}}]$$

$$\Leftrightarrow \sum_{k\in[K]}\min_{\alpha_k\in[\underline{\alpha},\overline{\alpha}]}\sum_{j\in[N]}(a_{ij}^k x_j - b_i^k)z_k + \sum_{j\in[N]} a_{ij}^0 x_j - b_i^0 \geq 0 \quad \forall i\in[M]$$

$$\Leftrightarrow \sum_{k\in\overline{[K]}}\sum_{j\in[N]}(a_{ij}^k x_j - b_i^k)\overline{\alpha}_k + \sum_{k\in\underline{[K]}}\sum_{j\in[N]}(a_{ij}^k x_j - b_i^k)\underline{\alpha}_k + \sum_{j\in[N]} a_{ij}^0 x_j - b_i^0 \geq 0 \quad \forall i\in[M],$$

where the last equivalence follows from applying the conditions in (22) and (23). Hence, $\boldsymbol{\xi}(\boldsymbol{\alpha})$ as defined in (24) is indeed a decision-independent worst case outcome. □

A.6. Proof for Theorem 4

Proof: Note that from the uncertainty set $\mathcal{U}(\boldsymbol{\alpha})$, the lower and upper limits of $\tilde{z}_k$, $k\in[K]$ are $\sum_{\ell\in[L(k)]}\zeta_k^\ell\underline{\alpha}_k^\ell$ and $\sum_{\ell\in[L(k)]}\zeta_k^\ell\overline{\alpha}_k^\ell$, respectively. Note also that

$$\ln\mathbb{P}\left[\sum_{\ell\in[L(k)]}\zeta_k^\ell\underline{\alpha}_k^\ell \leq \tilde{z}_k \leq \sum_{\ell\in[L(k)]}\zeta_k^\ell\overline{\alpha}_k^\ell\right] = \ln\left(\sum_{\ell\in[L(k)]}\left(\lambda_k^\ell\overline{\alpha}_k^\ell - \lambda_k^{\ell-1}\underline{\alpha}_k^\ell\right)\right).$$

Since $\ln(\delta)$ is a concave function on domain $\delta > 0$, it follows that

$$\ln(\delta) \leq \ln(\gamma) + \frac{1}{\gamma}(\delta-\gamma) = \ln(\gamma) - 1 + \frac{\delta}{\gamma}$$

for all $\gamma > 0$ and that equality is achieved trivially at $\gamma = \delta$. Moreover, we observe that

$$\sum_{\ell\in[L(k)]}\left(\lambda_k^\ell\overline{\alpha}_k^\ell - \lambda_k^{\ell-1}\underline{\alpha}_k^\ell\right)\in\mathcal{C}_k,$$

since

$$\sum_{\ell\in[L(k)]}\overline{\alpha}_k^\ell = 1,\ \sum_{\ell\in[L(k)]}\underline{\alpha}_k^\ell = 1,\ \sum_{\ell\in[L(k)]}\ell(\overline{\alpha}_k^\ell - \underline{\alpha}_k^\ell) \geq 0.$$

Hence, we have

$$\ln\left(\sum_{\ell\in[L(k)]}\left(\lambda_k^\ell\overline{\alpha}_k^\ell - \lambda_k^{\ell-1}\underline{\alpha}_k^\ell\right)\right) \leq \ln(\gamma) - 1 + \sum_{\ell\in[L(k)]}\frac{1}{\gamma}\left(\lambda_k^\ell\overline{\alpha}_k^\ell - \lambda_k^{\ell-1}\underline{\alpha}_k^\ell\right)$$

for all $\gamma > 0$ and note that equality is achieved when

$$\gamma = \sum_{\ell\in[L(k)]}\left(\lambda_k^\ell\overline{\alpha}_k^\ell - \lambda_k^{\ell-1}\underline{\alpha}_k^\ell\right)\in\mathcal{C}_k.$$

The remaining constraints follow trivially from Theorem 3. □

B. S-model with recourse

We consider a satisficing problem based on a linear optimization problem with recourse, where we have (reproduced from (3)):

$$\mathcal{T}(\boldsymbol{x}) = \left\{ \boldsymbol{z} \in \mathcal{W} \,\middle|\, \begin{array}{l} \exists \boldsymbol{y}(\boldsymbol{\zeta}) \in \mathbb{R}^{N_2} \ \forall \boldsymbol{\zeta} \in \mathcal{W} \ : \\ \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}) \end{array} \right\}.$$

