Mathematical Programming (2022) 196:309–345
https://doi.org/10.1007/s10107-022-01823-6

FULL LENGTH PAPER

Series B

Probability estimation via policy restrictions, convexification, and approximate sampling

Ashish Chandra[1] · Mohit Tawarmalani[1]

Received: 22 May 2020 / Accepted: 14 April 2022 / Published online: 20 May 2022
© Springer-Verlag GmbH Germany, part of Springer Nature and Mathematical Optimization Society 2022

Abstract

This paper develops various optimization techniques to estimate probability of events where the optimal value of a convex program, satisfying certain structural assumptions, exceeds a given threshold. First, we relate the search of affine/polynomial policies for the robust counterpart to existing relaxation hierarchies in MINLP (Lasserre in Proceedings of the international congress of mathematicians (ICM 2018), 2019; Sherali and Adams in A reformulation–linearization technique for solving discrete and continuous nonconvex problems, Springer, Berlin). Second, we leverage recent advances in Dworkin et al. (in: Kaski, Corander (eds) Proceedings of the seventeenth international conference on artificial intelligence and statistics, Proceedings of machine learning research, PMLR, Reykjavik, 2014), Gawrychowski et al. (in: ICALP, LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018) and Rizzi and Tomescu (Inf Comput 267:135–144, 2019) to develop techniques to approximately compute the probability binary random variables from Bernoulli distributions belong to a specially-structured union of sets. Third, we use convexification, robust counterpart, and chance-constrained optimization techniques to cover the event set of interest with such set unions. Fourth, we apply our techniques to the network reliability problem, which quantifies the probability of failure scenarios that cause network utilization to exceed one. Finally, we provide preliminary computational evaluation of our techniques on test instances for network reliability.

Keywords Uncertainty quantification · Approximate counting and sampling · Robust counterpart · Moment problem · Chance-constrained optimization

Mathematics Subject Classification 90C17 · 60F10 · 68M15 · 90C26

✉ Mohit Tawarmalani
mtawarma@purdue.edu
http://web.ics.purdue.edu/∼mtawarma/

Ashish Chandra
aashishc@purdue.edu

1 Krannert School of Management, Purdue University, West Lafayette 47906, USA

1 Introduction

This paper focuses on *probability quantification* (PQ), which is to estimate $\Pr\big(d(\mathbb{X}) = \min_y f(\mathbb{X}, y) > \varrho\big)$, where $\mathbb{X} \in \mathbb{R}^m$ is a random variable, $\varrho \in \mathbb{R}$, $y \in \mathbb{R}^n$, and $f : \mathbb{R}^{n+m} \to \mathbb{R}$. We begin our analysis with the general problem and successively impose structural assumptions on $\mathbb{X}$ and $\min_y f(\mathbb{X}, y)$. PQ plays an important role in reliability problems arising in diverse industries. A particular example is the *network reliability problem* (NR), which estimates the probability of bad failures, *i.e.,* failures where, across all links, the maximum ratio of the traffic relative to capacity–referred to as *maximum link utilization* (MLU)–exceeds one. A closely related problem, abbreviated as (RNR), certifies that MLU is below one across all failure states [10, 25, 39]. There is emerging literature on NR and a rising interest in service-level agreements [2, 9].

The general optimal uncertainty quantification (OUQ) problem allows $d(\mathbb{X})$ to be any response function [14, 32], while PQ focuses on the case where $d(\mathbb{X})$ is the value function of a convex program. PQ is hard to solve because RNR is already NP-Hard for a budgeted uncertainty set [9, 41]. Inequalities due to Markov, Chebyshev, and Chernoff are often used to bound probabilities of events, modeled as PQ. Moreover, semidefinite relaxations have been used to derive bounds on probabilities of sets described using polynomial inequalities [8, 22]. In practice, however, this probability is often estimated using Monte Carlo (MC) simulation, which may require many samples to obtain reliable estimates, if $\Pr(d(\mathbb{X}) > \varrho)$ is small. Instead, we partition the uncertainty set and, using an affine policy for y, prune the sampling region.

In Sect. 2, we use moments of the underlying distribution to upper bound $\Pr(\min_y f(\mathbb{X}, y) > \varrho)$. Here, we relate affine and polynomial policies for y to relaxation techniques in nonlinear programming. In the rest of the paper, we assume that $\mathbb{X}$ is a binary vector. We show that the m-level relaxation from the reformulation–linearization technique (RLT) computes the probability exactly and improve the bound from lower-level relaxations using a concave-envelope construction algorithm. In Sect. 3 and onwards, we assume that each $\mathbb{X}_i$ is a Bernoulli random variable. We partition $[0, 1]^m$ into finitely many polytopes, referred to as *low-weight polytopes* (LWPs), where each defining constraint has small coefficients. We then utilize recent advances in sparsification [13, 33] to develop techniques that estimate the probability that $\mathbb{X}$ lies in a *sliced low-weight polytope* (SLWP), an intersection of an LWP with a general inequality. Then, we use indicators of SLWPs to improve the bound from the concave-envelope construction algorithm. In Sect. 4, we outer-approximate $\{x \in \mathcal{L} : \min_{y \in Y(x)} f(x, y) > \varrho\}$, where $\mathcal{L}$ is an LWP, with a union of SLWPs using the robust counterpart and chance-constrained optimization techniques. We utilize approximate sampling techniques from [12] to estimate the probability that $\mathbb{X}$ lies in such a union and devise an approximate sampler for this set. To our knowledge, this gives the first polynomial-time approximation technique for PQ when y is restricted to an affine policy. This is useful, for example, in NR where quick network response dictates that y is anyway restricted to an affine policy and prior probability estimations still relied on branch-and-bound methods [9]. The NR problem is analyzed in Sect. 5. Finally, in Sect. 5.1, we evaluate our algorithms on NR instances.

Notation: For $m \in \mathbb{N}$, we represent $\{1, \dots, m\}$ as $[m]$; $\text{Vert}(\mathscr{P})$: vertex set of a polytope $\mathscr{P}$; For $\mathcal{F} \subseteq \mathbb{R}^m$, $\mathbb{1}_{\mathcal{F}}(x)$ represents the indicator of $\mathcal{F}$; $\Pr(\mathbb{X} \in \mathcal{F})$ is abbreviated as $\Pr(\mathcal{F})$; $\mathbb{Z}^+$ denotes the set of positive integers.

2 Upper bounding using concave overestimator

In this section, we are interested in the PQ problem, *i.e.,*we estimate $\Pr_*(d(\mathbb{X}) = \min_y f(\mathbb{X}, y) > \varrho)$, where $y \in \mathbb{R}^n, \varrho \in \mathbb{R}$ and $\mathbb{X} \in \mathbb{R}^m$ is a random variable realizing values from a given polytope $\mathscr{P}$ defined as in (1), with any given probability distribution (denoted by subscript $*$), supported over $\mathscr{P}$.

$$\mathscr{P} = \{x \in \mathbb{R}^m : \mathfrak{C}x \leq \mathfrak{d}\}, \text{ where } \mathfrak{C} \in \mathbb{R}^{l \times m} \text{ and } \mathfrak{d} \in \mathbb{R}^l. \tag{1}$$

If $\hat{d} : \mathbb{R}^m \to \mathbb{R}$ is concave, overestimates $d(\cdot)$, and is non negative, we can overestimate $\Pr_*\big(d(\mathbb{X}) > \varrho\big)$ as follows [18]:

$$\Pr_*(d(\mathbb{X}) > \varrho) \leq \Pr_*(\hat{d}(\mathbb{X}) \geq \varrho) \leq \frac{\mathbb{E}_*[\hat{d}(\mathbb{X})]}{\varrho} \leq \frac{\hat{d}(\mathbb{E}_*[\mathbb{X}])}{\varrho} \tag{2}$$

for $\mathbb{E}_*[\mathbb{X}] = (\mathbb{E}_*[\mathbb{X}_i])_{i=1}^m$, where, the first inequality is because $\hat{d}(\mathbb{X}) \geq d(\mathbb{X})$, the second inequality is by Markov's inequality, and the third inequality is by concavity of $\hat{d}(\cdot)$. Let $\mathcal{F} = \{x \in \mathscr{P} : d(x) > \varrho\}$ then, $\Pr_*(d(\mathbb{X}) > \varrho) = \Pr_*(\mathbb{1}_{\mathcal{F}}(\mathbb{X}) \geq 1)$. If $\hat{\mathbb{1}}(\mathbb{X})$ is a concave overestimator of $\mathbb{1}_{\mathcal{F}}(\mathbb{X})$, then from (2),

$$\Pr_*(d(\mathbb{X}) > \varrho) \leq \hat{\mathbb{1}}(\mathbb{E}_*[\mathbb{X}]). \tag{3}$$

We next, discuss ways to derive $\hat{d}(\cdot)$ and $\hat{\mathbb{1}}(\cdot)$. We argue that convex relaxations of the robust problem, $\max_{x \in \mathscr{P}} \min_y f(x, y)$ yield the desired concave overestimators of $\min_y f(x, y)$. In PQ, we use $\mathbb{X}$ to emphasize that it is random, while in the robust problem, we use x instead. For the PQ problem defined above, we assume $f(x, y)$ is linear in (x, y) and that its domain is restricted to $\{(x, y) \mid y \in Y(x)\}$, for $Y(x)$ defined as:

$$Y(x) = \big\{y \in \mathbb{R}^n : Ay + Bx \leq_{\mathbb{K}} c\big\}, \tag{4}$$

where $A \in \mathbb{R}^{p \times n}, B \in \mathbb{R}^{p \times m}, c \in \mathbb{R}^p$, $\mathbb{K}$ is a closed convex pointed cone in $\mathbb{R}^p$, and $a \leq_{\mathbb{K}} b$ implies $b - a \in \mathbb{K}$. Let $K' = \{(x, y, y', \lambda) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \mid y' \geq \mathfrak{h}(x, y, \lambda), \lambda \geq 0, Ay + Bx \leq_{\mathbb{K}} c\lambda\}$, $\mathfrak{h}(x, y, \lambda) = \lambda f\big(\frac{x}{\lambda}, \frac{y}{\lambda}\big)$, and $\mathfrak{h}(x, y, 0) = \lim_{\lambda \downarrow 0} f\big(\frac{x}{\lambda}, \frac{y}{\lambda}\big)$. Our assumption that $f(x, y)$ is linear is without loss of generality (wlog) whenever K' is a closed convex pointed cone. Indeed, for $f(x, y)$ whose epigraph is nonempty, closed, convex and does not contain vertical lines, it follows easily that K' is closed and convex (see Theorem 8.2 in [34]). The assumption of linearity of $f(x, y)$ is wlog because we can rewrite $\min_y f(x, y)$ as $\min_{y,y'}\{e'^{\intercal}(y, y') \mid (x, y, y', 1) \geq_{K'} 0\}$, where $e' = (0, \dots, 0, 1) \in \mathbb{R}^{n+1}$. Thus, wlog we can assume $f(x, y) = e^{\intercal} y$ for $e \in \mathbb{R}^n$.

We now discuss how convex relaxations for $\max_{x\in\mathscr{P}} \min_{y\in Y(x)} e^\intercal y$ assist in constructing $\hat{d}(x)$, which can be used in (2) to bound $\Pr_*(\mathcal{F})$. Given $x \in \mathscr{P}$, let $\mathrm{CP}(x)$ be the conic program $\min_y\{e^\intercal y \mid Ay \leq_{\mathbb{K}} c - Bx\}$ and $\mathrm{CD}(x)$ its dual $\max_w\{w^\intercal(c - Bx) \mid w^\intercal A = e^\intercal, w^\intercal \leq^*_{\mathbb{K}} 0\}$ for $w \in \mathbb{R}^p$. Throughout, we assume that the following holds:

(A1) There is no duality gap between $\mathrm{CP}(\cdot)$ and $\mathrm{CD}(\cdot)$.

We will make additional assumptions as and when required for deriving our results. These assumptions are listed in Appendix A for ease of reference.

See Theorem 1.4.2 in [5] for conditions when Assumption (A1) holds. Specifically, if $\mathbb{K} = \mathbb{R}^p_+$, $\mathrm{CP}(\cdot)$ and $\mathrm{CD}(\cdot)$ exhibit no duality gap if either of them is feasible.

Using duality, we rewrite $d(x) = \min_{y\in Y(x)} e^\intercal y$, as $d(x) = \max_w\{w^\intercal(c - Bx) \mid w^\intercal A = e^\intercal, w^\intercal \leq^*_{\mathbb{K}} 0\}$. Let (R) be any convex relaxation of $\max_{w,x}\{w^\intercal(c - Bx) \mid w^\intercal A = e^\intercal, w^\intercal \leq_{\mathbb{K}^*} 0, x \in \mathscr{P}\}$ in an extended space (x, w, W) where, for each feasible $(\bar{x}, \bar{w})$, there is a $(\bar{x}, \bar{w}, \bar{W})$ feasible to (R) such that the objective evaluates to $\bar{w}^\intercal(c - B\bar{x})$ or higher. By partially maximizing (R) with (w, W), we obtain $\hat{d}(x)$ that is concave (see for example, Proposition 2.22(a) in [36]). If $\varrho > 0$ and $d(x)$ takes negative values, we instead dualize $\min_y \max\{0, f(x, y)\}$ to construct $\hat{d}(x)$. Below we discuss the construction of one such (R) as obtained in (5) using the reformulation linearization technique (RLT).

RLT, which is a commonly used relaxation technique for nonlinear programs, can be used to construct $\hat{d}(\cdot)$. Affine policies, where y is restricted to $P^\intercal x + q$, for $P \in \mathbb{R}^{m\times n}$ and $q \in \mathbb{R}^n$ are used to relax $\max_{x\in\mathscr{P}} \min_{y\in Y(x)} f(x, y)$ [4, 7] and their connection to RLT has been explored in [10, 19, 42]. We review this relation (and extend it to allow for conic inequalities in (4)) so as to derive $\hat{d}(x)$ and, thereby, bound $\Pr_*(\mathcal{F})$. Let $\Gamma^* = \max_{x\in\mathscr{P}} \min_{y\in Y(x)} f(x, y)$ and Ψ be the restriction of this optimization problem where $y(x) = P^\intercal x + q$. Let $\Psi_1 = \min_{P,q} \Psi$, then observe that $\Gamma^* \leq \Psi_1$. We use RLT to relax $\max_{x\in\mathscr{P}} d(x)$ as follows:

$$
\begin{aligned}
(\mathrm{R}): \max_{w,W,x} \quad & w^\intercal c - \mathrm{Tr}(BW) && \text{(5a)}\\
& WA = xe^\intercal && \text{(5b)}\\
& \mathfrak{d}_r w^\intercal - \mathfrak{C}_r W \leq_{\mathbb{K}^*} 0 \quad \forall r \in [l] && \text{(5c)}\\
& w^\intercal A = e^\intercal && \text{(5d)}\\
& \mathfrak{C}x \leq \mathfrak{d} && \text{(5e)}\\
& w^\intercal \leq_{\mathbb{K}^*} 0, && \text{(5f)}
\end{aligned}
$$

where we have exploited Assumption (A1) to dualize the inner problem as $\mathrm{CD}(x)$: $\max_w\{w^\intercal(c - Bx) \mid w^\intercal A = e^\intercal, w^\intercal \leq_{\mathbb{K}^*} 0\}$ and introduced W to linearize $xw^\intercal$. Constraints (5d) and (5f) are from $\mathrm{CD}(x)$, Constraint (5e) models $\mathscr{P}$, Constraint (5b) is obtained by pre-multiplying $w^\intercal A = e^\intercal$ with x, and Constraints (5c) are obtained by post-multiplying $\mathfrak{C}x \leq \mathfrak{d}$ with $w^\intercal \leq_{\mathbb{K}^*} 0$. The formulation Γ^O below is then obtained by dualizing (R).

Proposition 1 *Let Γ^O be obtained by dualizing (R) as in (5), then Γ^O overestimates Ψ_1 and thus Γ^*, where*

$$\Gamma^O = \min_{q,U,P,\underline{\Theta}} \quad e^\intercal q + \mathfrak{d}^\intercal \underline{\Theta} \tag{6a}$$

$$AP^\intercal - U^\intercal \mathfrak{C} = -B \tag{6b}$$

$$U^\intercal \mathfrak{d} + Aq \leq_{\mathbb{K}} c \tag{6c}$$

$$-Pe + \mathfrak{C}^\intercal \underline{\Theta} = 0 \tag{6d}$$

$$\underline{\Theta} \geq 0, U_r \geq_{\mathbb{K}} 0 \qquad \forall r \in [l], \tag{6e}$$

$P \in \mathbb{R}^{m\times n}$, $q \in \mathbb{R}^n$, $\underline{\Theta} \in \mathbb{R}^l$, and $U \in \mathbb{R}^{l\times p}$ is such that for all $r \in [l]$, U_r denotes its row vector. Moreover, the above problem finds the optimal affine policy when the uncertainty set $\mathscr{P}$ is non empty and $\mathbb{K} = \mathbb{R}^p_+$. □

In general, Γ^* may be infinite. For example, if there is an x such that $Y(x)$ is empty, then $\Gamma^* = \infty$. Otherwise, $Y(x)$ is non-empty for all x and $\Gamma^* < \infty$. This is usually referred to as complete recourse [40]. If in addition, there is a $(w, \bar{x})$ such that $w^\intercal A = e^\intercal$, $w^\intercal \leq_{\mathbb{K}^*} 0$, and $\bar{x} \in \mathscr{P}$, then it follows by weak duality that $\Gamma^* \geq \min_{y\in Y(\bar{x})} e^\intercal y \geq w^\intercal(c - B\bar{x}) > -\infty$, and so Γ^* is finite.

In Appendix C, we relate polynomial policies for y to higher levels of RLT hierarchy. These policies yield better candidates for $\hat{d}(\cdot)$, when bounding $\Pr_*(d(\mathbb{X}))$. In Sect. 2.1, we derive relaxations for $\hat{\mathbb{1}}(\cdot)$ instead and show that higher levels of the hierarchy yield better bounds via (3).