The satisficing decision criterion in (5) can then be written as:

$$\nu(\boldsymbol{x}) = \max_{\boldsymbol{\alpha} \in \mathcal{S}} \left\{ \rho(\boldsymbol{\alpha}) \,\middle|\, \begin{array}{l} \forall \boldsymbol{z} \in \mathcal{U}(\boldsymbol{\alpha}), \ \exists \boldsymbol{y} \in \mathbb{R}^{N_2} \ : \\ \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{z}) \end{array} \right\}. \tag{39}$$

The corresponding S-model (with recourse) is written as:

$$\begin{array}{rl} \max & \rho(\boldsymbol{\alpha}) \\ \text{s.t.} & \forall \boldsymbol{z} \in \mathcal{U}(\boldsymbol{\alpha}), \ \exists \boldsymbol{y} \in \mathbb{R}^{N_2} \ : \ \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{z}) \\ & \boldsymbol{x} \in \mathcal{X} \\ & \boldsymbol{\alpha} \in \mathcal{S}. \end{array} \tag{40}$$

Unfortunately, in such a case the S-model is not computationally tractable (since even for a fixed $\boldsymbol{\alpha}$, the problem involves checking the feasibility of $\boldsymbol{x}$ in an adjustable robust optimization problem, which itself is NP-hard). To achieve tractability, a common approach in the literature is to use approximations such as affine decision rules and their extensions (see e.g., Zhang et al. (2016)). However, such approximation schemes do not necessarily result in being consistent with the satisficing decision criteria of Definition 1. We illustrate this in the following example.

Example 1. A straightforward approach to approximate $\nu(\boldsymbol{x})$ in (39) is based on a static decision rule defined as:

$$\hat{\nu}(\boldsymbol{x}) = \max_{\boldsymbol{\alpha} \in \mathcal{S}} \left\{ \rho(\boldsymbol{\alpha}) \mid \exists \boldsymbol{y} \in \mathbb{R}^{N_2} : \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{z}) \ \forall \boldsymbol{z} \in \mathcal{U}(\boldsymbol{\alpha}) \right\} \tag{41}$$

Let us consider a simple case where $\mathcal{T}(x) = \{z \in \mathcal{W} | \ \exists y(z) : \mathbb{R} \mapsto \mathbb{R} \text{ such that } z \leq y(z) \leq z + x\}$, with $\mathcal{W} = [0,1]$ being the support of z. Since $\mathcal{T}(0)=\mathcal{T}(1)$, we must have $\nu(0)=\nu(1)$ according to Definition 1 for any satisficing decision criterion ν. Using (41), for $x = 1$, we can have $y = 1$ satisfying the non-adjustable constraints for all $z \in \mathcal{W}$, since $1 \leq y \leq 0 + 1$. However, for the case $x = 0$, clearly no such y exists, since this requires $1 \leq y \leq 0 + 0$. Consequently $\hat{\nu}(0) \neq \hat{\nu}(1)$, and is hence not consistent with the satisficing decision criteria. □

Nevertheless, in some special situations, applying the static decision rule as in the above example turns out to be exact. This is for instance when the T-model has decision-independent worst case uncertainty (see Definition 2). In this case, let $\boldsymbol{\xi}(\boldsymbol{\alpha}) \in \mathcal{W}$ be a decision-independent worst case outcome, given $\boldsymbol{\alpha} \in \mathcal{S}$. Based on Proposition 1, the S-model in (40) can then be reduced as:

$$\begin{array}{rl} \max & \rho(\boldsymbol{\alpha}) \\ \text{s.t.} & \boldsymbol{A}(\boldsymbol{\xi}(\boldsymbol{\alpha}))\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{\xi}(\boldsymbol{\alpha})) \\ & \boldsymbol{x} \in \mathcal{X} \\ & \boldsymbol{\alpha} \in \mathcal{S}. \end{array} \tag{42}$$