2.1 Better bounds by lifting indicator function using functions

We construct a formulation for $\mathbb{1}_{\mathcal{F}}(x)$ as follows. Let $\mathcal{C}(x) = \{y \in \mathbb{R}^n : e^\intercal y \leq \varrho, Ay + Bx \leq_{\mathbb{K}} c\}$, and define $\Xi(x) = \{(y, \Phi) : \Phi = 0, y \in \mathcal{C}(x)\} \cup \{\Phi = 1, y = 0\}$, where $\Phi \in \mathbb{R}$. Then, let $\Delta(x) = \{(y, \Phi) : Ay + Bx(1 - \Phi) \leq_{\mathbb{K}} (1 - \Phi)c, e^\intercal y \leq (1 - \Phi)\varrho, \Phi \geq 0\}$, where $\Xi(x) \subseteq \Delta(x) \subseteq \mathrm{cl}(\mathrm{conv}(\Xi(x)))$, see Proposition 3.3.5 in [5]. Then, let $\mathcal{I}(x) = \min_{y,\Phi}\{\Phi \mid (y, \Phi) \in \Delta(x)\}$. We show that $\mathcal{I}(x)\mathbb{1}_{\mathscr{P}}(x) = \mathbb{1}_{\mathcal{F}}(x)$. First, we show that $\mathcal{I}(x) = 1$ if $\mathcal{C}(x) = \emptyset$ and 0 otherwise. Assume there exists $\underline{y} \in \mathcal{C}(x)$. Then, $(\underline{y}, 0) \in \Delta(x)$ and $\mathcal{I}(x) = 0$. Now, assume that $\mathcal{C}(x) = \emptyset$. Then, there does not exist $(\underline{y}, \underline{\Phi}) \in \Delta(x)$ for $\underline{\Phi} < 1$. Otherwise, $\left(\frac{\underline{y}}{(1-\underline{\Phi})}, 0\right) \in \Delta(x)$, which in turn implies that $\frac{\underline{y}}{(1-\underline{\Phi})} \in \mathcal{C}(x)$ and contradicts that $\mathcal{C}(x)$ is empty. Since $(0, 1) \in \Delta(x)$, it follows that $\mathcal{I}(x) = 1$. Recall that $x \in \mathcal{F}$ if and only if $x \in \mathscr{P}$ and $d(x) > \varrho$. The latter condition is equivalent to $\mathcal{C}(x) = \emptyset$. This shows that $\mathcal{I}(x)\mathbb{1}_{\mathscr{P}}(x) = \mathbb{1}_{\mathcal{F}}(x)$.

For the remainder of this paper, we assume that

(A2) The distribution of $\mathbb{X}$ is supported on a finite set of points $\mathscr{T}$ in $\mathscr{P}$.

To improve the probability bound, we use the concave envelope of $\mathbb{1}_{\mathcal{F}}(x)$ restricted to $\mathscr{T}$ over $\mathrm{conv}(\mathscr{T})$, which we denote as $\hat{\mathbb{1}}_E(\cdot)$. In doing so, we utilize that for computing $\Pr_*(\mathcal{F})$, we can limit attention to $\mathcal{F} \cap \mathscr{T}$ exploiting that the distribution is supported only on $\mathscr{T}$. For $x_0 \in \mathrm{conv}(\mathscr{T})$, $\hat{\mathbb{1}}_E(x_0) = \min_{a,b}\{a^\intercal x_0 + b \mid a^\intercal x^i + b \geq$

$\mathcal{I}(x^i)\ \forall x^i \in \mathscr{T}\}$, where $a, x_0, x^i \in \mathbb{R}^m$, and $b \in \mathbb{R}$ [38]. Unfortunately, the number of constraints depends on the cardinality of the support set $|\mathscr{T}|$, which can be large. For example, if $\mathscr{T} = \{0, 1\}^m$, there are 2^m constraints. Regardless, numerical experiments show that this bound can still be weak. It can, however, be improved by lifting $\mathscr{T}$ to a higher dimensional space. Additional variables improve the bound in (3) since more information about the probability distribution is captured in $\mathbb{E}_*[\mathbb{X}]$. In particular, if $\mathscr{T}$ are the vertices of the simplex, (3) is tight.

Proposition 2 *If $\mathscr{T}$ are the vertices of a simplex, $\hat{\mathbb{1}}_E(\mathbb{E}_*[\mathbb{X}]) = Pr_*(\mathcal{F})$.* □

Assume we have available expected values of a set of functions $\{\mathfrak{f}_\alpha(\mathbb{X})$ for $\alpha \in \bar{\Gamma} \subseteq \mathbb{N}^m\}$. Then, we determine $a \in \mathbb{R}^{|\bar{\Gamma}|}$ and $b \in \mathbb{R}$ so that $\mathfrak{f}(\mathbb{X}) := b + \sum_{\alpha \in \bar{\Gamma}} a_\alpha \mathfrak{f}_\alpha(\mathbb{X}) \geq \mathbb{1}_{\mathcal{F}}(\mathbb{X})$. Clearly, $\Pr_*(\mathcal{F}) = \mathbb{E}_*[\mathbb{1}_{\mathcal{F}}(\mathbb{X})] \leq \sum_{\alpha \in \bar{\Gamma}} a_\alpha \mathbb{E}_*[\mathfrak{f}_\alpha(\mathbb{X})] + b$, and the best such estimate is:

$$\min_{b,a} \left\{ \sum_{\alpha \in \bar{\Gamma}} a_\alpha \mathbb{E}_*[\mathfrak{f}_\alpha(\mathbb{X})] + b \,\middle|\, b + \sum_{\alpha \in \bar{\Gamma}} a_\alpha \mathfrak{f}_\alpha(z) \geq \mathbb{1}_{\mathcal{F}}(z)\ \forall z \in \mathscr{T} \right\}. \tag{7}$$

Here onwards, we limit our consideration to the case where $Y(x)$ is defined using linear inequalities. In particular, we assume that:

(A3) $\mathbb{K} = \mathbb{R}^p_+$.

We define $\mathcal{DI}(x)$ by dualizing the formulation for $\mathcal{I}(x)$, and observe that, for all x, $\mathcal{DI}(x) = \mathcal{I}(x)$ since $(w, v, \varphi) = (0, 0, 0)$ is feasible in the following:

$$\mathcal{DI}(x) = \max_{w,v,\varphi} \left\{ \varphi \mid w^\intercal A + ve^\intercal = 0,\ \varphi \leq w^\intercal (c - Bx) + v\varrho \leq 1,\ w, v \leq 0 \right\}, \tag{8}$$

where $w \in \mathbb{R}^p$ and $v \in \mathbb{R}$. Let $r(x, w, v) = \min\{w^\intercal(c - Bx) + v\varrho, 1\}$ if $(x, w, v) \in \mathscr{T} \times \mathcal{S}$, where $S = \{(w, v) \in \mathbb{R}^p_- \times \mathbb{R}_- : w^\intercal A + ve^\intercal = 0\}$, and $-\infty$ otherwise. Let $\underline{\mathscr{P}}$ be the problem $\max_{x,\underline{\varphi},v,w}\{\underline{\varphi} \mid \underline{\varphi} \leq h(x, w, v), (x, w, v) \in \mathrm{conv}(\mathscr{T}) \times \mathcal{S}\}$, where $h(x, w, v)$ is the concave envelope of $r(x, w, v)$ over $\mathrm{conv}(\mathscr{T}) \times \mathcal{S}$. Let $\hat{\mathbb{1}}(x) = \max_{v,w,\underline{\varphi}}\{\underline{\varphi} \mid \underline{\varphi} \leq h(x, w, v), (x, w, v) \in \mathrm{conv}(\mathscr{T}) \times \mathcal{S}\}$ *i.e.*, $\hat{\mathbb{1}}(x)$ is obtained by partially maximizing $\underline{\mathscr{P}}$ w.r.t $(w, v, \underline{\varphi})$. We show next that $\hat{\mathbb{1}}(\mathbb{E}_*[\mathbb{X}])$ is the bound in (3), obtained by using the concave envelope of $\mathbb{1}_{\mathcal{F}\cap\mathscr{T}}(\cdot)$ over $\mathrm{conv}(\mathscr{T})$.

Proposition 3 *Let $\hat{\mathbb{1}}(x)$ be the function obtained by partial maximization of $\underline{\mathscr{P}}$ w.r.t variables $(w, v, \underline{\varphi})$ and $\hat{\mathbb{1}}_E(x)$ be the concave envelope of $\mathbb{1}_{\mathcal{F}\cap\mathscr{T}}(\cdot)$ over conv$(\mathscr{T})$. Then, for all $x \in$ conv$(\mathscr{T})$, $\hat{\mathbb{1}}(x) = \hat{\mathbb{1}}_E(x)$.* □

As before, $\hat{\mathbb{1}}_E(x) = \min_{a,b}\{a^\intercal x + b \mid a^\intercal x^J + b \geq \mathcal{I}(x^J) = \mathcal{DI}(x^J), \forall x^J \in \mathscr{T}\}$. We briefly describe below a column generation algorithm that computes $\hat{\mathbb{1}}_E(x)$ [3, 38]. Let $\mathfrak{J} = \{J : x^J \in \mathscr{T}\}$. For $\underline{\mathfrak{J}} \subseteq \mathfrak{J}$, consider the relaxation where $\mathscr{T} = \bigcup_{J \in \underline{\mathfrak{J}}}\{x^J\}$. Then, the corresponding dual is:

$$\max_{\lambda} \left\{ \sum_{J \in \underline{\mathfrak{J}}} \lambda_J \mathbb{1}_{\mathcal{F}}(x^J) \,\middle|\, \sum_{J \in \underline{\mathfrak{J}}} \lambda_J x^J = x, \sum_{J \in \underline{\mathfrak{J}}} \lambda_J = 1, \lambda_J \geq 0\ \forall J \in \underline{\mathfrak{J}} \right\}. \tag{9}$$

To compute $\hat{\mathbb{1}}_E(x)$, let (a, b) be optimal dual solution to (9). Then, we find a $J \in \mathfrak{J} \setminus \underline{\mathfrak{J}}$, such that the reduced cost $\mathbb{1}_{\mathcal{F}}(x^J) - (a^\intercal x^J + b)$ of λ_J is positive and add J to $\underline{\mathfrak{J}}$. We use this algorithm in Sect. 3.1 to compute $\hat{\mathbb{1}}_E(x)$. As in (3), we will use $\hat{\mathbb{1}}_E\big(\mathbb{E}_*(\mathbb{X})\big)$ to overestimate $\Pr_*(\mathcal{F})$. However, inspired by Proposition 2, we will instead use (7) to compute a tighter bound by first lifting x to $\{\mathfrak{f}_\alpha(x) : \alpha \in \bar{\Gamma}\}$, which requires the choice of functions $\mathfrak{f}_\alpha$. We will, here onwards, make the following assumption.

(A4) $\mathscr{T} \subseteq \{0, 1\}^m$ and an inequality description of $\operatorname{conv}(\mathscr{T})$ is available.

For $\mathfrak{f}_\alpha$, we will use multilinear functions next and indicator functions of certain polytopes in Sect. 3.1. The bound in (7) requires $\mathbb{E}_*[\mathfrak{f}_\alpha(\mathbb{X})]$. Unfortunately, a naive computation is expensive since $|\mathscr{T}|$ may be exponential in the size of the problem. For example, $[0, 1]^m$ has 2^m extreme points. However, in Sect. 3, we will approximate $\mathbb{E}_*[\mathfrak{f}_\alpha(\mathbb{X})]$ in polynomial time for certain distributions Θ.

Now, we describe how to use multilinear functions as $\mathfrak{f}_\alpha$ so as to lift $\mathscr{T}$ into a higher-dimensional space. This idea is related to the use of moments to compute bounds on probability [23], where, invoking Putinar's Positivstellensatz [24], sum-of-squares of polynomials are used overestimate the indicator function. Instead, we use multilinear functions, exploiting that $\mathscr{T} \subseteq \{0, 1\}^m$. Let $\mathscr{T} = \{x^j\}_{j\in[s]}$. It is easy to see that an arbitrary function over $\mathscr{T}$ can be written as a multilinear function. For $J \subseteq [m]$, let $\mathfrak{X}^J \in \{0, 1\}^m$, be such that $\mathfrak{X}^J_i = 1$ if $i \in J$ and 0 otherwise. We associate with J a multilinear function, $\mathfrak{M}_J(x) = \prod_{i\in J} x_i \prod_{i\in J^C}(1 - x_i)$, where $J^C = [m]\setminus J$. Then, $\mathfrak{M}_J(x) = 0$ for all $x \in \{0, 1\}^m \setminus \{\mathfrak{X}^J\}$ and $\mathfrak{M}_J(\mathfrak{X}^J) = 1$. Therefore, $\mathbb{1}_{\mathcal{F}}(x) = M(x) := \sum_{J:\mathfrak{X}^J\in\mathcal{F}} \mathfrak{M}_J(x)$, a multilinear representation of the indicator function. If $M(x) = \sum_{\alpha\in\{0,1\}^m} g_\alpha x^\alpha$, we have $\mathbb{E}_\Theta[\mathbb{1}_{\mathcal{F}}(\mathbb{X})] = \sum_{\alpha\in\{0,1\}^m} g_\alpha b_\alpha$. Now, let $\bar{x} \in \mathscr{T}$. It follows from $\mathcal{DI}(x)$ that $\bar{x} \in \mathcal{F}$ if and only if there exists $(w_{\bar{x}}, v_{\bar{x}})$ feasible to (8) such that $w_{\bar{x}}^\intercal(c - B\bar{x}) + v_{\bar{x}}\varrho > 0$. We write $\Pr_\Theta(\mathcal{F}) = \mathbb{E}_\Theta[\mathbb{1}_{\mathcal{F}}(\mathbb{X})] = \sum_{x^i\in\mathscr{T}} \mathbb{1}_{\mathcal{F}}(x^i)\Pr_\Theta(\mathbb{X} = x^i)$ or

$$\Pr_\Theta(\mathcal{F}) = \max_{\{v^i, w^i, \varphi^i\}_i} \sum_{i=1}^{s} \varphi^i \tag{10a}$$

$$\varphi^i \le w^{i\intercal}(c - Bx^i) + v^i\varrho \le \Pr_\Theta(\mathbb{X} = x^i) \qquad \forall i \in [s] \tag{10b}$$

$$(w^i)^\intercal A + v^i e^\intercal = 0 \qquad \forall i \in [s] \tag{10c}$$

$$w^i, v^i \le 0 \qquad \forall i \in [s], \tag{10d}$$

where $w^i \in \mathbb{R}^p$ and $v^i \in \mathbb{R}$ $\forall i \in [s]$. For $J_i \subseteq [m]$:

$$\Pr_\Theta(\mathbb{X} = \mathfrak{X}^{J_i}) = \mathbb{E}_\Theta\left[\prod_{j\in J_i} x^i_j \prod_{j\in J_i^C}(1 - x^i_j)\right] = \sum_{J'\subseteq J_i^C} (-1)^{|J'|} b_{\alpha(J_i\cup J')}, \tag{11}$$

where $\alpha(J)$ is the indicator vector of $J \subseteq [m]$.Conversely, given Θ, $\mathbb{E}_\Theta[\mathbb{X}^\alpha]$ can be computed as $\sum_{J\subseteq\{j:\alpha_j=1\}} \Pr_\Theta(\mathbb{X} = \mathfrak{X}^J)$. Therefore, we write (10) as:

$$\Pr_\Theta(\mathcal{F}) = \max_{w,v,\mathfrak{p},\varphi} \sum_{J\subseteq[m]} \varphi^J \tag{12a}$$

$$\varphi^J \le (w^J)^\intercal(c - B\mathfrak{X}^J) + v^J \varrho \le \mathfrak{p}_J \qquad \forall J \subseteq [m] \tag{12b}$$

$$(w^J)^\intercal A + v^J e^\intercal = 0 \qquad \forall J \subseteq [m] \tag{12c}$$

$$w^J, v^J \le 0 \qquad \forall J \subseteq [m] \tag{12d}$$

$$\sum_{J\subseteq[m]} \mathfrak{p}_J (\mathfrak{X}^J)^\alpha = b_\alpha \qquad \forall \alpha \in \{0,1\}^m, \tag{12e}$$

where (12e) constrains $\mathfrak{p}_J$ to be $\Pr_\Theta(\mathbb{X} = \mathfrak{X}^J)$ for $J \subseteq [m]$. In fact, (12) is related to the m^{th} level RLT relaxation of $\max_{x\in\mathscr{T}} \mathcal{DI}(x)$, where the latter expands and linearizes the following formulation after substituting $x_i^2 = x_i$:

$$\max_{w,v,x,\varphi} \quad \varphi \tag{13a}$$

$$\varphi\mathfrak{M}_J(x) \le \mathfrak{M}_J(x)(w^\intercal(c - Bx) + v\varrho) \le \mathfrak{M}_J(x) \qquad \forall J \subseteq [m] \tag{13b}$$

$$\mathfrak{M}_J(x)(w^\intercal A + ve^\intercal) = 0 \qquad \forall J \subseteq [m] \tag{13c}$$

$$\mathfrak{M}_J(x)w, \mathfrak{M}_J(x)v \le 0 \qquad \forall J \subseteq [m] \tag{13d}$$

$$\mathfrak{M}_J(x)x \in \mathfrak{M}_J(x)\,\mathrm{conv}(\mathscr{T}) \qquad \forall J \subseteq [m]. \tag{13e}$$

Theorem 1 *Let φ^R be the m^{th} level RLT relaxation obtained by expanding expressions in (13), substituting x_i^2 with x_i for all $i \in [m]$, and linearizing the monomials x^α, for $\alpha \in \{0,1\}^m$ as x_α. Let $\varphi^R(b_\alpha)$ be the maximum value of the RLT relaxation when x_α are fixed to b_α and optimization is performed with respect to the remaining variables. Then, $\varphi^R(b_\alpha) = Pr_\Theta(\mathcal{F})$.* □

3 Probability estimation via weighted counting

Here, we develop techniques to compute $\mathbb{E}_\Theta[\mathfrak{f}_\alpha(\mathbb{X})]$, where $\mathfrak{f}_\alpha$ could be a multilinear function or the indicator of a special polytope, which we define below.

Definition 1 A polytope $\mathcal{L} = \{x \in [0,1]^m : \mathcal{A}x \le b\}$ for $\mathcal{A} \in \mathbb{Z}^{t\times m}, b \in \mathbb{Z}^t$ is a *low weight polytope* (LWP), if t is a constant, and entries in $\mathcal{A}$, b are bounded by polynomials in m. We refer to the constraints given by $\mathcal{A}x \le b$ as low-weight constraints.

Definition 2 Given a *general inequality* $\sum_{i=1}^m w_i x_i \ge C$, for $C, \{w_i\}_{i=1}^m \in \mathbb{Z}$, and a LWP, $\mathcal{L} \subseteq \mathbb{R}^m$, we define a *Sliced low weight polytope* (SLWP) as $\{x \in \mathbb{R}^m : \sum_{i=1}^m w_i x_i \ge C, x \in \mathcal{L}\}$, where $\mathcal{L}$ is referred to as the underlying LWP.

Here onwards, we will assume that:

(A5) $\mathscr{P} = \text{conv}(\mathscr{T}) = [0,1]^m$ and $\mathbb{X} \in \{0,1\}^m$, with distribution $\Theta = \bigotimes_{i=1}^m \text{Bernoulli}(p_i)$ (tensor product of m independent Bernoulli distributions). Moreover, we assume that $p_i = \frac{a_i}{n_i}$, where $a_i, n_i \in \mathbb{N}$, and $\text{GCF}(a_i, n_i) = 1$.