C. Piecewise linear approximations for T-models with log-concave probability density functions

For any $\mathcal{D}'_k \subseteq \mathcal{D}_k = \left\{(\underline{\delta}, \overline{\delta}) \in \mathcal{W}_k^2 \ : \ \underline{\delta} < \overline{\delta}\right\}$, we note that each term $\ln \mathbb{P}[\underline{\alpha}_k \leq \tilde{z}_k \leq \overline{\alpha}_k]$ in the objective function of (20) can be written as follows:

$$
\begin{aligned}
&\ln \mathbb{P}\left[\underline{\alpha}_k \leq \tilde{z}_k \leq \overline{\alpha}_k\right] \\
&= \max\left\{p_k \in \mathbb{R} \mid p_k \leq (F_k(\underline{\alpha}_k, \overline{\alpha}_k))\right\} \\
&= \max\left\{p_k \in \mathbb{R} \mid p_k \leq \min_{(\underline{\delta}, \overline{\delta}) \in \mathcal{D}_k} \left(F_k(\underline{\delta}, \overline{\delta})\right) + F_k^1(\underline{\delta}, \overline{\delta})(\underline{\alpha}_k - \underline{\delta}) + F_k^2(\underline{\delta}, \overline{\delta})(\overline{\alpha}_k - \overline{\delta})\right\} \\
&\leq \max\left\{p_k \in \mathbb{R} \mid p_k \leq \ \left(F_k(\underline{\delta}, \overline{\delta})\right) + F_k^1(\underline{\delta}, \overline{\delta})(\underline{\alpha}_k - \underline{\delta}) + F_k^2(\underline{\delta}, \overline{\delta})(\overline{\alpha}_k - \overline{\delta}) \ \forall (\underline{\delta}, \overline{\delta}) \in \mathcal{D}'_k\right\}
\end{aligned}
$$

where the function F_k is as defined in Corollary 1, and F_k^1, and F_k^2 on the domain $\mathcal{D}_k$ are defined as

$$
F_k^1(\underline{\delta}, \overline{\delta}) = \frac{\partial}{\partial \underline{\delta}} F_k(\underline{\delta}, \overline{\delta}) = -\frac{f_k(\underline{\delta})}{\mathbb{P}\left[\underline{\delta} \leq \tilde{z}_k \leq \overline{\delta}\right]}, \ \text{and} \ \ F_k^2(\underline{\delta}, \overline{\delta}) = \frac{\partial}{\partial \overline{\delta}} F_k(\underline{\delta}, \overline{\delta}) = \frac{f_k(\overline{\delta})}{\mathbb{P}\left[\underline{\delta} \leq \tilde{z}_k \leq \overline{\delta}\right]}
$$

The second equality above follows from the first-order necessary and sufficient conditions of concave functions, and by noting that for any given $(\underline{\alpha}_k, \overline{\alpha}_k) \in \mathcal{D}_k$ the minimum is trivially achieved by choosing $(\underline{\delta}, \overline{\delta}) = (\underline{\alpha}_k, \overline{\alpha}_k)$. This thus establishes the equivalent reformulation of $F_k(\underline{\alpha}_k, \overline{\alpha}_k)$ as a point-wise minimum of a set of affine functions. Hence, a relaxation of the problem can be achieved by replacing the original objective function with the last construct in the inequality above, for some $\mathcal{D}'_k$, for each $k \in K$. The constraints:

$$
p_k \leq \ F_k(\underline{\delta}, \overline{\delta}) + F_k^1(\underline{\delta}, \overline{\delta})(\underline{\alpha}_k - \underline{\delta}) + F_k^2(\underline{\delta}, \overline{\delta})(\overline{\alpha}_k - \overline{\delta}) \ \forall (\underline{\delta}, \overline{\delta}) \in \mathcal{D}'_k \tag{43}
$$

which are affine in the decision variables, can be efficiently implemented on general purpose solvers. In practice, the problem can be solved in an iterative cut generation procedure (if all variables are continuous) or in a branch-and-cut procedure (if some of the variables are integer) (see, e.g., Wolsey 1998), adding cuts (43) whenever they are violated. From a computational point of view, this approach is likely to be particularly attractive when the number of violated cuts is small.

D. Model Formulations for Stochastic Maximum Coverage Problem

In this section we provide the model formulations for the Stochastic Maximum Coverage Problem (SMCP) not shown in full in the main body of the paper.

D.1. Formulations for Two-stage SMCP

We first consider the P-model for the two-stage SMCP, in which demand is allocated after the actual customer demands are observed.

Let binary variables x_i be 1 if facility i has been selected, and 0 otherwise. The second stage demand allocation from facility i to customer j is denoted by $y_{ij}(\tilde{\boldsymbol{z}})$, where $y_{ij}(\boldsymbol{z}) : \mathbb{R}^I \mapsto \mathbb{R}$ is a measurable function that depends on the realization of the uncertain customers' demand quantities $\tilde{\boldsymbol{z}}$. Function $y_{ij}(\cdot)$ is optimized over the family of all measurable functions $\mathcal{R}(J,1)$ that map from $\mathbb{R}^J$ to $\mathbb{R}$. The formulation for the two-stage P-model for the SMCP is as follows.

$$\begin{array}{lll} \max & \ln \mathbb{P}\left[\sum_{i\in\mathcal{I}_j} y_{ij}(\tilde{\boldsymbol{z}}) \geq b_j^0 + \sum_{k\in[K]} b_j^k \tilde{z}_k \quad \forall j \in [J]\right] & \\ \text{s.t.} & \sum_{j\in\mathcal{J}_i} y_{ij}(\boldsymbol{z}) \leq c_i x_i & \forall \boldsymbol{z}\in\mathcal{W}, i\in[I] \\ & \sum_{i\in[I]} f_i x_i \leq \beta & \\ & y_{ij}(\boldsymbol{z}) \geq 0 & \forall \boldsymbol{z}\in\mathcal{W},\ i\in[I],\ j\in\mathcal{J}_i \\ & y_{ij}\in\mathcal{R}(J,1) & \forall i\in[I],\ j\in\mathcal{J}_i \\ & x_i\in\{0,1\} & \forall i\in[I], \end{array} \tag{44}$$

where the objective function maximizes the probability that the uncertain demand can be met. The first set of constraints are the facility capacity constraints, while the second constraint restricts the construction costs to the available budget.

T-models for the SMCP under decision-independent worst-case uncertainty. For the case where T-models assume decision-independent worst-case uncertainty as by Definition 2, we evaluate two implementations. First, model $\text{T}_{\mathbb{P}}$-2, assuming log-concave probability densities (see § 3.1); and model T-2, based on the discrete distribution case (see § 3.2) by using an empirical distribution of $\tilde{\boldsymbol{z}}$, given by the samples $\hat{\boldsymbol{\zeta}}^\ell \in \mathbb{R}^K$, $\ell\in[L]$, each with equal probability of realization. We assume $\tilde{\boldsymbol{z}}$ themselves are independently distributed, and furthermore each $\tilde{z}_k$ always impact all customer demands either positively or negatively (see Appendix E for full details). The random factors can therefore be divided into two disjoint sets $\underline{\mathcal{K}}$ and $\overline{\mathcal{K}}$, which satisfies the decision-independent worst-case uncertainty (Definition 2, and see also Proposition 2). The two variants of T-models for the two-stage SMCP can then be formulated as follows.

$$\begin{array}{lll} \max & \sum_{k\in\underline{[K]}}\sum_{\ell\in[L]} \ln\left(1-\frac{\ell}{L}\right)\underline{\alpha}_k^\ell + \sum_{k\in\overline{[K]}}\sum_{\ell\in[L]} \ln\left(\frac{\ell}{L}\right)\overline{\alpha}_k^\ell & \\ \text{s.t.} & \sum_{i\in\mathcal{I}_j} y_{ij} \geq b_j^0 + \sum_{k\in\underline{[K]}}\sum_{\ell\in[L]} b_j^k\zeta_k^\ell\underline{\alpha}_k^\ell + \sum_{k\in\overline{[K]}}\sum_{\ell\in[L]} b_j^k\zeta_k^\ell\overline{\alpha}_k^\ell & \forall j\in[J] \\ & \sum_{j\in\mathcal{J}_i} y_{ij} \leq c_i x_i & \forall i\in[I] \\ & \sum_{i\in[I]} f_i x_i \leq \beta & \\ & \sum_{\ell\in[L]} \underline{\alpha}_k^\ell = 1 & \forall k\in\underline{[K]} \\ & \sum_{\ell\in[L]} \overline{\alpha}_k^\ell = 1 & \forall k\in\overline{[K]} \\ & y_{ij}\geq 0 & \forall i\in[I],\ j\in\mathcal{J}_i \\ & x_i\in\{0,1\} & \forall i\in[I] \\ & \underline{\alpha}_k^\ell, \overline{\alpha}_k^\ell \in\{0,1\} & \forall \ell\in[L],\ k\in[K]. \end{array} \tag{Model T-2}$$

where $(\zeta_k^1,\dots,\zeta_k^L)$ denotes the ordered statistics of the empirical samples $(\hat{\zeta}_k^1,\dots,\hat{\zeta}_k^L)$ for each factor $k\in[K]$.