In fact, we will study $\Pr_\Theta(\mathcal{F})$ by first partitioning $[0,1]^m$ into LWPs, $\{\mathcal{L}^{\mathfrak{b}}\}_{\mathfrak{b}=1}^F$. Then, we will obtain $\Pr_\Theta(\mathcal{F})$ indirectly by quantifying $\Pr_\Theta(\mathcal{F} \cap \mathcal{L}^{\mathfrak{b}})$ instead. For concreteness, for $\mathfrak{b} = 1, \ldots, F$, consider $\mathcal{L}^{\mathfrak{b}} = \{x \in [0,1]^m : \sum_{j=1}^m x_j = \mathfrak{b}\}$ as the LWPs that partition $[0,1]^m$. Then, we write $\Pr_\Theta(\mathcal{F}) = \sum_{\mathfrak{b}=1}^F \Pr_\Theta(\mathcal{F} \cap \mathcal{L}^{\mathfrak{b}})$. In this section, we will develop techniques to approximate each term on the right-hand-side of the above expression. We will also use overestimates of $\Pr_\Theta(\mathcal{F} \cap \mathcal{L}^{\mathfrak{b}})$ to overestimate $\Pr_\Theta(\mathcal{F})$. To do so, we outer-approximate $\mathcal{F} \cap \mathcal{L}^{\mathfrak{b}}$ using Proposition 1 to derive an affine policy $y = P^{*\mathsf{T}}x + q^*$ where x is restricted to $\mathcal{L}^{\mathfrak{b}}$. Such a policy overestimates $d(x)$ by $f(x, P^{*\mathsf{T}}x + q^*)$ for each x such that $P^{*\mathsf{T}}x + q^* \in Y(x)$. This implies that we can outer-approximate $\mathcal{F} \cap \mathcal{L}^{\mathfrak{b}}$ by a union of SLWPs, derived from (since SLWPs do not allow strict inequalities) $D_0, \ldots, D_p$, where $D_0 = \{x \in \mathcal{L}^{\mathfrak{b}} : f(x, P^{*\mathsf{T}}x + q^*) > \varrho\}$ and, for all $r \in [p]$, $D_r = \{x \in \mathcal{L}^{\mathfrak{b}} : A_r(P^{*\mathsf{T}}x + q^*) + B_r x > c_r\}$. We will explore this in more detail in Sect. 4.

We will compute $\Pr_\Theta(\mathcal{F} \cap \mathcal{L})$, where $\mathcal{L}$ is a LWP as in Definition 1. In particular, we will be interested in a deterministic algorithm, that given $\epsilon_s > 0$, overestimates (resp. underestimates) $\Pr_\Theta(\sum_{i=1}^m w_i \mathbb{X}_i \geq C, \mathbb{X} \in \mathcal{L})$ for $\{w_i\}_{i=1}^m, C \in \mathbb{Z}$ with a relative error of $(1+\epsilon_s)$ (resp. $(1-\epsilon_s)$) in time, that is polynomial in the size of input data and $\frac{1}{\epsilon_s}$. Such an algorithm is a fully polynomial time approximation scheme (FPTAS) for the computation of this probability [16]. We remark that the special case that counts $\{0,1\}^m$ solutions to an inequality $\sum_{j=1}^m \hat{w}_j x_j \leq \hat{C}$, with $\{\hat{w}_i\}_{i=1}^m, \hat{C} \geq 0$ is known to be #P-complete. To see that this counting problem is a special case, let $p_i = \frac{1}{2}$, $\mathcal{L} = [0,1]^m$, $w_i = -\hat{w}_j$, and $C = -\hat{C}$. We leverage recent developments in knapsack counting and counting paths in a directed acyclic graph (DAG) to develop an FPTAS [13, 29, 33]. For any given SLWP, S, we let $|S|_\Theta$ represent the cardinality of its $\{0,1\}$ solutions where, in agreement with Θ, there are a_i ways in which x_i can be one, and $n_i - a_i$ ways in which it is zero. Then, $|S|_\Theta = \sum_{x \in S \cap \{0,1\}^m} \prod_{i:x_i=1} a_i \prod_{i:x_i=0} (n_i - a_i)$, and $\Pr_\Theta(S) = \frac{|S|_\Theta}{\prod_i n_i}$, where a_i and n_i are as in (A5).

(A6) The weights of the general inequality defining each SLWP are non-negative *i.e.*,$\{w_i\}_{i=1}^m \in \mathbb{Z}_{\geq 0}$.

This assumption is without loss of generality (wlog) since if (A6) does not hold, then we may create another instance that satisfies the above conditions by defining $\mathbb{X}_i' = 1 - \mathbb{X}_i$ if $w_i < 0$ and $\mathbb{X}_i' = \mathbb{X}_i$ otherwise. Moreover, with this assumption, we can assume that $C \in \mathbb{Z}^+$. Otherwise, the inequality $\sum_{i=1}^m w_i \mathbb{X}_i \geq C$ is redundant. To keep notation simpler and to fix ideas, we will, at the outset, consider the SLWP, S_s as in (14),

$$S_s = \left\{ x : \sum_{j=1}^m w_j x_j \geq C, x \in \mathcal{L} \right\}, \text{ where } \mathcal{L} = \left\{ x : \sum_{i=1}^{K_2} x_i - \sum_{i=1}^{K_1} x_{K_2+i} = \mathfrak{b} \right\}, \quad (14)$$

$K_1 + K_2 = m$ and $\mathfrak{b} \geq 0$. In order to approximate $|S_s|_\Theta$, we construct a DAG in $\mathbb{Z}^2$, where, $\forall (i, j) \in [m]^2$, there is a vertex associated with each lattice point (i, j). There are arcs that connect $(i-1, j)$ to (i, j) for $i < m$, $(i-1, j-1)$ to (i, j) when $i \leq K_2$, and $(i-1, j+1)$ to (i, j) when $i > K_2$. The arcs from $(i-1, j)$ to (i, j) are associated with a tuple $(0, n_i - a_i)$ while the remaining arcs are associated with (w_i, a_i). Given a function, $f : \mathbb{Z}^+ \to \mathbb{Z}^+$, Gawrychowski et al. [13] construct a $1 + \epsilon_s$ function approximation of $f^{\leq}(\cdot)$, where $f^{\leq}(x) = \sum_{y \leq x} f(y)$. We adapt their definitions and properties [13] to our setting.

Definition 3 Given $f : \mathbb{Z}^+ \to \mathbb{Z}^+$ and an approximation parameter $\epsilon_s > 0$, a function $F : \mathbb{Z}^+ \to \mathbb{Z}^+$ is a $1+\epsilon_s$ function approximation of f if: $f(x) \leq F(x) \leq (1+\epsilon_s) f(x)$ for all x. A function $\bar{F} : \mathbb{Z}^+ \to \mathbb{Z}^+$ is a $1 + \epsilon_s$ sum-approximation of f if, for all x, $f^{\geq}(x) \leq \bar{F}^{\geq}(x) \leq (1 + \epsilon_s) f^{\geq}(x)$, where $f^{\geq}(x)$ (resp. $\bar{F}^{\geq}(x)$) is defined as $\sum_{y \geq x} f(y)$ (resp. $\sum_{y \geq x} \bar{F}(y)$).

Definition 4 Given $f(\cdot)$, and a shifting parameter $h > 0$, the shifting of $f(\cdot)$ by h is defined as $f_{|h}(x)$, where $f_{|h}(x) = f(x-h)$ if $x \geq h$ and 0 otherwise.

Lemma 1 [13] *Given $\epsilon_s > 0$, let F and G be a $(1+\epsilon_s)$ sum-approximations of f and g respectively, then (i) A $(1+\delta_s)$ sum-approximation of F is a $(1+\delta_s)(1+\epsilon_s)$ sum-approximation of f, (ii) $F + G$ is a $(1+\epsilon_s)$ sum-approximation of $f + g$, (iii) $F_{|w}$ is a $(1+\epsilon_s)$ sum-approximation of $f_{|w}$ for any $w > 0$, (iv) αF is a $(1+\epsilon_s)$ sum-approximation of αf.* □

Definition 5 The number of pairs $(x, f(x))$ in the function representation of f is defined as the size of $f(\cdot)$ i.e. $|f(\cdot)|$.

Function sparsification to obtain $\bar{F}^{\geq}$: Given a function $f(\cdot) : \mathbb{Z}^+ \to \mathbb{Z}^+$, and a sparsification parameter $\delta_s > 0$, we construct $\bar{F}$, a $1+\delta_s$ sum-approximation function of f. We partition the values of $f^{\geq}$ into segments $[r_i, r_{i+1})$, where $r_0 = 0$, $r_{i+1} = \max\{r_i + 1, \lfloor (1+\delta_s) r_i \rfloor\}$ for all $i \geq 0$. Let $c_i = \max_c \{c \mid f^{\geq}(c) \geq r_i\}$. For any c, we define $\text{pred}(c) = \max_i \{c_i \mid c_i < c\}$ and define $\bar{F}^{\geq}(c) = f^{\geq}(\text{pred}(c) + 1)$, where $\bar{F}^{\geq}(c) = \lim_{x \to -\infty} f^{\geq}(x)$ if $\text{pred}(c) = -\infty$.

Lemma 2 [13] *Given a sparsification parameter $\delta_s > 0$, $\bar{F}^{\geq}$ constructed by the above procedure is a $(1+\delta_s)$ function approximation of $f^{\geq}$.* □

Since the sequence of values in r grows by a factor $1+\delta_s$ each time, it follows that $|\bar{F}|$ is no more than $\log_{1+\delta_s} M$, where M is the largest value of $f^{\geq}$. In order to approximate $|S_s|_\Theta$, for all (i, j), we store a sparsified version of $s((i, j), *)$ defined below. Given a non negative integer, $\tilde{c}$ and (i, j), $s((i, j), \tilde{c})$ is the total number of directed acyclic paths (DAPs) from $(0, 0)$ to (i, j) of total path weight equal to $\tilde{c}$. Then, $s((i, j), \tilde{c})$ satisfies the recursion:

$$s((i,j),\tilde{c}) = \begin{cases} \left| \left\{ (x)_{l=1}^{i} : \sum_{l=1}^{i} w_l x_l = \tilde{c},\ \sum_{l=1}^{K_2} x_l - \sum_{l=K_2+1}^{i} x_l = j \right\} \right|_\Theta & i > K_2 \\ \left| \left\{ (x)_{l=1}^{i} : \sum_{l=1}^{i} w_l x_l = \tilde{c},\ \sum_{l=1}^{i} x_l = j \right\} \right|_\Theta & i \leq K_2. \end{cases} \tag{15}$$

Moreover, for any (i, j) and $\tilde{c}$, we define $s^{\geq}((i, j), \tilde{c}) = \sum_{y \geq \tilde{c}} s((i, j), y)$. Given $\delta_s > 0$, leveraging the sparsification procedure above, we obtain $\tilde{s}^{\geq}((i, j), \cdot)$, as a $(1 + \delta_s)^{i-1}$ function approximation of $s^{\geq}((i, j), \cdot)$.

Theorem 2 *Given S_s, Θ as in (14), (A5) respectively, and an error parameter $\epsilon_s \in (0, 1)$, we can deterministically compute a $(1 + \epsilon_s)$ relative error approximation of $|S_s|_\Theta$, in time $O\big(\epsilon_s^{-1}\underline{\Theta}m^2\big(\xi \ln(m/\xi) + m \ln \mathcal{T}\big)\big)$, where $\mathcal{T} \geq \max_i n_i$, $\xi = \min\big\{K_2, K_1 + \mathfrak{b}, \frac{m}{2}\big\}$ and $\underline{\Theta} = \min\{K_2 - \mathfrak{b}, K_1 + \mathfrak{b}\}$.* □

Similar techniques yield a $(1 - \epsilon_s)$ relative error approximation of $|S_s|_\Theta$; see Appendix I. We extend Theorem 2 to SLWP, $S_{t+1} = \{x : \sum_{j=1}^m w_j x_j \geq C, x \in \mathcal{L}_t\}$, where $\mathcal{L}_t$, is the underlying LWP is described by t equality constraints, with $\{w_i^l\}_{i=1}^m$ denoting the weights of the l^{th} $(l \in [t])$ constraint.

Theorem 3 *Consider the SLWP, S_{t+1} described above. Let the first t constraints describing S_{t+1} correspond to $\mathcal{L}_t$, and the $(t + 1)^{st}$ be a general inequality. Given an error parameter $\epsilon_s \in (0, 1)$, and Θ as in (A5), there exists an FPTAS which deterministically computes a $(1 + \epsilon_s)$ relative error approximation of $|S_{t+1}|_\Theta$.* □

Let γ be as defined in the proof of Theorem 3 so that $\gamma \geq \max_k w_k^l - \min_k w_k^l$ for all $l \in [t]$. Then, we can relax the requirement that the constraints defining $\mathcal{L}_t$ are equality constraints. This is because, there are at most $(m\gamma)^t$ possible nodes in the last slice of the graph constructed in the proof of Theorem 3. Since, w_k^l for $k \in [m]$ and $l \in [t]$ were assumed to be polynomial in m and t is a constant, the number of nodes is polynomial in the problem input. Therefore, we can run the algorithm of Theorem 3 on each of these nodes which satisfy $\mathcal{L}_t$. In other words, we can extend Theorem 3 to handle inequality constraints. Here onwards, whenever we create sparsified function approximations, we will refer to Theorem 3, although a better time complexity can be obtained for this construction using Theorem 2, when the underlying SLWP is of the form (14).

We now describe an algorithm to approximately sample $\{0, 1\}^m$ solutions according to distribution Θ, from a given SLWP, $S_\Omega = \big\{x : \sum_{i=1}^m w_i x_i \geq C, x \in \mathcal{L}_\Omega\big\}$, where the underlying LWP, $\mathcal{L}_\Omega = \{x : \Omega x \leq C_\Omega\}$ for $\Omega \in \mathbb{Z}^{t \times m}$ and $C_\Omega \in \mathbb{Z}^t$. Later, we will leverage this algorithm in Sect. 4.1 to obtain a randomized approximation scheme for computing the probability of a union of SLWPs, which in turn will be constructed to overapproximate $\Pr_\Theta(\mathcal{F} \cap \mathcal{L})$, where $\mathcal{L}$ is the underlying LWP of the SLWPs in the union.

Since, we are interested in $\{0, 1\}^m$ solutions of S_Ω, we write the solution set as $\bigcup_J S(J)$, where $S(J) = \{x \in \{0, 1\}^m : \sum_{i=1}^m w_i x_i \geq C, \Omega x = J\}$. The algorithm first uses Theorem 3 for setting up $\tilde{s}^{\geq}((\cdot, \cdot), \cdot)$ with a chosen parameter δ_s. Then, for each generation, it requires m steps. We denote the generated random variable as $\widetilde{\mathbb{X}}$. We begin by choosing J with probability $\frac{\tilde{s}^{\geq}((m,J),C)}{\sum_{J'} \tilde{s}^{\geq}((m,J'),C)}$. At iteration t of the algorithm, the values $(\widetilde{\mathbb{X}}_k)_{k=m-t+2}^m$ are fixed. Then, let $i = m - t + 1$, $c(t) = C - \sum_{k=i+1}^m w_k \widetilde{\mathbb{X}}_k$,

$j^J(t) = J - \Omega_{.,i+1:m}\widetilde{\mathbb{X}}_{.,i+1:m}$, and

$$\widetilde{p}_t^J = \frac{\tilde{s}^{\geq}\Big(\big(i-1, j^J(t)\big), c(t)\Big)(n_i - a_i)}{\tilde{s}^{\geq}\Big(\big(i-1, j^J(t)\big), c(t)\Big)(n_i - a_i) + \tilde{s}^{\geq}\Big(\big(i-1, j^J(t) - \Omega_{.,i}\big), c(t) - w_i\Big)a_i}.$$

At each iteration, the algorithm generates a uniform random variable $\mathbb{U}_t$ in [0, 1] and, if $\mathbb{U}_t \geq \widetilde{p}_t^J$, it sets $\widetilde{\mathbb{X}}_i = 1$. Otherwise, it sets $\widetilde{\mathbb{X}}_i = 0$.

Theorem 4 *Consider the* $\{0,1\}^m$ *solutions of* S_Ω *as defined above with the underlying distribution* Θ *as in (A5). Let* $\{w_i^l\}_{i=1}^m$ *denote the weights of the* l^{th} *constraint defining* $\mathcal{L}_\Omega$*, and assume that an error parameter* $\epsilon_s \in (0,1)$ *is given. Then, after initial setup that requires* $O(\epsilon_s^{-1} m^{t+4}\gamma^t \ln \mathcal{T})$ *time, we can generate a* $\{0,1\}^m$ *solution from* S_Ω *with a probability which is different from the true probability by a relative factor of* $(1 \pm \epsilon_s)$ *in time* $O\big(m + (m\gamma)^t\big)$*, where* $\mathcal{T} \geq \max_i n_i$ *and* $\gamma \geq \max_k w_k^l - \min_k w_k^l$ *for all* $l \in [t]$. □

As remarked earlier, Theorems 3 and 4 will be used to estimate $\Pr_\Theta(\mathcal{F})$. In particular, they are the key ingredients of the algorithm that estimates the probability of a union of SLWPs that cover $\mathcal{F}$. Moreover, after Proposition 3, we discussed the need to compute $\mathbb{E}_\Theta[\mathfrak{f}_\alpha(\mathbb{X})]$. Theorem 3 provides a $(1+\epsilon_s)$ relative approximation for this quantity when $\mathfrak{f}_\alpha = x^\alpha$ for $\alpha \in \{0,1\}^m$. These functions are especially useful, as shown in Theorem 1 to develop tight approximations for $\Pr_\Theta(\mathcal{F})$, and, similarly in approximating $\Pr_\Theta(\mathcal{F} \cap \mathcal{L}^{\mathfrak{b}})$, where $\mathcal{L}^{\mathfrak{b}}$, for $\mathfrak{b} \in \{1, \ldots, F\}$, is a partitioning of $\{0,1\}^m$. In fact, Theorem 3 can be used to approximate $\Pr_\Theta(\mathbb{X}^\alpha = 1, \mathbb{X} \in U)$ for any SLWP U. Indeed, $x^\alpha = 1$ if and only if $x_i = 1$ whenever $\alpha_i = 1$. Therefore, the set $\{x : x^\alpha = 1, x \in U\}$ is itself an SLWP and amenable to probability estimation.

3.1 Indicators of SLWPs to improve concave envelope bound

For any SLWP, U, we can use Theorem 3 to overestimate, to any accuracy, $\mathbb{E}_\Theta[\mathbb{1}_U(\mathbb{X})]$. So, we may use $\mathbb{1}_U(x)$ as $\mathfrak{f}(x)$ in (7) to improve the bound for $\Pr_\Theta(\mathcal{F})$ as obtained in (3). In this section, we describe Algorithm 1, which derives this improved bound and uses the following subroutines.

(1.) FIND- VIOLATE($(a,b),V$): Given a linear function $a^\intercal x + b$ and a set $V \subseteq \mathscr{T}$, FIND- VIOLATE returns an optimal solution to: $\max_{x\in V}\{\mathcal{DI}(x) - a^\intercal x - b\}$ where $\mathcal{DI}(x)$ is as in (8), if bounds on v, w are available so that this optimization problem can be solved as an integer program. Otherwise, FIND- VIOLATE returns an optimal solution to $\max_{x\in V}\{\mathcal{DI}(x) \mid a^\intercal x + b < 1\}$, where we additionally impose that $\|(v,w)\| \leq 1$ for some norm $\|\cdot\|$.

2. OVERESTIMATOR($\mathbb{E}_\Theta[\mathbb{X}]$, $\{\hat{\mathbb{E}}[\mathbb{1}_{U_i}(\mathbb{X})]\}_{i=1}^l$,$\{U_i\}_{i=1}^l$): Let conv$(\mathscr{T})$ be $\{x : C^0 x \geq d^0\}$, and consider its l non empty subsets $U_i = \{x : C^i x \geq d^i\}$, where $C^i \in \mathbb{R}^{k\times m}$ and $d^i \in \mathbb{R}^k$ $\forall i = \{0, \ldots, l\}$. Let $\hat{\mathbb{E}}[\mathbb{1}_{U_i}(\mathbb{X})]$ overestimate $\mathbb{E}_\Theta[\mathbb{1}_{U_i}(\mathbb{X})]$ for all $i \in [l]$. Then, consider $\min_{a,b,\pi}\big\{a^\intercal \mathbb{E}_\Theta[\mathbb{X}] + b + \sum_{i=1}^l \pi_i \hat{\mathbb{E}}[\mathbb{1}_{U_i}(\mathbb{X})] \;\big|\; a^\intercal x + b \geq 0\ \forall x \in \text{conv}(\mathscr{T});\ a^\intercal x + b + \pi_i \geq 1\ \forall x \in U_i,\ \forall i \in [l];\ \pi_i \geq 0\,\forall i \in [l]\big\}$, and use duality to write $a^\intercal x + b + \pi_i \geq 1\,\forall x \in U_i$

Algorithm 1 Bound $\Pr_{\Theta}(\mathcal{F})$ using indicators of SLWPs

Initialize: $V_P^1 = \text{conv}(\mathscr{T})$; $a = 0$, $b = 0$; $k = 1$
Input: Input Θ.
Output: A bound on $\Pr_{\Theta}(\mathcal{F})$

procedure BOUND
  $(w^*, v^*, x^*) \leftarrow$ FIND- VIOLATE$((a, b), V_P^k)$
  while $w^{*\intercal}(c - Bx^*) + v^*\varrho > 0$ **do**
    $\big(U_k, \hat{\mathbb{E}}[\mathbb{1}_{U_k}(\mathbb{X})]\big) \leftarrow$ SOLUTION- COUNTER$(\nu^* = (w^*, v^*), V_P^k)$
    $(a^*, b^*, \pi^*) \leftarrow$ OVERESTIMATOR$(\mathbb{E}_{\Theta}[\mathbb{X}], \{\hat{\mathbb{E}}[\mathbb{1}_{U_i}(\mathbb{X})]\}_{i=1}^k, \{U_i\}_{i=1}^k)$
    $k \leftarrow k + 1$
    $V_P^k \leftarrow V_P^{k-1} \cap \{x \mid w^{*\intercal}(c - Bx) + v^*\varrho \le 0\}$
    $(w^*, v^*, x^*) \leftarrow$ FIND- VIOLATE$((a^*, b^*), V_P^k)$
  return $a^{*\intercal}\mathbb{E}_{\Theta}[\mathbb{X}] + b^* + \sum_{i=1}^k \hat{\mathbb{E}}[\mathbb{1}_{U_i}(\mathbb{X})]$

as: $\max_{u^i \in \mathbb{R}_+^k}\{(u^i)^\intercal d^i + b + \pi_i - 1 \mid (u^i)^\intercal C^i = a^\intercal\} \ge 0$. Thus, OVERESTIMATOR solves:

$$\begin{aligned}
\min_{\pi,a,b,u} \quad & a^\intercal \mathbb{E}_{\Theta}[\mathbb{X}] + b + \sum_{i=1}^{l} \pi_i \hat{\mathbb{E}}[\mathbb{1}_{U_i}(\mathbb{X})] \\
& (u^0)^\intercal C^0 = a^\intercal,\ (u^0)^\intercal d^0 + b \ge 0 \\
& (u^i)^\intercal C^i = a^\intercal,\ (u^i)^\intercal d^i + b + \pi_i \ge 1 && \forall i \in [l] \\
& u^i, \pi_i \ge 0 && \forall i
\end{aligned}$$

Setting $\pi_i = 0$ for all $i \in [l]$, it follows that the optimal value of the above problem is no more than the concave envelope bound in Sect. 2.1. Moreover, since $(a, b, \pi) = (0, 0, \mathbf{1}_l)$, where $\mathbf{1}_l$ is an l-dimensional vector of ones, is feasible, it follows that the bound is no more than $\sum_{i=1}^{l} \hat{\mathbb{E}}[\mathbb{1}_{U_i}(\mathbb{X})]$.

3. SOLUTION- COUNTER$(\nu = (v', w'), V)$: Given $(v', w') \in \arg\max \mathcal{DI}(x')$ and $V \subseteq \mathscr{T}$, where $x' \in V$ is such that $\mathcal{DI}(x') > 0$. Let $V_s = \{x \in V : w'^\intercal(c - Bx) + v'\varrho > 0\}$, then from (8), $V_s \subseteq \mathcal{F}$, and $\Pr_{\Theta}(V_s)$ contributes to $\Pr_{\Theta}(\mathcal{F})$. SOLUTION- COUNTER uses the algorithm in Theorem 3 to overestimate $\mathbb{E}_{\Theta}[\mathbb{1}_{V_s}(\mathbb{X})]$ as $\hat{\mathbb{E}}[\mathbb{1}_{V_s}(\mathbb{X})]$. Since, it can only handle one general inequality, V is first transformed to a LWP, making V_s a SLWP, and then $\hat{\mathbb{E}}[\mathbb{1}_{V_s}(\mathbb{X})]$ is computed. It returns V_s and $\hat{\mathbb{E}}[\mathbb{1}_{V_s}(\mathbb{X})]$.

We mention that Algorithm 1 does not rely on Assumption (A5) except when we use Theorem 3 to estimate $\mathbb{E}_{\Theta}[\mathbb{1}_{U_i}(\mathbb{X})]$ or $\mathbb{E}_{\Theta}[\mathbb{X}]$.

Remark 1 To use a strict inequality, such as $a^\intercal x + b < 1$ in FIND- VIOLATE (resp. $-w^\intercal(c - Bx) - v\varrho < 0$ in SOLUTION- COUNTER, with c, B, and ϱ integer), we utilize that (a, b) (resp. (w, v)) are rational numbers, being optimal solutions to a linear program. We can, therefore, scale (a, b) (resp. (w, v)) so that they are integer and then increment b (resp. decrement v) by one and enforce the weak inequality. □

Remark 2 If $\max_{x \in V}\{\mathcal{DI}(x) - a^\intercal x - b\}$ is solved, for $V \subseteq \mathscr{T}$, then FIND- VIOLATE finds an $x \in \mathcal{F}$, such that $a^\intercal x + b$ has the lowest value. Instead,

$\max_{x\in V}\{\mathcal{DI}(x) \mid a^\intercal x + b < 1\}$ finds an $x \in V$ where $a^\intercal x + b$ evaluates to a value less than one. For practical reasons, we can replace $a^\intercal x + b < 1$ with $a^\intercal x + b \le 1 - \epsilon$. If FIND- VIOLATE does not find a violating (w^*, v^*, x^*), then $\Pr_\Theta(\mathcal{F}) \le \frac{1}{1-\epsilon} a^\intercal \mathbb{E}_\Theta[\mathbb{X}] + \frac{b}{1-\epsilon} + \sum_{i=1}^{l} \pi_i \hat{\mathbb{E}}[\mathbb{1}_{U_i}(\mathbb{X})]$. OVERESTIMATOR can also be adapted to account for the relative approximation error in the computation of $\mathbb{E}_\Theta[\mathbb{X}]$ using Theorem 3. In particular, assume that y_i, $i \in [m]$ are available such that $\frac{y_i}{1+\epsilon_s} \le \mathbb{E}_\Theta[\mathbb{X}_i] \le y_i(1+\epsilon_s)$. Since $a_i \mathbb{E}_\Theta[\mathbb{X}_i] \le \max\{\frac{a_i y_i}{1+\epsilon_s}, a_i y_i(1+\epsilon_s)\}$, we introduce variables z_i, $i = 1, \dots, m$, require that $z_i \ge \frac{a_i y_i}{1+\epsilon_s}$ and $z_i \ge a_i y_i (1+\epsilon_s)$. Then, we minimize $\sum_{i=1}^{m} z_i + b + \sum_{i=1}^{l} \pi_i \hat{\mathbb{E}}[\mathbb{1}_{U_i}(\mathbb{X})]$ instead. □

Remark 3 Algorithm 1 terminates in at most $|\mathscr{T}|$ iterations of the loop starting at Step 3. This is because, at Step 7, $|V_P^k|$ is strictly decreasing in k. The finiteness can also be shown as follows. For a fixed x^*, FIND- VIOLATE solves a linear program to determine (w^*, v^*). We may, therefore, assume that $(w^*, v^*) \in \text{Vert}(S_{w,v}(x^*))$, where $S_{w,v}(x^*) = \{w^\intercal A + ve^\intercal = 0, w^\intercal(c - Bx^*) + v\varrho \le 1, v, w \le 0\}$ because $S_{w,v}(x^*)$ does not contain lines and the optimal value is finite. Moreover, the vertices of $S_{w,v}$ are extreme rays of $P_{w,v}$, where $P_{w,v} = \{w^\intercal A + ve^\intercal = 0, w, v \le 0\}$. Therefore, it follows that the number of iterations of the loop is bounded by the number of extreme rays of $P_{w,v}$. □

As presented, in Algorithm 1, FIND- VIOLATE solves an integer program (or MINLP) at each iteration to find an $x \in \mathcal{F}$. As an alternative, we bypass OVERESTIMATOR so that it returns $(a, b, \pi) = (0, 0, \mathbf{1}_l)$ and combine the search in a single branch & bound tree in Gurobi [17]. Whenever an integer feasible solution (w^*, v^*, x^*) is found at Step 2 or Step 8 of Algorithm 1, we use the lazy-constraint callback function and add a cut $(w^*)^\intercal(c - Bx) + v^*\varrho \le 0$ which eliminates x^* in Step 7 of Algorithm 1 and continue to find next violation.

Remark 4 If (w_k^*, v_k^*, x_k^*) is obtained at the k^{th} iteration of Algorithm 1, using FIND-VIOLATE((0,0),V_P^k), then $\mathcal{F} = \bigcup_{k=1}^{L} U_k$, where $U_k = \{x \in V_P^k : (w_k^*)^\intercal(c - Bx) + v_k^*\varrho > 0\}$, and L is the number of iterations until the algorithm terminates. After any intermediate iteration $t < L$, we obtain $\{U_k\}_{k=1}^{t}$, such that $\bigcup_{k=1}^{t} U_k \subseteq \mathcal{F}$. We use this to obtain a lower estimate for $\Pr_\Theta(\mathcal{F})$. To do so, we may first transform the constraints defining V_P^k for all $k \in [t]$, to low weight constraints and obtain $\underline{V}_P^k$ as an inner-approximation to V_P^k, *i.e.*, $\underline{V}_P^k \subseteq V_P^k$. Then, each U_k for $k \in [t]$ is a SLWP. For $\epsilon_s \in (0, 1)$, we may use Appendix I to obtain a $(1 - \epsilon_s)$ relative approximation of $\Pr_\Theta(U_k)$ for all $k \in [t]$. Given an ϵ_g, this assists in deriving a $1 - \epsilon_g$ relative approximation confidence estimate of $\Pr_\Theta(\cup_{k=1}^{t} U_k)$ (see Sect. 4.1), a lower bound for $\Pr_\Theta(\mathcal{F})$. □

Clearly, a drawback of this approach is that at each iteration, FIND- VIOLATE solves an integer program (or MINLP) which is NP-hard. Moreover, a cover for $\mathcal{F}$ or an overestimate for $\Pr_\Theta(\mathcal{F})$ is obtained only at the termination of the algorithm. As such, this approach is not well-suited for obtaining an overestimate to $\Pr_\Theta(\mathcal{F})$ when L is large (see for instance our computational experience on network routing for D(3) - Deltacom $\mathfrak{b} = 3$ in Table 1). Next, we propose a scheme which derives an overestimate for $\Pr_\Theta(\mathcal{F})$ in polynomial time.

4 Policy restrictions to outer-approximate $\mathcal{F}$ by a union of SLWPs

In Sect. 2, we discussed how a candidate $\hat{d}(x)$ in (2) is constructed by restricting y to be an affine function of x and related this construction to RLT relaxations for the corresponding nonlinear formulation. Now, we will elaborate on our discussion following Assumption (A5) to show that such affine policy restrictions naturally lend themselves to an outer-approximation of $\mathcal{F}$ via a union of SLWPs. Assume $y \in Y(x)$ as defined in (4), is restricted to be an affine function, $\delta(\mathbb{X})$, of $\mathbb{X}$. Then, for an LWP $\mathcal{L}$, $\mathcal{F} \cap \mathcal{L} \subseteq S' = D_0 \cup \bigcup_{r=1}^{p} D_r$, where $D_r = \{x \in \mathcal{L} : A_r\delta(x) + B_r x > c_r\}$ for all $r \in [p]$ and $D_0 = \{x \in \mathcal{L} : e^\intercal \delta(x) > \varrho\}$. Clearly, $\Pr_\Theta(\mathcal{F}) \le \sum_{r=0}^{p} \Pr_\Theta(D_r)$. Here, we discuss how this estimate can be improved. To do so, we may project out a few y variables, thereby removing the restriction that they are affine in $\mathbb{X}$. Second, we bound $\Pr_\Theta(S') = \sum_{r=0}^{p} \Pr_\Theta\big(D_r \cap \bigcap_{j=0}^{r-1} D_j^C\big) \le \sum_{r=0}^{p} \Pr_\Theta\big(D_r \cap \bigcap_{j\in J_r} D'^C_j\big)$, where $J_r \subseteq \{0, \ldots, r-1\}$, $|J_r| \le \kappa$ for some constant κ, and $D'_r \subseteq D_r$. For tractability, we choose D'^C_r as an approximation of D_r^C defined using low-weight constraints. For example, let $D_r^C = \{x \in \mathcal{L} : \sum_{j=1}^{m} w_j x_j \le C\}$. Then, we use standard approximation techniques to define D'^C_r. For a given ϵ, and for $q = \max_{j=1}^{m} |w_j|$, we replace w_j with $w'_j = \left\lceil \frac{w_j m}{q\epsilon} \right\rceil$ and C with $C' = \min\left\{m\left\lceil \frac{m}{\epsilon} \right\rceil, \left\lceil \frac{Cm}{q\epsilon} \right\rceil\right\}$. This way, $|w'_j| \le \left\lceil \frac{m}{\epsilon} \right\rceil$, for all j, *i.e.,*all coefficients are bounded by a polynomial in m and $\frac{1}{\epsilon}$ and we can use Theorem 3 to approximate $\Pr_\Theta\big(D_r \cap \bigcap_{j\in J_r} D'^C_j\big)$. Thus, we can outer-approximate $\mathcal{F}$ with $\bigcup_{r=0}^{p} D_r \cap \mathcal{B}_r$, where, for each r, $\mathcal{B}_r = \bigcap_{j\in J_r} D'^C_j$ is a LWP, D_r is equivalently written as a SLWP (since SLWPs do not allow strict inequality) and, so, $D_r \cap \mathcal{B}_r$ is a SLWP. Later in Sect. 4.2, we discuss various ways to derive such an affine policy to cover the set of interest using a union of SLWPs.

4.1 Randomized approximation scheme for probability of union of SLWPs

We discussed in Sects. 3 and 4 that affine policies can be used to outer-approximate $\mathcal{F} \cap \mathcal{L}^{\mathfrak{b}}$ for $\mathfrak{b} \in [F]$, and thereby $\mathcal{F}$ via a union of SLWPs. Similar SLWPs also arise in Algorithm 1, as stated in Remark 4. Assume $S := \bigcup_{l=1}^{L} S_l \supseteq \mathcal{F} \cap \mathcal{L}^{\mathfrak{b}}$ is such an outer-approximation. By Assumption (A5), the support of Θ is restricted to binary points, $\{0, 1\}^m$. Moreover, any point generated by Theorem 4 belongs to $\{0, 1\}^m$. In this section, whenever we say that $x \in S_l$, it should be understood that $x \in \{0, 1\}^m \cap S_l$ since these are the only vectors generated by our algorithm and $\Pr_\Theta(x) = 0$ for all $x \notin \{0, 1\}^m$. In this section, we will use sampling to obtain a $1 + \epsilon_g$ relative approximation confidence estimate of $\Pr_\Theta(S)$. To this end, we adapt the algorithm of [12, 20] to our setting and present it as Algorithm 2. This algorithm requires three input parameters, where α_g and ϵ_g specify the accuracy of the approximation, and δ_g specifies the significance level. It also assumes the existence of the following polynomial time subroutines.

Algorithm 2 Generalized KLM [12] to estimate $\Pr_\Theta(S)$ for $S = \bigcup_{l=1}^{L} S_l$

Input parameters: α_g, ϵ_g and δ_g.
Output: $\widehat{\Pr}(S)$ such that with probability $(1-\delta_g)$,
$(1-\alpha_g)(1-\epsilon_g)\Pr_\Theta(S) \le \widehat{\Pr}(S) \le (1+\alpha_g)(1+\epsilon_g)\Pr_\Theta(S)$

procedure GENERALIZED- KLM
  $T \leftarrow \frac{3L}{\alpha_g^2}\ln\left(\frac{2}{\delta_g}\right)$
  $\widehat{\Pr}(S_l) \leftarrow$ SLWP- PROB$(\epsilon_g, S_l, \Theta)\ \forall l \in [L]$
  for $t = 1$ to T **do**
    Choose $i_0 \in [L]$ with probability $\frac{\widehat{\Pr}(S_{i_0})}{\sum_{i=1}^{L}\widehat{\Pr}(S_i)}$
    $(\widetilde{\mathbb{X}} \in S_{i_0}) \leftarrow$ SAMPLE- ASSIGN$(\epsilon_g, S_{i_0}, \Theta)$
    for $j = 1$ to L **do**
      if SATISFACTION$(\widetilde{\mathbb{X}}, S_j)$ is True **then**
        $t(\widetilde{\mathbb{X}}) \leftarrow j$
        break
    if $t(\widetilde{\mathbb{X}}) = i_0$ **then**
      $Z_t \leftarrow 1$
    else
      $Z_t \leftarrow 0$
  Return $\widehat{\Pr}(S) = \frac{1}{T}\sum_{t=1}^{T} Z_t \sum_{l=1}^{L}\widehat{\Pr}(S_l)$

(1.) SLWP- PROB$(\epsilon_g, S_l, \Theta)$: This procedure computes $\widehat{\Pr}(S_l)$ which is a $(1 \pm \epsilon_g)$ relative error approximation of $\Pr_\Theta(S_l)$. Using Theorem 3, we compute:

$$\Pr_\Theta(S_l) \le \widehat{\Pr}(S_l) \le \left(1 + \frac{\epsilon_g}{3}\right)\Pr_\Theta(S_l). \tag{16}$$

(2.) SAMPLE- ASSIGN$(\epsilon_g, S_l, \Theta)$: This procedure generates $\widetilde{\mathbb{X}}_l \in S_l$ so that:

$$(1-\epsilon_g)\Pr_\Theta(\mathbb{X} = \widetilde{x}) \le \Pr(\widetilde{\mathbb{X}}_l = \widetilde{x})\widehat{\Pr}(S_l) \le (1+\epsilon_g)\Pr_\Theta(\mathbb{X} = \widetilde{x}), \tag{17}$$

for every $\widetilde{x} \in S_l$.
We show that the sampling algorithm of Theorem 4 with $\epsilon_s = \frac{\epsilon_g}{3}$ meets this condition. If $\widetilde{x} \notin S_l$, then, by (J.3) and $\Pr_\Theta(\mathbb{X} = \widetilde{x} | \mathbb{X} \in S_l) = 0$, it follows that $\Pr(\widetilde{\mathbb{X}}_l = \widetilde{x}) = 0$. For $\widetilde{x} \in S_l$, (17) holds because

$$\begin{aligned}(1-\epsilon_s)\Pr_\Theta(\mathbb{X} = \widetilde{x}) &\le (1-\epsilon_s)\Pr_\Theta(\mathbb{X} = \widetilde{x})\frac{\widehat{\Pr}(S_l)}{\Pr_\Theta(S_l)} \le \Pr(\widetilde{\mathbb{X}}_l = \widetilde{x})\widehat{\Pr}(S_l)\\ &\le (1+\epsilon_s)\Pr_\Theta(\mathbb{X} = \widetilde{x})\frac{\widehat{\Pr}(S_l)}{\Pr_\Theta(S_l)}\\ &\le (1+3\epsilon_s)\Pr_\Theta(\mathbb{X} = \widetilde{x}),\end{aligned} \tag{18}$$

where the first and last inequality are by (16) and $(1+\epsilon_s)^2 \le 1 + 3\epsilon_s$ for $\epsilon_s \in (0, 1)$. The second and third inequalities are by (J.3).