D.2. Sample average approximation model formulations

We consider two types of sample average approximation (SAA) optimization models, that assume as input a sample of L scenarios of the demand observations, obtained using the affine factor model with the samples $\hat{\zeta}_k^\ell$ for each random factor $\tilde{z}_k$, $\ell=1,\dots,L$. The first type of SAA models corresponds to the P-model objective for the single- and two-stage SMCP, maximizing the number of feasible scenarios in the given sample. The model involves one demand satisfaction constraint for each scenario $\ell \in [L]$. Binary variables p^ℓ take value 1 if demand scenario $\ell \in [L]$ is covered by the set of selected facilities, and 0 otherwise. Since the P-model objective is monotone in the feasibility probability $\mathbb{P}\left[\sum_{i\in\mathcal{I}_j} y_{ij}(\tilde{\boldsymbol{z}}) \geq b_j^0 + \sum_{k\in[K]} b_j^k \tilde{z}_k \quad \forall j \in [J]\right]$, we consider the following SAA models for the single-stage and two-stage SMCP:

$$
\begin{aligned}
\max\ & \frac{1}{L}\sum_{\ell\in[L]} p^\ell \\
\text{s.t.}\ & \sum_{i\in\mathcal{I}_j} y_{ij} \geq \left(b_j^0 + \sum_{k\in[K]} b_j^k \hat{\zeta}_j^\ell\right) p^\ell && \forall j\in[J],\ \ell\in[L] \\
& \sum_{j\in\mathcal{J}_i} y_{ij} \leq c_i x_i && \forall i\in[I] \\
& \sum_{i\in[I]} f_i x_i \leq \beta \\
& y_{ij} \geq 0 && \forall i\in[I],\ j\in\mathcal{J}_i \\
& x_i \in\{0,1\} && \forall i\in[I] \\
& p^\ell \in\{0,1\} && \forall \ell\in[L],
\end{aligned}
\qquad \text{(Model P-1)}
$$

to

$$
\begin{aligned}
\max\ & \frac{1}{L}\sum_{\ell\in[L]} p^\ell \\
\text{s.t.}\ & \sum_{i\in\mathcal{I}_j} y_{ij}^\ell \geq \left(b_j^0 + \sum_{k\in[K]} b_j^k \hat{\zeta}_j^\ell\right) p^\ell && \forall j\in[J],\ \ell\in[L] \\
& \sum_{j\in\mathcal{J}_i} y_{ij}^\ell \leq c_i x_i && \forall i\in[I],\ \ell\in[L] \\
& \sum_{i\in[I]} f_i x_i \leq \beta \\
& y_{ij}^\ell \geq 0 && \forall i\in[I],\ j\in\mathcal{J}_i,\ \ell\in[L] \\
& x_i \in\{0,1\} && \forall i\in[I] \\
& p^\ell \in\{0,1\} && \forall \ell\in[L].
\end{aligned}
\qquad \text{(Model P-2)}
$$

Conceivably, due to the additional binary variables p^ℓ, the model will be difficult to solve when the number of scenarios L is large. Nevertheless, for reasonably sized instances, Model P-1 can be implemented and solved directly using general-purpose MIP solvers.