(3.) SATISFACTION(x, S_l): Given a x and S_l, this subroutine trivially checks if $x \in S_l$ by using the inequalities that define S_l.

Then, Theorem 1 in [12] shows that for any $\alpha_g, \epsilon_g, \delta_g$, Algorithm 2 computes $\widehat{\Pr}(S) = \frac{1}{T}\sum_{t=1}^{T} Z_t \sum_{l=1}^{L} \widehat{\Pr}(S_l)$ such that, with probability $(1-\delta_g)$,

$$(1-\alpha_g)(1-\epsilon_g)\Pr_\Theta(S) \le \widehat{\Pr}(S) \le (1+\alpha_g)(1+\epsilon_g)\Pr_\Theta(S). \tag{19}$$

We show that the sampling algorithm yields a fully polynomial almost uniform sampler (FPAUS) for $S = \bigcup_{i=1}^{L} S_i$. A FPAUS for a set S is a randomized algorithm that takes as input a tolerance δ_u and generates a random variable $\mathbb{G} \in S$ so that $d_{\text{TV}} \le \delta_u$, in time polynomial in the problem size (m) and $\log(\frac{1}{\delta_u})$, where $d_{\text{TV}} = \frac{1}{2}\sum_{x\in S} |\Pr_\Theta(\mathbb{X} = x|\mathbb{X} \in S) - \Pr(\mathbb{G} = x)|$ is the total variation distance between the true distribution and the sampling distribution. Wlog we assume $S_1 \neq \emptyset$ and, in particular, $x' \in S_1 \cap \{0,1\}^m$. We run Algorithm 2, and terminate with $\widetilde{\mathbb{X}}$ the first time it encounters $Z_t = 1$ if this occurs in no more than $T' = 2L\ln\left(\frac{5}{2\delta_u}\right)$ iterations. Otherwise, we return x'. We denote the random vector generated by this procedure as $\mathbb{G}$. For any $x \in S$, let $t(x) = \min_i\{i \mid x \in S_i\}$, *i.e.,*among $\{S_1, \ldots, S_L\}$, the first set that x belongs to has index $t(x)$. Observe that, for

$$\begin{aligned}
\Pr(Z_t = 0) = 1 - \Pr(Z_t = 1) &= 1 - \frac{\sum_{x\in S}\Pr(\widetilde{\mathbb{X}}_{t(x)} = x)\widehat{\Pr}(S_{t(x)})}{\sum_{i=1}^{L}\widehat{\Pr}(S_i)} \\
&\le 1 - \frac{(1-\epsilon_g)}{(1+\epsilon_g)}\frac{\sum_{x\in S}\Pr_\Theta(\mathbb{X} = x)}{\sum_{i=1}^{L}\Pr_\Theta(S_i)} \\
&\le 1 - \frac{1}{2}\frac{\max_{i\in[L]}\Pr_\Theta(S_i)}{\sum_{i=1}^{L}\Pr_\Theta(S_i)} \le 1 - \frac{1}{2L}.
\end{aligned}$$

The first inequality follows from (16) and (18), the second inequality because, for $\epsilon \in (0, 1/3)$, $\frac{1-\epsilon}{1+\epsilon} \ge \frac{1}{2}$ and $\sum_{x\in S}\Pr_\Theta(\mathbb{X} = x) = \Pr_\Theta(S) \ge \Pr_\Theta(S_i)$ for all i. Let $Z' = 0$ iff $Z_{t'} = 0$ for all $t' \in [T']$. Then,

$$\Pr(Z' = 0) = \Pr(Z_{t'} = 0, t' = 1, \ldots, T') \le \left(1 - \frac{1}{2L}\right)^{T'} \le \frac{2\delta_u}{5}. \tag{20}$$

Now,

$$\begin{aligned}
\Pr(\mathbb{G} = x \text{ and } Z' = 1) &= \sum_{t'=1}^{T'} \frac{\Pr(\widetilde{\mathbb{X}}_{t(x)} = x)\widehat{\Pr}(S_{t(x)})}{\sum_{i=1}^{L}\widehat{\Pr}(S_i)} \prod_{t''<t'} \Pr(Z_{t''} = 0) \\
&= \sum_{t'=1}^{T'}\left(\frac{\Pr(\widetilde{\mathbb{X}}_{t(x)} = x)\widehat{\Pr}(S_{t(x)})}{\sum_{\widetilde{x}\in S}\Pr(\widetilde{\mathbb{X}}_{t(\widetilde{x})} = \widetilde{x})\widehat{\Pr}(S_{t(\widetilde{x})})}\Pr(Z_{t'} = 1)\prod_{t''<t'}\Pr(Z_{t''} = 0)\right) \\
&= \frac{\Pr(\widetilde{\mathbb{X}}_{t(x)} = x)\widehat{\Pr}(S_{t(x)})}{\sum_{\widetilde{x}\in S}\Pr(\widetilde{\mathbb{X}}_{t(\widetilde{x})} = \widetilde{x})\widehat{\Pr}(S_{t(\widetilde{x})})}\Pr(Z' = 1),
\end{aligned} \tag{21}$$

where the first equality follows by observing that the process stops at t' iteration with $\mathbb{G} = x$ and $Z' = 1$, if $Z_{t''} = 0$ for $t'' < t'$, $Z_{t'} = 1$, and x is generated at t'. This happens if $S_{t(x)}$ is chosen at Step 5 of Algorithm 2 and x is generated at Step 6 of

Algorithm 2 by sampling $S_{t(x)}$. The second equality uses

$$\Pr(Z_{t'} = 1) = \sum_{i=1}^{L} \frac{\widehat{\Pr}(S_i) \sum_{\widetilde{x}:t(\widetilde{x})=i} \Pr(\widetilde{\mathbb{X}}_i = \widetilde{x})}{\sum_{l=1}^{L} \widehat{\Pr}(S_l)} = \sum_{\widetilde{x}\in S} \frac{\Pr(\widetilde{\mathbb{X}}_{t(\widetilde{x})} = \widetilde{x})\widehat{\Pr}(S_{t(\widetilde{x})})}{\sum_{l=1}^{L} \widehat{\Pr}(S_l)},$$

and the third equality follows from the independence of $Z_{t'}$ with $Z_{t''}$ for $t'' < t'$ and $\Pr(Z' = 1) = \sum_{t'=1}^{T'} \Pr(Z_{t'} = 1) \prod_{t''<t'} \Pr(Z_{t''} = 0)$. It follows from (J.3) and (16) that:

$$(1 - 3\epsilon_g)\frac{\Pr_\Theta(\mathbb{X} = x)}{\Pr_\Theta(S)} \le \frac{\Pr(\widetilde{\mathbb{X}}_{t(x)} = x)\widehat{\Pr}(S_{t(x)})}{\sum_{\widetilde{x}\in S} \Pr(\widetilde{\mathbb{X}}_{t(\widetilde{x})} = \widetilde{x})\widehat{\Pr}(S_{t(\widetilde{x})})} \le (1 + 4\epsilon_g)\frac{\Pr_\Theta(\mathbb{X} = x)}{\Pr_\Theta(S)}, \tag{22}$$

where we used that $\frac{1-\epsilon_g}{(1+\epsilon_g)^2} \ge 1 - 3\epsilon_g$ and $\frac{(1+\epsilon_g)^2}{(1-\epsilon_g)} \le (1 + 4\epsilon_g)$ for $\epsilon_g \in (0, \frac{1}{5})$. Choose $\epsilon_g = \frac{2\delta_u}{5}$. Therefore, by (21), (20), and, (22), it follows that if $x \ne x'$:

$$\left(1 - \frac{8}{5}\delta_u\right) \frac{\Pr_\Theta(\mathbb{X} = x)}{\Pr_\Theta(S)} \le \Pr(\mathbb{G} = x, Z' = 1) \le \left(1 + \frac{8}{5}\delta_u\right) \frac{\Pr_\Theta(\mathbb{X} = x)}{\Pr_\Theta(S)}, \tag{23}$$

where we used $1 - \frac{8}{5}\delta_u \le (1 - \frac{6}{5}\delta_u)(1 - \frac{2\delta_u}{5})$ and $1 + 4\epsilon_g = 1 + \frac{8}{5}\delta_u$. Thus, $d_{\text{TV}} = \frac{1}{2}\sum_{x\in S} |\Pr_\Theta(\mathbb{X} = x|\mathbb{X} \in S) - \Pr(\mathbb{G} = x)| \le \frac{1}{2}\sum_{x\in S} |\Pr_\Theta(\mathbb{X} = x|\mathbb{X} \in S) - \Pr(\mathbb{G} = x, Z' = 1)| + \frac{\Pr(Z'=0)}{2} \le \frac{4\delta_u}{5}\sum_{x\in S} \Pr_\Theta(\mathbb{X} = x|\mathbb{X} \in S) + \frac{\delta_u}{5} = \delta_u$.

Assume we have $S = \bigcup_{i=1}^{L'} S'_i$ such that $\mathcal{F} \cap \mathcal{L} \subseteq S$, where $\mathcal{L}$ is a LWP. The FPAUS for S can be used to estimate $\Pr_\Theta(\mathcal{F})$ using rejection sampling. Let $\theta_\mathcal{F}$ and $\beta_\mathcal{F}$ be given parameters that will be used in the specification of the accuracy of estimation and the confidence level respectively. We run $T'' = \frac{2+\theta_\mathcal{F}}{\theta_\mathcal{F}^2} \ln \frac{2}{\beta_\mathcal{F}}$ iterations of the FPAUS described above. For each $t' \in [T'']$, let $Z'_{t'} = 0$ if $Z' = 0$ or $\mathbb{G} \notin \mathcal{F} \cap \mathcal{L}$. Otherwise, we define $Z'_{t'} = 1$. Therefore, it follows that $\mathbb{E}[Z'_{t'}] = \tilde{p} = \Pr(\mathbb{G} \in \mathcal{F} \cap \mathcal{L}, Z' = 1)$, which by (23) yields:

$$(1 - \epsilon_u)\frac{\Pr_\Theta(\mathcal{F} \cap \mathcal{L})}{\Pr_\Theta(S)} \le \tilde{p} \le (1 + \epsilon_u)\frac{\Pr_\Theta(\mathcal{F} \cap \mathcal{L})}{\Pr_\Theta(S)}, \tag{24}$$

where we have used $\epsilon_u = \frac{8}{5}\delta_u$. Let $Z'' = \sum_{t'=1}^{T''} Z'_{t'}$ and $p = \frac{\Pr_\Theta(\mathcal{F}\cap\mathcal{L})}{\Pr_\Theta(S)}$. Let $\epsilon_\mathcal{F} = \frac{\theta_\mathcal{F}}{\tilde{p}}$. By the 2-sided Chernoff bound, we have $\Pr(|Z'' - \tilde{p}| \ge \theta_\mathcal{F}) = \Pr(|Z'' - \tilde{p}| \ge \epsilon_\mathcal{F}\tilde{p}) \le 2\exp\left(-\frac{\epsilon_\mathcal{F}^2}{2+\epsilon_\mathcal{F}}\tilde{p}T''\right) \le 2\exp\left(-\frac{\theta_\mathcal{F}^2}{2+\theta_\mathcal{F}}T''\right) = \beta_\mathcal{F}$. Therefore, $\Pr(Z'' - \theta_\mathcal{F} \le \tilde{p} \le Z'' + \theta_\mathcal{F}) \ge (1 - \beta_\mathcal{F})$. Combining with (24) and (19), we have $\Pr\left(\frac{(Z''-\theta_\mathcal{F})\widehat{\Pr}(S)}{(1+\epsilon_u)(1+\alpha_g)(1+\epsilon_g)} \le \Pr_\Theta(\mathcal{F} \cap \mathcal{L}) \le \frac{(Z''+\theta_\mathcal{F})\widehat{\Pr}(S)}{(1-\epsilon_u)(1-\alpha_g)(1-\epsilon_g)}\right) \ge (1 - \delta_g)(1 - \beta_\mathcal{F})$.

4.2 Deriving affine policies to cover $\mathcal{F} \cap \mathcal{L}$ with a union of SLWPs

Earlier in Sect. 4, we discussed how the affine policy $\delta(x)$ assists in outer-approximating $\mathcal{F}$ with a union of SLWPs. In this section, we discuss two ways of obtaining affine policies that utilize chance-constrained programming techniques [26]. Let $\mathbb{V}$ be a discrete random variable such that $\Pr(\mathbb{V} = \tau_i) = p^i$ for $i \in \{1, \ldots, \bar{s}\}$. Consider the problem $\max\{\frac{1}{\widetilde{\mathcal{L}}} \sum_{i=1}^{\bar{s}} z_i \tau_i \mid \sum_{i=1}{}^{\bar{s}} z_i \leq \widetilde{\mathcal{L}}, 0 \leq z_i \leq p^i\}$. Let $\bar{\tau} = \max\{\tau_i \mid \Pr(\mathbb{V} \geq \tau_i) \geq \widetilde{\mathcal{L}}\}$. Clearly, an optimal solution sets $z_i = p^i$ if $\tau_i > \bar{\tau}$, $z_i = \frac{\widetilde{\mathcal{L}} - \Pr(\mathbb{V} > \bar{\tau})}{\Pr(\mathbb{V} = \bar{\tau})}$ if $\tau_i = \bar{\tau}$, and $z_i = 0$ otherwise. This models expected value of the largest supports of $\mathbb{V}$ with a cumulative probability of $\widetilde{\mathcal{L}}$, which is typically referred to as the conditional value-at-risk (CVaR). Then, as is standard, this quantity is computed by solving the dual C_D: $\min_{l, v_i}\{l + \frac{1}{\widetilde{\mathcal{L}}} \sum_{i=1}^{\bar{s}} p^i v_i \mid v_i \geq 0\, \forall i \in [\bar{s}], l + v_i \geq \tau_i\, \forall i \in [\bar{s}]\}$ [5]. So, for a sample, $\{x^i\}_{i \in [\bar{s}]}$, we use C_D to minimize CVaR [35]:

$$\text{P-CVaR:} \min_{l, v, \tau} \Bigg\{ l + \frac{1}{\widetilde{\mathcal{L}}} \sum_{i=1}^{\bar{s}} p^i v_i \mid \forall i \in [\bar{s}],\, f(x^i, \delta(x^i)) \leq \varrho + \tau_i \bar{w}_0,$$

$$A\delta(x^i) + Bx^i \leq c + \tau_i \bar{w},\, \tau_i \geq v_i + l,\, v_i \geq 0 \Bigg\}, \tag{25}$$

where we have scaled τ_i in the constraints using $\bar{w} \in \mathbb{R}^p$, a vector of arbitrary positive weights, and a positive $\bar{w}_0 \in \mathbb{R}$. As before, p^i is the probability of scenario x^i and $\bar{s}$ is the number of sampled scenarios. The variables v_i for $i \in [\bar{s}]$ and l are used to model C_D as above.

In contrast, Bernstein approximation does not require explicit enumeration of scenarios [6, 31]. Recall that x is not in $\mathcal{F} \cap \mathcal{L}$ if the affine policy satisfies all the constraints and evaluates to an objective no larger than ϱ. We previously outer-approximated $\mathcal{F} \cap \mathcal{L}$ as a union of sets, each one obtained when one of these conditions is violated. Say, this violated inequality is given by $w^\intercal x > C$ where w_i and C depend, possibly, on the affine coefficients used to derive $\delta(x)$. Then, under various technical conditions, when the cumulant generating function of $\sum_{i=1}^m w_i \mathbb{X}_i$ over $\mathcal{L}$ is known, we can use Bernstein approximation to derive $\delta(x)$. Observe that if $\mathcal{L} = [0, 1]^m$ then $\Pr_\Theta(\sum_{i=1}^m w_i \mathbb{X}_i \geq C) \leq \bar{\epsilon}$ if there is a $t' > 0$ such that (t', w) satisfy the following convex constraint [4]:

$$-C + \sum_{i=1}^m t' \log\Big(\frac{n_i - a_i}{n_i} + \frac{a_i}{n_i} e^{\frac{w_i}{t'}}\Big) \leq t' \log(\bar{\epsilon}). \tag{26}$$

We derive a similar approximation for a specially structured Θ, where each set in the outer-approximation is of the form S',

(A7) $S' = \big\{ \sum_{i=1}^m w_i x_i \geq C, \sum_{i=1}^m x_i = \mathfrak{b},$ where $w_i \geq 0\ \forall i \big\}$ and $\Pr_\Theta(\mathbb{X}_i = 1) = p$ for all i.