The second SAA model used for comparison aims at minimizing the total expected demand shortfall

$$
\sum_{j\in[J]} \mathbb{E}\left[\left(\left(\sum_{i\in\mathcal{I}_j} y_{ij}(\tilde{\boldsymbol{z}})\right) - \left(b_j^0 + \sum_{k\in[K]} b_j^k \tilde{z}_k\right)\right)^+\right].
$$

Let s_j^ℓ be the demand shortfall for the demand of customer j at scenario ℓ. The approximation for the single-stage problem variant is formulated as:

$$\begin{aligned}
\min\ & \frac{1}{L}\sum_{j\in[J]}\sum_{\ell\in[L]} s_j^\ell \\
\text{s.t.}\ & \sum_{i\in[I]} y_{ij} + s_j^\ell \geq \left(b_j^0 + \sum_{k\in[K]} b_j^k \hat{\zeta}_j^\ell\right) p^\ell && \forall j\in[J],\ \ell\in[L] \\
& \sum_{j\in[J]} y_{ij} \leq c_i x_i && \forall i\in[I] \\
& \sum_{i\in[I]} f_i x_i \leq \beta \\
& y_{ij} \geq 0 && \forall i\in\mathcal{I}_j,\ j\in[J] \\
& s_j^\ell \geq 0 && \forall j\in[J],\ \ell\in[L] \\
& x_i\in\{0,1\} && \forall i\in[I].
\end{aligned} \tag{Model E-1}$$

The corresponding two-stage model SMCP is:

$$\begin{aligned}
\min\ & \frac{1}{L}\sum_{j\in[J]}\sum_{\ell\in[L]} s_j^\ell \\
\text{s.t.}\ & \sum_{i\in[I]} y_{ij}^\ell + s_j^\ell \geq \left(b_j^0 + \sum_{k\in[K]} b_j^k \hat{\zeta}_j^\ell\right) p^\ell && \forall j\in[J],\ \ell\in[L] \\
& \sum_{j\in[J]} y_{ij}^\ell \leq c_i x_i && \forall i\in[I],\ \ell\in[L] \\
& \sum_{i\in[I]} f_i x_i \leq \beta \\
& y_{ij}^\ell \geq 0 && \forall i\in\mathcal{I}_j,\ j\in[J],\ \ell\in[L] \\
& s_j^\ell \geq 0 && \forall j\in[J],\ \ell\in[L] \\
& x_i\in\{0,1\} && \forall i\in[I].
\end{aligned} \tag{Model E-2}$$

Like Models P-1 and P-2, Models E-1 and E-2 can be implemented and solved directly using general purpose MIP solvers. Note that a computational advantage of Models E-1 and E-2 over Models P-1 and P-2 is that they avoid the use of additional binary variables for modeling the objective function.

E. Generation of Problem Instances for SMCP

We generate the problem instances for the SMCP as follows. The number of customers J is taken from $J\in\{100,250,500\}$; the number of candidate facility locations is chosen accordingly from $I\in\{0.5J,J,2J\}$. The arcs between customers and facilities have been generated randomly such that each customer is connected to $A\%$ of the I facilities, where the parameter $A\in\{20,40\}$ can be regarded as a surrogate for the level of facility-customer connectivity assumed in the problem.

We generate instances with independent and correlated customer demands. The demands are modeled as affine functions of independently distributed random factors $\tilde{z}_1,\ldots,\tilde{z}_K$, with $K>J$, written as:

$$b_j(\tilde{\boldsymbol{z}}) = b_j^0 + b_j^j\tilde{z}_j + \sum_{k=J+1}^{K} b_j^k\tilde{z}_k\ \forall j\in[J],$$