Although classical Bernstein approximation does not allow side constraints, we use the recent analysis of 1-negatively correlated random variables in [11] to derive this extension. To do so, we upper bound $\mathbb{E}_\Theta[\exp(t\mathbb{Y}) \mid \sum_{i=1}^m \mathbb{X}_i = \mathfrak{b}]$, where $\mathbb{Y} = \sum_{i=1}^m w_i \mathbb{X}_i$. Assume, for all $i \in [m]$, $\mathbb{X}'_i$ are independent random variables such that $\Pr(\mathbb{X}'_i = 1) = \mathfrak{b}/m$ and $\mathbb{Y}' = \sum_{i=1}^m w_i \mathbb{X}'_i$. Then, for $I \subseteq [m]$, if $|I| \le \mathfrak{b}$, we have $\Pr_\Theta(\forall i \in I, \mathbb{X}_i = 1 \mid \sum_{i=1}^m \mathbb{X}_i = \mathfrak{b}) = \frac{\binom{m-|I|}{\mathfrak{b}-|I|}}{\binom{m}{\mathfrak{b}}} \le \left(\frac{\mathfrak{b}}{m}\right)^{|I|}$. Instead, if $|I| > \mathfrak{b}$, $\Pr_\Theta(\forall i \in I, \mathbb{X}_i = 1 \mid \sum_{i=1}^m \mathbb{X}_i = \mathfrak{b}) = 0 \le \left(\frac{\mathfrak{b}}{m}\right)^{|I|}$. This implies that: $\mathbb{E}_\Theta[\mathbb{Y}^l \mid \sum_{i=1}^m \mathbb{X}_i = \mathfrak{b}] = \mathbb{E}_\Theta\left[\sum_{\mathfrak{i}\in[m]^l} \prod_{j=1}^l w_{\mathfrak{i}_j} \mathbb{X}_{\mathfrak{i}_j} \mid \sum_{i=1}^m \mathbb{X}_i = \mathfrak{b}\right] \le \sum_{\mathfrak{i}\in[m]^l} \prod_{j=1}^l w_{\mathfrak{i}_j} \prod_{i\in\{\mathfrak{i}_1,\ldots,\mathfrak{i}_l\}} \Pr(\mathbb{X}'_i = 1) = \mathbb{E}\left[\sum_{\mathfrak{i}\in[m]^l} \prod_{j=1}^l w_{\mathfrak{i}_j} \mathbb{X}'_{\mathfrak{i}_j}\right] = \mathbb{E}[\mathbb{Y}'^l]$, where the first inequality uses that $\prod_{j=1}^l w_{\mathfrak{i}_j}$ is non-negative. Therefore, for all $t > 0$, $\mathbb{E}_\Theta[\exp(t\mathbb{Y}) \mid \sum_{i=1}^m \mathbb{X}_i = \mathfrak{b}] \le \mathbb{E}[\exp(t\mathbb{Y}')]$ and

$$\Pr_\Theta(S') \le \frac{\mathbb{E}_\Theta\left[\exp(t\mathbb{Y}) \mid \sum_{i=1}^m \mathbb{X}_i = \mathfrak{b}\right]}{\exp(tC)} \le \exp(-tC)\mathbb{E}\left[\exp\left(t\mathbb{Y}'\right)\right]. \tag{27}$$

Then, with $t' = \frac{1}{t}$, $\Pr_\Theta(S') \le \overline{\epsilon}$ if there exists $t' > 0$ such that

$$-C + \sum_{i=1}^m t' \log\left(1 - \frac{\mathfrak{b}}{m} + \frac{\mathfrak{b}}{m} e^{\frac{w_i}{t'}}\right) \le t' \log(\overline{\epsilon}). \tag{28}$$

Now, with $\bar{p}_i = \frac{a_i}{n_i}$ (resp. $\bar{p}_i = \frac{\mathfrak{b}}{m}$), the following formulation can be used to model (26) (resp. (28)), where $K_{\exp}$ denotes the exponential cone:

$$\Bigg\{(u, v, \theta) : \; -C + \sum_{i=1}^m \theta_i \le t' \log \overline{\epsilon}, \; (1 - \bar{p}_i)u_i$$
$$+\bar{p}_i v_i \le t'(u_i, t', -\theta_i), (v_i, t', w_i - \theta_i) \in K_{\exp} \text{ for } i \in [m]\Bigg\}. \tag{29}$$

5 Case study: network reliability problem

Consider a graph, $G(V, E)$, where V and E are the set of nodes and edges in G. Let $d : V \times V \to \mathbb{R}$ be the traffic between node-pairs and $c : E \to \mathbb{R}$ the link capacities. For link $\langle i, j\rangle \in E$, $x_{ij} = 1$ if $\langle i, j\rangle$ fails and 0 otherwise. Given x, the network routes traffic by solving a *multicommodity flow problem* as in (30) that minimizes *maximum link utilization*, $\text{MLU}(x)$, where $\text{MLU}(x) = \max_{\langle i,j\rangle\in E} U_{ij}$ and U_{ij} is the ratio of the traffic on $\langle i, j\rangle$ to c_{ij}.

$$\text{MLU}(x) = \min_{y,U} \quad U \tag{30a}$$

$$\sum_{t \in V} y_{ijt} \leq U c_{ij}(1 - x_{ij}) \qquad \forall \langle i, j \rangle \in E \quad (30b)$$

$$\sum_{j \in V} y_{ijt} - \sum_{j \in V} y_{jit} = d_{it} - \sum_{j \in V} d_{jt} \delta_{(i=t)} \qquad \forall i, t \in V \quad (30c)$$

$$y_{ijt}, U \geq 0 \qquad \begin{array}{l} \forall \langle i, j \rangle \in E \\ \forall t \in V, \end{array} \quad (30d)$$

where, for all $\langle i, j \rangle \in E$ and $t \in V$, y_{ijt} is the flow destined to node t on link $\langle i, j \rangle$, and $\delta_{(i=t)}$ is 1 if $i = t$ and 0 otherwise. Constraints (30b) and (30c) model capacity and flow balance constraints respectively. We will consider the uncertainty set of $\mathfrak{b}$ simultaneous link failures $\mathscr{X}_{\mathfrak{b}} = \{x \in \{0, 1\}^{|E|} : \sum_{\langle i, j \rangle \in E} x_{ij} = \mathfrak{b}\}$, which is of interest to network architects [9, 10, 25, 39]. Further, consistent with network failure measurements [15, 27], we assume that links fail independently of one another, *i.e.,* if $\mathbb{X}$ is a $|E|$-dimensional binary random vector representing the state of the links then $\mathbb{X} \sim \bigotimes_{i=1}^{|E|} \text{Bernoulli}(p_i)$. We are then interested in computing $\Pr_{\Theta}(\mathcal{F}_{\text{NR}})$, where $\mathcal{F}_{\text{NR}} = \{x \in \mathscr{X}_{\mathfrak{b}} : \text{MLU}(x) > 1\}$, a set we often refer to as set of "bad failures". Clearly, this is a special case of the PQ problem introduced in Sect. 2, where $\mathscr{P} = \{x : \sum_{\langle i, j \rangle \in E} x_{ij} = \mathfrak{b}\}$, $d(x) = \text{MLU}(x)$, $\varrho = 1$ and $\mathbb{X} \sim \Theta = \bigotimes_{i=1}^{|E|} \text{Bernoulli}(p_i)$ is binary random vector. It is easy to check that this setting satisfies our assumptions (A2)–(A6). Here, for the results in Sect. 3, we treat $\mathscr{P} = [0, 1]^m$ and interpret $\mathcal{F}_{\text{NR}}$ as the intersection of the set of interest with $\mathscr{X}_{\mathfrak{b}}$, an LWP. The related robust network reliability problem (RNR) computes $\max_{x \in \mathscr{X}_{\mathfrak{b}}} \text{MLU}(x)$ and can be written by dualizing $\text{MLU}(x)$ as follows:

$$\text{(RNR):} \quad \max_{\lambda, v, x} \quad \sum_{\substack{i,t \in V \\ i \neq t}} v_{it} d_{it} - \sum_{t \in V} v_{tt} \left(\sum_{j \in V} d_{jt} \right) \quad (31a)$$

$$v_{it} - v_{jt} \leq \lambda_{ij} \qquad \forall \langle i, j \rangle \in E, \forall t \in V \quad (31b)$$

$$\sum_{\langle i, j \rangle \in E} \lambda_{ij} c_{ij} (1 - x_{ij}) \leq 1 \quad (31c)$$

$$\lambda_{ij} \geq 0 \qquad \forall \langle i, j \rangle \in E \quad (31d)$$

$$x \in \mathscr{X}_{\mathfrak{b}}. \quad (31e)$$

Moreover, Assumption (A1) is satisfied because, for any x, $(\lambda, v) = (0, 0)$ is a feasible solution. Consider a relaxation of RNR, obtained using RLT, where, as in Proposition 1, (31b) and (31d) are multiplied with constraints defining $\mathscr{X}_{\mathfrak{b}}$. This relaxation produces a weak bound [10, 28]. However, this can be remedied (see [10]) by first lifting (30) to an equivalent higher-dimensional formulation, Slack-MLU(x), before performing RLT. Slack-MLU(x) is obtained from (30) by (i) introducing additional new variables $a_{ij} \geq 0 \; \forall \langle i, j \rangle \in E$, (ii) replacing d with d' where, for all $(i, j) \in V \times V$, $d'_{ij} = d_{ij} + a_{ij} \delta_{\langle i, j \rangle \in E}$, (iii) replacing (30b) with $\sum_{t \in V} y_{ijt} \leq U c_{ij}(1 - x_{ij}) + a_{ij} \; \forall \langle i, j \rangle \in E$. For completeness, we prove the validity in Appendix K.

Proposition 4 *For a given $x = \{x_{ij}\}_{\langle i,j\rangle \in E}$, consider the multi-commodity flow problem in (30) and its lifted version described above as Slack-MLU(x). Then, given a feasible solution to either of the formulations, a feasible solution to the other can be constructed that has the same utilization.* □

So, it follows from Proposition 4 that $\Pr_\Theta(\mathcal{F}_{\text{NR}})$ can be expressed as $\Pr_\Theta(\{x \in \mathscr{X}_{\mathfrak{b}} : \nexists (y, a)$ feasible to Slack-MLU(x) with $U = 1\})$. Let $g(x)$ represent the indicator function of $\{x \in \mathscr{X}_{\mathfrak{b}}$: Slack-MLU$(x) > 1\}$. Then,

$$g(x) = \max_{\lambda, v} \quad \mathfrak{z}(\lambda, v, x) \tag{32a}$$

$$-\lambda_{ij} + v_{it} - v_{jt} \le 0 \quad \forall \langle i, j\rangle \in E,\ \forall t \in V \tag{32b}$$

$$\lambda_{ij} - v_{ij} + v_{jj} \le 0 \quad \forall \langle i, j\rangle \in E \tag{32c}$$

$$\mathfrak{z}(\lambda, v, x) \le 1 \tag{32d}$$

$$\lambda_{ij} \ge 0 \quad \forall \langle i, j\rangle \in E, \tag{32e}$$

where $\mathfrak{z}(\lambda, v, x) = \sum_{\langle i,j\rangle \in E} \lambda_{ij} c_{ij}(x_{ij} - 1) + \sum_{i,t \in V: i \ne t} d_{it} v_{it} - \sum_{t \in V} v_{tt}\big(\sum_{j \in V} d_{jt}\big)$.

Proposition 5 *Given $x \in \mathscr{X}_{\mathfrak{b}}$, we may equivalently require that $v_{ij} = \lambda_{ij}$ for all $\langle i, j\rangle \in E$ and $v_{tt} = 0$ for all $t \in V$ in (32).* □

5.1 Computational evaluation on network reliability test instances

We estimate $\Pr_\Theta(\mathcal{F}_{\text{NR}})$ on three network topologies (i) Geant ($|V| = 32$, $m = 100$), (ii) Highwind ($|V| = 16$, $m = 58$), and (iii) Deltacom ($|V| = 103$, $m = 302$), taken from the topology zoo [21]. We abbreviate them as G, H and D respectively. As in [9], we recursively removed one-degree nodes in the original topologies and used the gravity model [43] to generate traffic matrices with MLU in [0.6, 0.67]. Geant has $c_{ij} \in [1, 100]$ Gbps, whereas $c_{ij} = 1$ for the other topologies. Undirected links (i, j) were replaced with two directed links $i \to j$ and $j \to i$ of the same capacity. Unless mentioned otherwise, we report the probability $\Lambda_{\text{NR}}(\mathfrak{b}) := \Pr_\Theta\{\mathcal{F}_{\text{NR}} \mid \mathbb{X} \in \mathscr{X}_{\mathfrak{b}}\}$ instead of $\Pr_\Theta(\mathcal{F}_{\text{NR}})$. This is obtained from $\Pr_\Theta(\mathcal{F}_{\text{NR}})$ by estimating $\Pr_\Theta(\mathbb{X} \in \mathscr{X}_{\mathfrak{b}})$ using Theorem 2 (resp. Corollary 1) when an overestimate (resp. underestimate) is needed.

Our algorithms were implemented in Python, the LPs and IPs were solved using Gurobi 8.0 [17], while formulations using Bernstein approximation (29) were solved using MOSEK 9.1 [1]. The CPU used was Intel Xeon E5-2623 @3.00 GHz. We assume that all $\mathfrak{b}$-*simultaneous link failure scenarios* ($\mathfrak{b}$ failures) occur with equal probability and report conditional probabilities, which are much higher than the unconditional ones. We only report bounds that use our algorithms from Sects. 3 and 4.1. This is because bounds from relaxation of RNR using Proposition 4 or using the concave envelope building algorithm of Sect. 2.1 are not tight enough to be useful for our NR instances.

We denote a problem instance as T($\mathfrak{b}$), where T abbreviates the topology and $\mathfrak{b}$ is the number of failures. In Table 1, column labeled True reports the ground truth values

Table 1 Deterministic Bonferroni estimates to $\varLambda_{\mathrm{NR}}(\mathfrak{b}) = \Pr_{\Theta}(\mathcal{F}_{\mathrm{NR}} \mid \mathbb{X} \in \mathscr{X}_{\mathfrak{b}})$

T($\mathfrak{b}$)	True	$\sum$-Counting (V, G, C, B, W)	$\sum$-Bernstein (V, G, C, B, W)
H(1)	0.035	(0.035, 0.793, 0.035, 0.069, 0.035)	(0.052, 4.276, 1.310, 0.345, 1.741)
H(2)	0.075	(0.075, 1.653, 0.175, 0.179, 0.174)	(0.140, 8.119, 3.345, 1.590, 3.817)
H(3)	0.122	(0.123, 2.563, 0.412, 0.351, 0.410)	(0.234, 11.255, 5.224, 3.400, 5.680)
G(1)	0.040	(0.040, 1.050, 0.040, 0.090, 0.040)	(0.060, 2.130, 1.320, 0.360, 2.780)
G(2)	0.088	(0.089, 1.252, 0.140, 0.204, 0.145)	(0.166, 3.829, 2.834, 1.488, 4.272)
G(3)	0.142	(0.146, 1.665, 0.295, 0.391, 0.311)	(0.273, 5.325, 4.220, 3.174, 5.599)
D(1)	0.017	(0.020, 0.858, 0.364, 0.083, 0.017)	(0.030, 14.417, 3.755, 0.715, 3.911)
D(2)	0.037	(0.057, 4.848, 0.775, 0.205†, 0.084)	(0.114, 28.345, 8.147, 2.248, 7.206)
D(3)	–	(–, 10.690, 1.231, 0.419‡, 0.202)	(–, 39.669, 12.165, 4.397, 10.171)

† estimates are obtained by sparsifying with $\delta_s = 0.005$
‡ respresent the $\delta_s = 0.05$

of $\varLambda_{\mathrm{NR}}(\mathfrak{b})$ obtained by enumeration. Since we report probabilities when network performance is unacceptable, a lower value corresponds to better network performance. For D(3), enumeration was not possible in a reasonable time. We consider 5 types of approximations of the failure set, and report rigorous upper bounds on $\varLambda_{\mathrm{NR}}(\mathfrak{b})$ using these approximations.

For the first approximation, called the V-cut approximation (abbreviated as V), we use the constraints referred to as V-cuts that are identified by FIND- VIOLATE, and obtained as in Step 3 of Algorithm 1. For all the other approximations we derive policies to upper bound $\varLambda_{\mathrm{NR}}(\mathfrak{b})$.

Our second approximation, the G-cut approximation (abbreviated as G) uses Propositions 6 (in Appendix C), 4, and 5 to derive a restricted version of the affine policy. Numerical computations have shown that this formulation (detailed in Appendix L) results in good bounds for RNR [9, 10]. The corresponding formulation is referred to as Gen-R3. To derive this policy, we set $\mathfrak{b} = 1$ in Gen-R3. Then, we fix U to 1 in Constraint (L.1b) and negate the constraint to obtain Constraint (L.2), which we refer to as G-cut. Using this constraint, we identify scenarios where the Constraint (L.1b) can only be satisfied with $U > 1$. Since these scenarios cover all the scenarios where MLU exceeds 1, we overestimate the probability of a bad failure by the probability that a failure scenario violates (L.1b). Similarly, the cuts we describe next are also used to overestimate the probability of bad failures.

Next, we describe the C-cut approximation (abbreviated as C). This approximation uses the affine policy obtained by solving the following problem, which is derived by specializing (25) to NR:

$$\min_{v,l,r,p,a} \quad l + \frac{1}{\widetilde{\mathcal{L}}} \sum_{i=1}^{\bar{s}} p^i v_i \tag{33a}$$

$$\sum_{t \in V} r_{et} + \sum_{l \in E} p_{el} x_l^i \le U_i c_e (1 - x_e^i) + a_e x_e^i \qquad \forall e \in E, i \in [\bar{s}] \tag{33b}$$

$$U_i - 1 \geq v_i + l,\, v_i \geq 0 \qquad \forall i \in [\bar{s}]$$
$$(L.1c),\, (L.1d),\, (L.1e), \tag{33c}$$

where $\bar{s}$ is the number of sampled scenarios and p^i is the probability of scenario $x^i \in \mathscr{X}_{\mathfrak{b}}$. Variables r_{et}, p_{el} are as defined for (L.1), and variables v_i for $i \in [\bar{s}]$ and l are as defined for (25). We choose $\widetilde{\mathcal{L}} = 5$ and use the resulting policy in Constraint (L.2) to obtain a cut, which we refer to as the C-cut.

Our fourth approximation (abbreviated as B) uses a B-cut. This cut uses the affine policy obtained by solving the Bernstein approximation described in Sect. 4.2. This approximation requires that Assumption (A7) is satisfied, or that the failure probability of each link is the same. For the B-cut approximation, we solve (29) iteratively choosing $\bar{\epsilon}$ so that MLU is below 1.

The last approximation, W-cut approximation (abbreviated as W) uses the policy obtained by solving Gen-R3, but by restricting failure scenarios to those single-failures where MLU is below 1 (see [9]). The cut obtained using this policy is referred to as W-cut. The C- and W-cut approximations can be improved by adding other good failure (failures x' with $\text{MLU}(x') \leq 1$) scenarios, an extension we do not implement. These cuts suffice to show that our framework is flexible enough so that it can be used to estimate $\Lambda_{\text{NR}}(\mathfrak{b})$ in a variety of ways. In the column labeled $\sum$-Counting of Table 1, we use the counting algorithm of Theorem 2, to obtain a rigorous upper bound for $\Lambda_{\text{NR}}(\mathfrak{b})$ by summing up the probabilities for each of the SLWPs in the union. For V-cuts, we do not report the bound for D(3), since Algorithm 1 did not terminate within 36 hours of CPU time. In column labeled $\sum$-Bernstein, we report a similar bound, but compute the probability of each SLWP by optimizing (28), after replacing t' with $\frac{1}{t}$. Note that the bounds in $\sum$-Bernstein are much weaker than those obtained in $\sum$-Counting using Theorem 2. Also, this bound can only be obtained when the probability of link failures is the same. We remark that we did not use a sparsifier for computing entries in $\sum$-Counting except those marked with † and ‡, where we chose δ_s as 0.005 and 0.05 respectively. This is because, unlike other cuts, the B-cuts are dense, and thus the approximation benefits, in terms of computational time, from the use of sparsifier. To illustrate, for G(3), the Bonferroni estimate without the sparsifier takes 521 CPU seconds, while it takes 150 s (resp. 12 s) with $\delta_s = 0.005$ (resp. $\delta_s = 0.05$).

The rigorous bounds in Table 1 are conservative. Recall that each of the cuts, G, C, B, and W, is associated with a policy. Theorem 4 gives an algorithm to sample scenarios where an edge in the network violates the capacity constraint. This algorithm is then used as the Sample- Assign subroutine within Algorithm 2 to sample scenarios where at least one edge violates the capacity constraint. Such a sampling algorithm only samples scenarios where the corresponding policy does not perform well. Using this algorithm, we improve the Bonferroni estimate given in Table 1 by directly estimating the probability of the union of scenarios where at least one edge exceeds capacity. These improved estimates are given in Table 2, where we generated 100,000 samples with $Z_t = 1$ for each problem instance and each cut. Recall that V-cuts do not correspond to a specific policy. Rather, each V-cut is obtained by solving a nonlinear integer program that gives a dual solution to $\text{MLU}(x)$ described in (30). When we report the results for these cuts, we also report the number of V-cuts along with the probability of union of V-cuts.

As mentioned above, for G, C, B, and W cuts, the reported probabilities in Table 2 are estimates of the probability that a network using a certain routing policy does not perform well. On the other hand, with V-cuts, whenever Algorithm 1 terminates, Table 2 estimates a lower bound on the probability that no routing policy can achieve an MLU of less than one. Therefore, when Algorithm 1 terminates, the probability associated with V-cuts in Table 2 is expected to be close to the ground truth as is indeed the case. However, when Algorithm 1 does not terminate, as in the case of D(3), the reported probability in Table 2 is a lower estimate of the probability of bad failures. We mention that rejection sampling, as described at the end of Sect. 4.1, can be used to estimate the fraction of bad failures discovered using the V-cuts. In particular, we recognize that the policy corresponding to any of the cuts is more restrictive than a fully flexible network response. Although our discussion below applies to any of the cuts, we use W-cuts to illustrate the usage of rejection sampling. In particular, there may be an alternate routing strategy that allows the network to perform well even when the policy associated with W-cut fails to achieve the desired performance. Using Algorithm 2, we sample failures where the policy associated with W-cut fails, and run MLU(x) as in (30), on each of these failure scenarios to estimate the probability of such failures where an alternate routing recovers the network performance. Rejecting these scenarios, we are left with the sampled scenarios where the network does not perform well. Consider D(3). We generated 10,000 scenarios using Algorithm 2 where the policy associated with W-cuts does not perform well. Among these, the network can handle 40.5% of the scenarios using a different routing strategy. As a result, our estimate for the probability of scenarios that this network cannot handle is $0.405 \times 0.15 \approx 0.061$ which is also the estimate obtained using the V-cuts. This suggests that the 432 V-cuts that we derived using Algorithm 1 are able to identify almost all scenarios where the network does not perform well.

Although, we have reported conditional failure probabilities throughout, unconditional probabilities over $[0, 1]^m$ are obtained either by dropping the constraint $\sum_{i=1}^{m} x_i = \mathfrak{b}$ or by aggregating the results by varying $\mathfrak{b}$. As mentioned above, the estimates of these unconditional probabilities can be further improved using rejection sampling on the samples generated by Algorithm 2. For instance, consider $p = 0.001$ as the probability of each link failure for Geant. Using B-cuts, we generated 10,000 samples via Algorithm 2, to find that 79.49% of the sampled scenarios were bad. In contrast, only 0.40% of all scenarios are bad. Our computations show that our algorithms prune the sample space significantly and the probability estimates we obtain are reasonably close to the ground truth values.

6 Conclusions

In this paper, we developed methods to estimate the probability that the optimal value of a convex program, satisfying certain structural assumptions, exceeds a given threshold. We used convexification, robust counterpart, and chance-constrained optimization techniques to cover the event set of interest by a union of sets and devised new approx-

Table 2 Estimates for $\Lambda_{\text{NR}}(\mathfrak{b})$ via Algorithm 2

T($\mathfrak{b}$)	True	Union Pr.:	V	G	C	B	W
H(1)	0.035		(0.035, 1)	0.328	0.035	0.035	0.035
H(2)	0.075		(0.075, 12)	0.593	0.147	0.093	0.153
H(3)	0.122		(0.122, 33)	0.778	0.311	0.185	0.326
G(1)	0.040		(0.040, 2)	1	0.040	0.050	0.040
G(2)	0.088		(0.088, 50)	0.999	0.121	0.113	0.127
G(3)	0.142		(0.142, 71)	1	0.232	0.228	0.248
D(1)	0.017		(0.017, 3)	0.480	0.192	0.066	0.017
D(2)	0.037		(0.037, 179)	0.998	0.366	0.156†	0.068
D(3)	–		(0.061, 432)	0.997	0.514	0.295‡	0.150

† estimates are obtained by sparsifying with $\delta_s = 0.005$ ‡ respresent the $\delta_s = 0.05$)

imate sampling and counting techniques to estimate the probability of this union. Our techniques effectively prune uninteresting scenarios from the sample space. To our knowledge, this is the first work to use affine policies with approximate counting techniques to derive bounds on probability quantification problems. We considered the network reliability (NR) problem which determines the probability of failures where network utilization exceeds one. Our computational results on NR are encouraging and, to our knowledge, the first non-trivial bounds obtained in polynomial time on the probability of bad failure scenarios.

Acknowledgements We acknowledge Shabbir Ahmed for his insightful comments at Dagstuhl on the use of Markov inequality for OUQ and Sanjay G. Rao for extensive discussions on NR. The second author would like to acknowledge the funding provided by NSF CMMI-1727989 and AFOSR 21RT0453

A List of assumptions

Here, we will briefly describe the assumptions we make in different parts of the paper.

1. In Sect. 2, for relating RLT to affine and polynomial policies, we assume:
   - There is no duality gap between CP$(\cdot)$ and CD$(\cdot)$ (A1).
2. For deriving the column generation algorithm and to show convergence of RLT at the m^{th} level in Theorem 1, we assume in Sect. 2.1 that:
   - The distribution of $\mathbb{X}$ is supported on a finite set of points $\mathscr{T}$ in $\mathscr{P}$ (A2).
   - $\mathbb{K} = \mathbb{R}^p_+$ (A3).
   - $\mathscr{T} \subseteq \{0, 1\}^m$ and an inequality description of conv$(\mathscr{T})$ is available (A4).

   Additionally, when $\mathscr{T}$ consists of the vertices of a simplex, we show in Proposition 2 that the concave envelope of the indicator function can be used to compute $\Pr_*(\mathcal{F})$. The column generation algorithm also assumes that expectations of a set of functions of the random variable $\mathbb{X}$, denoted as $\{\mathfrak{f}_\alpha(\mathbb{X}), \alpha \in \bar{\Gamma} \subseteq \mathbb{N}^m\}$, are known.

3. In Sects. 3 and 4 we devise counting and sampling algorithms by assuming that:
   - $\mathscr{P} = \text{conv}(\mathscr{T}) = [0,1]^m$ and $\mathbb{X} \in \{0,1\}^m$, with distribution $\Theta = \bigotimes_{i=1}^m \text{Bernoulli}(p_i)$ (tensor product of m independent Bernoulli distributions). Moreover, we assume that $p_i = \frac{a_i}{n_i}$, where $a_i, n_i \in \mathbb{N}$, and $\text{GCF}(a_i, n_i) = 1$ (A5).
   - Without loss of generality, the weights of the general inequality defining each Sliced low weight polytope (SLWP) are non-negative *i.e.*,$w_i \in \mathbb{Z}_{\geq 0}$ for all $i \in [m]$ (A6).
4. We derive the Bernstein approximation by assuming in Sect. 4.2 that:
   - $S' = \left\{\sum_{i=1}^m w_i x_i \geq C, \sum_{i=1}^m x_i = \mathfrak{b}, \text{ where } w_i \geq 0 \ \forall i \in [m]\right\}$ and $\Pr_\Theta(\mathbb{X}_i = 1) = p$ for all $i \in [m]$ (A7).

B Proof of Proposition 1

Let $y = P^\top x + q$. For y to be feasible, $(AP^\top + B)x + Aq \leq_{\mathbb{K}} c$ for all x satisfying $\mathfrak{C}x \leq \mathfrak{d}$. Then, $(AP^\top+B)x+Aq-c \leq_{\mathbb{K}} U^\top\mathfrak{C}x - U^\top\mathfrak{d} = U^\top(\mathfrak{C}x-\mathfrak{d}) \leq_{\mathbb{K}} 0$, where the first inequality follows from (6b) and (6c) and the last inequality because $U^\top(\mathfrak{C}x - \mathfrak{d})$, by (6e) is a non-positive conic combination of vectors in $\mathbb{K}$. Moreover, the objective, $e^\top(P^\top x+q) = \underline{\Theta}^\top\mathfrak{C}x+e^\top q \leq \underline{\Theta}^\top\mathfrak{d}+e^\top q$, where the first equality is from (6d) and the second inequality is because $\underline{\Theta} \geq 0$ and $\mathfrak{C}x \leq \mathfrak{d}$. This shows that the feasible solutions in (6) describe an affine policy and the objective function value overestimates that of the corresponding affine policy. We now show that relaxation is exact when $\mathbb{K} = \mathbb{R}^p_+$ and $\mathscr{P} \neq \emptyset$. For an affine policy to be feasible, $(A_k P^\top + B_k)x + A_k q - c_k \leq 0$ for all x satisfying $\mathfrak{C}x \leq \mathfrak{d}$ and for all $k \in [p]$, where $A_k^\top \in \mathbb{R}^n$, $B_k^\top \in \mathbb{R}^m$ represent the k^{th} row of A and B respectively, and $c_k \in \mathbb{R}$ represents the k^{th} entry of c. In other words, for $a^\top = -(A_k P^\top + B_k)$ and $b = A_k q - c_k$, it follows that $\{x \mid a^\top x < b, \mathfrak{C}x \leq \mathfrak{d}\} = \emptyset$. By Farkas' Lemma, one of $\mathscr{S}_1$ and $\mathscr{S}_2$ is therefore feasible, where

$$\mathscr{S}_1 := \left\{(\lambda, \mu) \in \mathbb{R}_{++} \times \mathbb{R}^l_+ \;\middle|\; \lambda a^\top + \mu^\top\mathfrak{C} = 0, \lambda b + \mu^\top\mathfrak{d} \leq 0\right\}$$
$$\mathscr{S}_2 := \left\{(\lambda, \mu) \in \mathbb{R}_{+} \times \mathbb{R}^l_+ \;\middle|\; \lambda a^\top + \mu^\top\mathfrak{C} = 0, \lambda b + \mu^\top\mathfrak{d} < 0\right\}.$$

In $\mathscr{S}_1$, to see the equivalence scale λ to 1 and set $\mu = \widetilde{U}_k$, the k^{th} column of U. In $\mathscr{S}_2$, we assume $\lambda = 0$, otherwise we obtain a solution to $\mathscr{S}_1$. Therefore, there is a non-negative μ such that $\mu^\top\mathfrak{C} = 0$ and $\mu^\top\mathfrak{d} < 0$, which, by Farkas' Lemma contradicts that $\mathscr{P} \neq \emptyset$. □

C Extension of Proposition 1 to consider polynomial policies

The design of such polynomial policies relates to the use of polynomial chaos expansion for structured representation of uncertainty in chance-constrained optimization; see [30] for its use in optimal power flow. Suppose $y_j = \sum_{\alpha\in\gamma_j} g_j^\alpha x^\alpha$ for $j \in [n]$, where $\alpha = (\alpha_1, \ldots, \alpha_m) \in \gamma_j \subseteq \mathbb{N}^m$, $g_j^\alpha \in \mathbb{R}$, and x^α represents the monomial

$x_1^{\alpha_1} \cdots x_m^{\alpha_m}$. Let $Y(x) = \{y \in \mathbb{R}^n : A(x)y + \mathscr{X}(x) \leq_{\mathbb{K}} c\}$, where $A(x)$ is a $p \times n$ matrix of polynomial functions, $\mathscr{X}(x)$ is a p sized vector of polynomials, such that $A(x)_{kj} = \sum_{\beta \in S_{kj}} a_{kj}^{\beta} x^{\beta}$ and $\mathscr{X}(x)_k = \sum_{\beta \in S_{k0}} a_{k0}^{\beta} x^{\beta}$ for some sets S_{kj} and S_{k0}. Let $S' {:=} \{\alpha' : \exists (j,k) \text{ such that } \alpha' = \alpha + \beta, \alpha \in \gamma_j, \beta \in S_{kj}\}$. Assume that $\mathscr{P} = \{x' : \mathfrak{C}' x' \leq \mathfrak{d}'\}$ is a linear relaxation of $\{x' : x'_{\alpha'} = x^{\alpha'} \forall \alpha' \in S', \mathfrak{C} x \leq \mathfrak{d}\}$. Let $\varsigma = \max_{x \in \mathscr{P}} \min_{y \in Y(x)} e^{\intercal} y$ and restrict y to a polynomial policy to define:

$$
\begin{aligned}
\Psi_1^* {:=} \min_{\xi, g} \quad & \xi \\
& \xi \geq e^{\intercal} \sum_{\alpha \in \gamma_j} g_j^{\alpha} x'_{\alpha} && \forall x' \in \{\mathfrak{C}' x \leq \mathfrak{d}'\} \\
& \sum_{j \in [n]} \sum_{\beta \in S_{kj}} \sum_{\alpha \in \gamma_j} a_{kj}^{\beta} g_j^{\alpha} x'_{\alpha+\beta} + \sum_{\beta \in S_{k0}} a_{k0}^{\beta} x'_{\beta} \leq_{\mathbb{K}} c_k && \forall k \in [p] \\
& && \forall x' \in \{\mathfrak{C}' x \leq \mathfrak{d}'\}.
\end{aligned}
$$

Then, assuming $\{x' : \mathfrak{C}' x' \leq \mathfrak{d}'\}$ is not empty and $\mathbb{K} = \mathbb{R}_+^p$, dualization allows us to succinctly express the constraints for all x' so that $\Psi_D^* = \Psi_1^*$, where:

$$
\begin{aligned}
\Psi_D^* = \min_{g, \Theta, U} \quad & \sum_{r \in [l]} \mathfrak{d}'_r \Theta_r + \sum_{j : 0 \in \gamma_j} e_j g_j^0 && && \text{(C.2a)} \\
& \sum_{r \in [l]} U_{rk} \mathfrak{C}'_{r\alpha} = \sum_{j \in [n]} \sum_{\alpha - \alpha' \in S_{kj}} \sum_{\alpha' \in \gamma_j} a_{kj}^{\alpha - \alpha'} g_j^{\alpha} + a_{k0}^{\alpha} && \forall k \in [p], \alpha \in \gamma_j && \\
& && && \text{(C.2b)} \\
& \sum_{r \in [l]} U_{rk} \mathfrak{d}'_r + \sum_{j : 0 \in S_{k,j}} a_{kj}^0 g_j^0 + \sum_{0 \in S_{k0}} a_{k0}^0 \leq c_k && \forall k \in [p] && \text{(C.2c)} \\
& \sum_{r \in [l]} \Theta_r \mathfrak{C}'_{r\alpha} = \sum_{j : \alpha \in \gamma_j} g_j^{\alpha} e_j && \forall \alpha \in \gamma_j && \text{(C.2d)} \\
& \Theta \geq 0, \; U_{\cdot k} \geq 0 && \forall k \in [p]. && \text{(C.2e)}
\end{aligned}
$$

Proposition 6 *Assume that $\bar{x} \in \mathscr{P}$, and there exists a $\bar{w} \in -\mathbb{K}^*$ such that $\bar{w}^{\intercal} A(\bar{x}) = e^{\intercal}$, and that strong duality holds for the inner problem, i.e.,*

$$
\min_{y \in Y(x)} e^{\intercal} y = CD(x) {:=} \max_{w} \left\{ w^{\intercal} \left(c - \mathscr{X}(x) \right) \,\middle|\, w^{\intercal} A(x) = e^{\intercal}, w \leq_{\mathbb{K}}^* 0 \right\},
$$

where $Y(x) = \{y \in \mathbb{R}^n : A(x)y + \mathscr{X}(x) \leq_{\mathbb{K}} c\}$, such that $A(x)_{kj}$ and $\mathscr{X}(x)_k$ for $k \in [p]$ and $j \in [n]$, are as discussed above. Then, if $\mathbb{K} = \mathbb{R}_+^p$, there is an RLT relaxation of $\varsigma = \max_{x \in \mathscr{P}} \min_{y \in Y(x)} e^{\intercal} y$ which dualizes (C.2) and has the same optimal value.

Proof By strong duality, $\varsigma = \max_{x \in \mathscr{P}} \text{CD}(x)$. We obtain the following constraints by taking products of equality constraints in $\text{CD}(x)$ with x^{α} and inequalities with $\mathfrak{C}' x' \leq \mathfrak{d}'$ that relax the monomial definitions:

$$
\sum_{k \in [p]} \sum_{\beta \in S_{kj}} w_k a_{kj}^{\beta} x^{\alpha+\beta} = x^{\alpha} e_j \; \forall \alpha \in \gamma_j, \forall j \in [n]; \quad \text{and} \quad (\mathfrak{d}'_r - \mathfrak{C}'_r x') w^{\intercal} \leq_{\mathbb{K}^*} 0.
$$

Upon linearization, we obtain:

$$\underline{\Delta} := \max_{w,x,w',w''} \quad w^{\mathsf{T}}(c - B'x') \tag{C.3a}$$

$$w'A' = M' \tag{C.3b}$$

$$\mathfrak{d}'_r w^{\mathsf{T}} - \mathfrak{C}'_r w'' \leq_{\mathbb{K}^*} 0 \qquad \forall r \in [l] \tag{C.3c}$$

$$\mathfrak{C}'x' \leq \mathfrak{d}' \tag{C.3d}$$

$$w \leq_{\mathbb{K}^*} 0, \tag{C.3e}$$

where: (i) $w''_{(\alpha,k)}$ linearizes $x^\alpha w^{\mathsf{T}}$ and $x'_\alpha w^{\mathsf{T}}$, (ii) whenever $\alpha' + \beta' = \alpha$, $j \in [n]$, $\alpha' \in \gamma_j$, and $\beta' \in S_{kj}$, $w'_{(j,\alpha'),(\beta',k)} = w''_{(\alpha,k)}$, (iii) for all $k \in [p]$, $j \in [n]$, $A'_{(\beta,k),j} = a^\beta_{kj}$ if $\beta \in S_{kj}$ and 0 otherwise, (iv) for all $\alpha \in \gamma_j$ and $j \in [n]$, $M'_{(j,\alpha)} = x^\alpha e_j$, and (v) for all $k \in [p]$, $B'_{(k,\beta)} = a^\beta_{k0}$ if $\beta \in S_{k0}$ and 0 otherwise. Let P', $\{U'_r\}_{r\in[l]}$, and Θ' be the dual variables to the equations (C.3b), (C.3c) and (C.3d) respectively. Given that $(\bar{w}, \bar{x})$ is feasible for $\max_{x\in\mathscr{P}} \mathrm{CD}(x)$, its relaxation (C.3) used to compute $\underline{\Delta}$ is also feasible. When $\mathbb{K} = \mathbb{R}^p_+$, (C.3) is a linear program and so has no duality gap. In general, its dual is:

$$\min_{\Theta',U',P'} \quad \sum_{j:0\in\gamma_j} P'_{j0} e_j + \sum_{r\in[l]} \Theta'_r \mathfrak{d}'_r \tag{C.4a}$$

$$-U'^{\mathsf{T}}\mathfrak{C}' + F' + B' = 0 \tag{C.4b}$$

$$U'^{\mathsf{T}}\mathfrak{d}' + L \leq_{\mathbb{K}} c \tag{C.4c}$$

$$h + \mathfrak{C}'^{\mathsf{T}}\Theta' = 0 \tag{C.4d}$$

$$\Theta' \geq 0, U'_r \geq_{\mathbb{K}} 0 \qquad \forall r \in [l], \tag{C.4e}$$

where, for $k \in [p]$ and $\alpha \in \gamma_j$, $F'_{\alpha,k} = \sum_j \sum_{\alpha'\in\gamma_j} \sum_{\beta'=\alpha-\alpha'\in S_{kj}} a^{\beta'}_{kj} P'_{\alpha' j}$, and, for $k \in [p]$, $L_k = \sum_{j\in[n]} \sum_{0\in\gamma_j} \sum_{k:0\in S_{kj}} a^0_{kj} P'_{0j}$. Finally, for $\alpha \in \gamma_j$, $h_\alpha = -\sum_{j:\alpha\in\gamma_j} P_{\alpha j} e_j$. When $\mathbb{K} = \mathbb{R}^p_+$, we obtain (C.2) by replacing (Θ', U', P') in (C.4) with (Θ, U, g). □

The equivalence in Propositions 1 and 6 holds when $\mathbb{K}$ has a tractable linear inequality representation. To reduce the more general case to that for $\mathbb{R}^p_+$, we write $U \in \mathbb{K}$ as $\mathcal{G}U \geq 0$ for some $\mathcal{G}$ and replace $Ay + Bx \leq_{\mathbb{K}} c$ with $\mathcal{G}Ay + \mathcal{G}Bx \leq \mathcal{G}c$.