For simplicity, we assume in all our instances that the constant term $b_j^0=0$ for all $j\in[J]$. For instances with independent customer demands, the parameters $b_j^k=0$ for all $k=J+1,\ldots,K$. For instances with correlated customer demands, the demand of each customer is composed of an individual demand term $b_j^j z_j$, plus up to 10 random factors that are randomly selected from a pool of shared demand factors. Among those, 20% are assumed to negatively impact the customer demands. We assume that the total number of random factors $K = J + \lceil\sqrt{J}\rceil$, so that the number of shared random factors $(K - J) = \lceil\sqrt{J}\rceil$. For example, with $J = 500$ customers, we assume that the demand of each customer is a linear combination of its own random factor, as well as some of the $\lceil\sqrt{500}\rceil = 23$ shared random factors. We further assume that the individual term accounts for approximately 30% of the entire demand, therefore $b_j^j = 0.3,\ \forall j \in [J]$. The coefficients of the random factors are uniformly chosen at random normalized in each linear combination, such that $\sum_{k\in[K]} b_j^k = 1,\ \forall j \in [J]$. Therefore, we need to choose $b_j^{J+1}, \ldots, b_j^K$ in a way such that $\sum_{k=J+1}^{K} b_j^k = 1 - b_j^j = 0.7$. In practice, it is reasonable to assume that the number of random factors each customer demand depends on, here referred to as $nRFC$, does not exceed 10. Each customer demand is therefore modeled by an individual random factor, as well as up to 10 random factors that it may have in common with other customers. For each customer, these 10 random factors are chosen randomly according to a uniform distribution from $[J+1, K]$. Note that the 10 random factors are randomly sampled with replacement, and sampling more than once the same random factor implies that the corresponding random coefficients are summed. While the above generated customer demands with significant correlation, we additionally generate the same instances without shared random factors, i.e., $nRFC = 0$ and $K = |J|$. The random factors themselves are generated as follows. All random factors in $[K]$ follow a continuous normal distribution with mean μ_k and standard deviation $\sigma_j = 0.5\mu_k$, where μ_k is chosen randomly according to a uniform distribution between 1 and 100. If the random outcome of a factor has been smaller than 0, it is set to 0. We assume that the individual random factors $z_1, \ldots, z_J$ have a positive influence on the customer demand, i.e., they can only take non-negative values. Among the shared random factors, we randomly choose 20% to have a negative influence on customer demands (i.e., they are non-positive) and the remaining to have a positive influence.

Here, a subtle detail impacts whether the resulting instances will allow the T-model to exploit demand-independent worst-case uncertainty or not. Specifically, if the 20% of individual random factors with negative impact on the demand are the same for all customers (i.e., if a random factor negatively impacts demand for one customer, then it never positively impacts demand for another customer), the T-model has demand-independent worst-case uncertainty. Instances generated in this way are used in the computational experiments for the two-stage problem variant. In contrast, if we randomly select the 20% of the random factors with negative impact separately for each customer (i.e., a random factor may negatively impact demand for one customer, but positively impact demand for another), then the T-model does not necessarily have demand-independent worst-case uncertainty. These instances are used in the experiments for the single-stage problem variant.

For the facilities that can be constructed, capacities are set sequentially to 500, 750 and 1,000 units, i.e., $c_{3i'} = 500$, $c_{3i'+1} = 750$ and $c_{3i'+2} = 1,000$ for the running index i'. As the average customer demand is about 50 units, a total construction budget of $50 \times J$ would, on average, ensure that all customer demands can be met, if each customer could be served by any facility (i.e., $A = 100$). However, as demands are uncertain and the number of service arcs between facilities and customers are restricted, we allow for an additional budget of B%, i.e., we set the total budget allowance $\beta = (1 + B/100) \times 50 \times J$. If not otherwise stated, we set $B = 5$ for the two-stage problems. For the single-stage problem, we allow a higher budget with $B = 150$, since demands are naturally more difficult to satisfy in this non-adjustable problem variant.

F. Computational Results for the Two-stage SMCP

In this section, we report on computational results for the two-stage SMCP defined in Appendix D.1. The studies will follow the same format as those for the single-stage problem variant. However, to enable a tractable two-stage T-model, we rely on demand-independent worst-case uncertainty. All computational experiments have therefore been carried out on instances, in which random factors either negatively impact demand for all customers, or positively impact demand for all customers (see Appendix E for details).

We compare the performance of the models T-2, P-2 and E-2. To correctly measure the P-model objective and account for the fact that demands can be reallocated after observing it, we evaluate the success rates and demand shortfalls for each of the scenarios based on the optimal demand allocation between the given set of facilities and the customer demands of the specific scenario. Table 5 reports average results for all instances which have been solved by the respective models. While Model E-2 may provide good results when sufficient time is available, Model P-2 would theoretically provide the best results when L is large (which, however, becomes computationally intractable). Owing to the ability to exploit the decision independent worst-case uncertainty, Model T-2 is significantly more tractable computationally (and even more so compared to Model T-1 in Section 4). For all instances that have either more than 100 customers or use more than 25 samples, Models P-2 and E-2 often hit the time-limit of 12 hours and report sub-optimal solutions, compromising the overall success rates and demand shortfalls. Model T-2 finds feasible solutions for all problem instances and number of samples, therefore consistently providing the best overall-performance among all three models.

		Model T-2				Model P-2				Model E-2			
J	L	succ. rate %	short fall	time (min)	# ns	succ. rate %	short fall	time (min)	# ns	succ. rate %	short fall	time (min)	# ns
100	5	63.68	0.1	2.3	0	63.58	0.1	0.6	0	61.32	0.1	75.7	0
	25	64.75	0.1	0.3	0	59.48	0.5	78.1	0	52.30	0.1	158.7	0
	50	60.05	0.1	1.3	0	52.42	1.0	214.2	0	46.77	0.1	243.4	0
	100	48.71	0.1	110.6	0	36.11	2.1	345.5	0	39.69	1.1	381.0	0
250	5	67.22	0.3	3.6	0	61.66	0.3	2.2	0	66.17	0.3	73.4	0
	25	65.32	0.3	0.3	0	46.84	3.6	365.6	0	49.96	0.9	365.8	0
	50	69.39	0.3	0.2	0	15.98	8.0	598.9	0	31.05	3.7	487.1	0
	100	67.92	0.3	0.3	0	4.30	9.7	698.9	0	19.14	7.0	671.7	0
500	5	77.34	0.3	1.5	0	77.33	0.3	37.7	0	77.33	0.3	24.3	0
	25	77.52	0.3	0.1	0	0.00	22.0	720.0	0	25.84	14.8	620.3	0
	50	77.51	0.3	0.2	0	5.17	20.5	693.9	0	15.90	14.8	720.0	0
	100	77.51	0.3	0.5	0	0.00	22.0	720.0	0	0.00	22.0	720.0	0

Table 5 Performance comparison for different instance sizes J and sample sizes L for two-stage SMCP: success rate, average demand shortfall (in 1000 units) and average computing time (in minutes) for the instances where the solver has found a feasible solution. The number of instances where the solver did not find any feasible solution is reported in column '# ns'. As L increases, Models P and E find feasible solutions for less instances, decreasing the average success rate and increasing the demand shortfall. In contrast, Model T-2 remains relatively stable.

We next investigate the performance of the sample-based models with respect to the problem and sample size, under multiple replications. Table 6 shows the average, minimum and maximum of the average success rates (over all 60 instances) for the different models and different sample sizes L. Instances for which no feasible solution has been found were counted as 0 success rates. For Models P-2 and E-2, we also report results for sample size $L = 5$, as these models were unable to handle large sample sizes in our computational studies. The T-2 results are shown for different sample sizes, from 5 to 1000. The average success rates and sample-to-sample variation tend to slightly decrease with higher sample sizes.

	Model P-2		Model E-2		Model T-2				
J	$L = 5$	$L = 25$	$L = 5$	$L = 25$	$L = 5$	$L = 25$	$L = 100$	$L = 500$	$L = 1000$
100	51.62	49.79	51.22	46.17	50.88	50.84	50.63	50.52	50.12
	[38.1, 54.5]	[49.1, 54.4]	[47.2, 54.4]	[43.5, 51.1]	[41.7, 54.5]	[45.4, 54.4]	[46.8, 54.4]	[46.6, 54.6]	[44.1, 54.5]
250	65.33	38.74	63.78	46.35	67.83	67.03	65.85	67.13	66.34
	[54.8, 70.3]	[23.3, 46.8]	[54.9, 67.9]	[40.5, 50.0]	[65.5, 69.6]	[63.0, 70.3]	[63.9, 68.2]	[61.8, 70.3]	[61.4, 68.9]
500	66.12	3.01	75.50	21.35	77.47	76.77	76.49	75.33	77.47
	[46.0, 77.6]	[0.0, 12.9]	[67.2, 77.6]	[18.4, 25.9]	[77.3, 77.6]	[72.4, 77.6]	[72.5, 77.6]	[69.1, 77.6]	[77.3, 77.6]

Table 6 Scalability study for two-stage SMCP, comparing average [minimum, maximum] success rates (over all 60 instances) among 10 replications with different sample sizes L.

The average computing times of Model T-2 across all instances are 69, 57, 64, 56, and 59 minutes when using 5, 25, 100, 500 and 1,000 demand scenarios, respectively. The results hence suggest that Model T-2 is highly scalable in the size of the samples used. However, it seems that, for this problem variant, using a large number of samples does add little value to the final solution quality.