D Proof of Proposition 2

By definition, $\hat{\mathbb{1}}_E(\mathbb{E}_*[\mathbb{X}]) = \max_\lambda \left\{\sum_{i\in\mathcal{I}} \lambda_i \mathbb{1}_{\mathcal{F}}(x^i) \middle| \sum_{i\in\mathcal{I}} \lambda_i x^i = \mathbb{E}_*[\mathbb{X}], \lambda \geq 0, \sum_{i\in\mathcal{I}} \lambda_i = 1\right\}$, where $\lambda = \{\lambda_i\}_{i\in\mathcal{I}}$ and $\{x^i\}_{i\in\mathcal{I}}$ are the extreme points in $\mathscr{T}$. There is a unique feasible solution with $\lambda_i = \Pr_*(\mathbb{X} = x^i)$. So, $\hat{\mathbb{1}}_E(\mathbb{E}_*[\mathbb{X}]) = \sum_{i\in\mathcal{I}} \Pr_*(\mathbb{X} = x^i)\mathbb{1}_{\mathcal{F}}(x^i) = \mathbb{E}_*[\mathbb{1}_{\mathcal{F}}(\mathbb{X})]$. □

E Proof of Proposition 3

We first write $\hat{\mathbb{1}}(x)$ as $\max_{(w,v)\in\mathcal{S}} h(x, w, v)$. Then, for any $\bar{x} \in \mathscr{T}$, $\max_{(w,v)\in\mathcal{S}} r(\bar{x}, w, v) = \mathbb{1}_{\mathcal{F}\cap\mathscr{T}}(\bar{x}) \geq 0$. Since $\max_{(w,v)\in\mathcal{S}} h(x, w, v)$ is concave and, for $x \in \text{conv}(\mathscr{T})$, is larger than $\mathbb{1}_{\mathcal{F}\cap\mathscr{T}}(x)$, it follows that $\hat{\mathbb{1}}(x) \geq \hat{\mathbb{1}}_E(x)$. For the converse, observe that, for all $(\bar{x}, w, v) \in \text{conv}(\mathscr{T}) \times \mathcal{S}$, $r(\bar{x}, w, v) \leq \mathbb{1}_{\mathcal{F}\cap\mathscr{T}}(\bar{x}) \leq \hat{\mathbb{1}}_E(x)$. Since $\hat{\mathbb{1}}_E(x)$ is concave, it follows that $h(\bar{x}, w, v) \leq \hat{\mathbb{1}}_E(x)$ and, so, $\hat{\mathbb{1}}(x) = \max_{(w,v)\in\mathcal{S}} h(x, w, v) \leq \hat{\mathbb{1}}_E(x)$. □

F Proof of Theorem 1

We prove the result by showing that $\varphi^R(b_\alpha)$ computes the optimal value in (12). Let $\mathscr{M}_J(x) := \prod_{j\in J} x_j$. Let α be defined so $\alpha_j = 1$ if $j \in J$ and 0 otherwise. Then, $\mathscr{M}_J(x) = x^\alpha$. Clearly, for any variable z, the functions $z\mathscr{M}_J(x)$ and $z\mathfrak{M}_J(x)$ for $J \subseteq [m]$ form bases of the same vector space of functions. Indeed, $z\mathfrak{M}_J(x) = \sum_{J':J\subseteq J'\subseteq[m]} (-1)^{|J'\setminus J|} z\mathscr{M}_{J'}(x)$. Conversely, we have $z\mathscr{M}_J(x) = \sum_{J':J\subseteq J'\subseteq[m]} z\mathfrak{M}_{J'}(x)$. Therefore, we write the RLT relaxation obtained from (13) equivalently without expanding the multilinear terms, instead linearizing $\varphi\mathfrak{M}_J(x)$, $w\mathfrak{M}_J(x)$, $v\mathfrak{M}_J(x)$, and $\mathfrak{M}_J(x)$ directly using φ^J, w^J, v^J, and $\mathfrak{p}_J$ respectively. Since the former basis includes 1, we must also require that $\sum_{J':J'\subseteq[m]} z\mathfrak{M}_{J'}(x) = z$ for each $z \in \{\varphi, w, v, 1\}$. When z is φ, this shows that the objective (12a) matches that in (13a). The substitution $x_i^2 = x_i$ replaces $x_i\mathfrak{M}_J(x)$ with $\mathfrak{X}_i^J\mathfrak{M}_J(x)$. This is linearized as $\mathfrak{X}_i^J\mathfrak{p}_J$ in (13e) while $\mathfrak{M}_J(x)w^\intercal Bx$ in (13b) is replaced with $(w^J)^\intercal B\mathfrak{X}^J$. The constraints (12b), (12c), and (12d) now follow easily from the linearizations of (13b), (13c) and (13d).

We show that the set defined by the linearization of (13e), denoted as X' is: $X = \left\{(\mathfrak{p}_J)_{J\subseteq[m]} : \mathfrak{p}_J \geq 0,\ J \subseteq [m];\ \sum_{J\subseteq[m]} \mathfrak{p}_J = 1;\ \mathfrak{p}_J = 0 \text{ if } \mathfrak{X}^J \notin \mathscr{T}\right\}$. Note that X models the probability distributions with support on $\mathscr{T}$. Because $x_i\mathfrak{M}_J(x)$ linearizes to $\mathfrak{X}_i^J\mathfrak{p}_J$, X' has the same variables as X. We first show that $X' \subseteq X$. Observe that $\sum_{J':J'\subseteq[m]} \mathfrak{M}_{J'}(x) = 1$, linearizes to $\sum_{J':J'\subseteq[m]} \mathfrak{p}_{J'} = 1$. Moreover, for any $j \in J$, (resp. $j \in J^C$), $x_j \geq 0$ (resp. $1 - x_j \geq 0$) is implied by conv$(\mathscr{T})$. Thus, the linearization of $x_j\mathfrak{M}_J(x) \geq 0$ (resp. $(1 - x_j)\mathfrak{M}_J(x) \geq 0$) is implied by (13e) and yields $\mathfrak{p}_J \geq 0$. Now, consider any $\mathfrak{X}^J \notin \text{conv}(\mathscr{T})$. Then, if $\mathfrak{p}_J > 0$, we obtain a contradiction since (13e) requires that $\mathfrak{p}_J\mathfrak{X}^J \in \mathfrak{p}_J\,\text{conv}(\mathscr{T})$. Therefore, $\mathfrak{p}_J = 0$ whenever $\mathfrak{X}^J \notin \mathscr{T}$. Now, we show that $X' \supseteq X$. Since X' is convex, it suffices to show that the extreme points of X are contained in X'. It can be verified that if $\mathfrak{X}^J \in \mathscr{T}$ then the solution $\mathfrak{p}_J = 1$ and $\mathfrak{p}_{J'} = 0$ for $J' \neq J$ is feasible to X'.

Finally, we show that $x_\alpha = b_\alpha$ is feasible to the linearization of (13e). Let $\bar{\mathfrak{p}}_J = \sum_{J'\subseteq J^C} (-1)^{|J'|} b_{\alpha(J\cup J')}$ for all $J \subseteq [m]$. Then, since b_α is the moment of x^α with support on $\mathscr{T}$, it follows that $\bar{\mathfrak{p}}_J \in X$. Let x_α linearize $\mathscr{M}_J(x)$, where $\alpha_j = 1$ if $j \in J$ and 0 otherwise. Observe that, with this linearization, (13e) yields an affine transform of X, say $T(X)$, in the space of x_α variables. Then, $x_\alpha = \sum_{J':J\subseteq J'\subseteq[m]} \bar{\mathfrak{p}}_{J'}$ is feasible to $T(X)$. However, it can be easily verified that $\sum_{J':J\subseteq J'\subseteq[m]} \bar{\mathfrak{p}}_{J'} = \sum_{J':J'\subseteq[m]} \bar{\mathfrak{p}}_{J'}(\mathfrak{X}^{J'})^\alpha = b_\alpha$. The first equality is because

$(\mathfrak{X}^J)^\alpha = 1$ if $J \subseteq J'$ and 0 otherwise, while the second equality follows since $\bar{\mathfrak{p}}_J$ is the probability distribution corresponding to the moments b_α. Thus, $x_\alpha = b_\alpha$ is feasible to $T(X)$. Then, it follows that $\varphi^R(b_\alpha)$ computes $\Pr_\Theta(\mathcal{F})$ as in (12). □

G Proof of Theorem 2

We denote the maximum value of $s^{\geq}((i,j),*)$ by M_i, where $s^{\geq}((i,j),c) = \sum_{\tilde{c}\geq c} s((i,j),\tilde{c})$. For any i, range$(j)_i$ is the range of possible values of j. For $i \leq K_2$, $M_i \leq \binom{i}{j}\mathcal{T}^i$, $j \in \left[\max\{0, i + \mathfrak{b} - K_2\}, \min\{K_2, K_1 + \mathfrak{b}, i\}\right]$, and range$(j)_i = \min\{K_2, i + \mathfrak{b} - 2K_2, K_1 + \mathfrak{b}, m - i\}$. For $i > K_2$, there is a $l \in \left[0, \min\{K_2 - j, i - K_2\}\right]$ so that we select $j + l$ (resp. $i - K_2 - l$) variables from $\{1,\ldots,K_2\}$ (resp. $\{K_2+1,\ldots,i\}$) to set to 1 (resp. 0). It follows that $M_i \leq \binom{i}{j+i-K_2}\mathcal{T}^i$, $j \in \left[\mathfrak{b}, \min\{K_2, m-i+\mathfrak{b}\}\right]$, and range$(j)_i = \min\{K_2 - \mathfrak{b}, m - i\}$. We choose a sparsification parameter, δ_s, to perform $(1+\delta_s)$ sparsification of each $s((i,j),\tilde{c})$. Observe that $\log_{(1+\delta_s)}\binom{i}{j} \leq \min\{\frac{i}{2}, j\}\log_{1+\delta_s}\frac{i\exp(1)}{\min\{\frac{i}{2},j\}}$. Since the time-complexity of summing, shifting, and querying function lists is bounded by their size, the time complexity is $O\big(m\underline{\Theta}\big(\xi\log_{1+\delta_s}(m/\xi) + m\log_{1+\delta_s}\mathcal{T}\big)\big)$. The time complexity in Theorem 2, follows by choosing $\delta = (1+\epsilon_s)^{1/m} - 1$, and using $\ln(1+\epsilon_s) \geq \epsilon_s/2$ for $\epsilon_s \in (0,1)$. □

H Proof of Theorem 3

Consider a $t+1$ dimensional DAG, where the $(l+1)^{\text{st}}$ dimension corresponds to the l^{th} low weight constraint. Let $s((i, j_1,\ldots,j_t),*)$ be the list of all pairs $(c, s((i, j_1,\ldots,j_t),c))$. For $\mathcal{T} \geq \max_i n_i$, if M is the maximum value of $s^{\geq}((i, j_1, \ldots, j_t),*)$, then $M \leq (2\mathcal{T})^i$ as there are 2^i solutions, each of which occurs at most $\mathcal{T}^i$ times. Moreover, let γ be such that $\gamma \geq \max_k w_k^l - \min_k w_k^l$ for all $l \in [t]$. Thus, the l^{th} low weight constraint at the i^{th} slice has at most $i\gamma$ values. Since there are t low weight constraints, the number of nodes with first coordinate i is bounded by $(m\gamma)^t$. Then, after a $1+\delta_s$ sparsification the cumulative length of lists is bounded by $m(m\gamma)^t\log_{1+\delta_s}(2\mathcal{T})^m$. For a $1+\epsilon_s$ approximation, with $1+\delta_s = (1+\epsilon_s)^{1/m}$, the time complexity is $O[\epsilon_s^{-1}m^{t+3}\gamma^t\ln\mathcal{T}]$. □

I A lower estimate for probability of 0–1 solutions to a SLWP

Given a function $f: \mathbb{Z}^+ \to \mathbb{Z}^+$ and an approximation parameter $\epsilon_s > 0$, we say $F: \mathbb{Z}^+ \to \mathbb{Z}^+$ (resp. $\underline{F}: \mathbb{Z}^+ \to \mathbb{Z}^+$) is a $(1-\epsilon_s)$ function approximation (resp. sum-approximation) of f if, for all x, $(1-\epsilon_s)f(x) \leq F(x) \leq f(x)$ (resp. $(1-\epsilon_s)f^{\geq}(x) \leq \underline{F}^{\geq}(x) \leq f^{\geq}(x)$). The properties in Lemma 2 follow easily for $(1-\epsilon_s)$ sum-approximation of functions. The sparsifier takes as input a function f and a parameter $\delta_s > 0$. We partition values of $f^{\geq}$ into $[r_{i+1}, r_i)$, where $r_0 = \max_c f^{\geq}(c)$ and if $r_i > 0$, then $r_{i+1} = \min\{r_i - 1, \lceil(1-\delta_s)r_i\rceil\}$. Let $c_i = \min_c\{c \mid f^{\geq}(c) \leq r_i\}$.

For any c, we define $l(c) = \min_i\{c_i \mid c_i > c\}$. If $l(c)$ is finite, $\underline{F}^{\geq}(c) = f^{\geq}(l(c)-1)$. Otherwise, $\underline{F}^{\geq}(c) = \lim_{x\to\infty} f^{\geq}(x)$. Then, $\underline{F}^{\geq}(x)$ is a $(1-\delta_s)$ approximation of $f^{\geq}(x)$. As a consequence of the Theorem 2, we obtain the time complexity to compute $|\underline{S_s}|_\Theta$ such that for any given $\epsilon_s \in (0,1)$, $(1-\epsilon_s)|S_s|_\Theta \leq |\underline{S_s}|_\Theta \leq |S_s|_\Theta$.

Corollary 1 *Given S_s, Θ as in (14), (A5) respectively, and an error parameter $\epsilon_s \in (0,1)$, we can deterministically compute a $1-\epsilon_s$ relative error approximation of $|S_s|_\Theta$ in time given as in Theorem 2.*

Proof When we use a $(1-\delta_s)$ sparsifier, the time to compute $|S_s|_\Theta$ is $O\big(m\underline{\Theta}\big[\xi \log_{\frac{1}{(1-\delta_s)}}\big(\frac{m}{\xi}\big) + m\log_{\frac{1}{1-\delta_s}} \mathcal{T}\big]\big)$. To control the approximation error, we set $(1-\delta_s)^m = (1-\epsilon_s)$. Then, we obtain the same time-complexity as in Theorem 2 using $\ln(1-\epsilon_s)^{-1} \geq \epsilon_s$. □

J Proof of Theorem 4

We write the solution set of S_Ω as $\bigcup_J S(J)$, where each $S(J) = \{x \in \{0,1\}^m : \sum_{i=1}^m w_i x_i \geq C, \Omega x = J\}$. For a given $\widetilde{x} \in \{0,1\}^m$, we first compute $\Pr_\Theta(\mathbb{X} = \widetilde{x}|\mathbb{X} \in S_\Omega)$. To do so, we will compute $\Pr_\Theta(\mathbb{X} = \widetilde{x} \mid \mathbb{X} \in S(J))$. Let $s_i(j,c) = \big\{x : \sum_{k=1}^i w_k x_k \geq c, \Omega_{\cdot,1:i} x_{1:i} = j\big\}$ and $s'_i(j,c) = s_i(j,c) \cap \{x : x_k = \widetilde{x}_k \forall k > i\}$. Define $c(i) = C - \sum_{k=i+1}^m w_k \widetilde{x}_k$ and $j^J(i) = J - \Omega_{\cdot,i+1:m}\widetilde{x}_{i+1:m}$. Clearly, if $\widetilde{x} \in S(J)$, $j^J(0) = 0$, $c(0) \leq 0$, and $s'_0\big(0, c(0)\big) = \{\widetilde{x}\}$. Observe that $s'_r\big(j^J(r), c(r)\big) \subseteq s'_{r+1}\big(j^J(r+1), c(r+1)\big)$ because if $x \in s'_r\big(j^J(r), c(r)\big)$, we have $x_k = \widetilde{x}_k$ for $k > r$, $\sum_{k=1}^r w_k x_k \geq c(r) = c(r+1) - w_{r+1}\widetilde{x}_{r+1}$, and $\Omega_{\cdot,1:r} x_{1:r} = j^J(r) = j^J(r+1) - \Omega_{\cdot,r+1}\widetilde{x}_{r+1}$, showing that $x \in s'_{r+1}\big(j^J(r+1), c(r+1)\big)$. Then, $s'_0\big(j^J(0), c(0)\big) \subseteq s'_1\big(j^J(1), c(1)\big) \subseteq \cdots \subseteq s'_m\big(j^J(m), c(m)\big) = S(J)$, and we have

$$\Pr_\Theta\left(\mathbb{X} = \widetilde{x}|\mathbb{X} \in S_\Omega\right) = \sum_J \left(\Pr_\Theta(\mathbb{X} \in S(J) \mid \mathbb{X} \in S_\Omega) \prod_{i=1}^m \Pr_\Theta\left(\mathbb{X} \in s'_{i-1}\big(j^J(i-1), c(i-1)\big)\Big|\mathbb{X} \in s'_i(j^J(i), c(i))\right)\right). \tag{J.1}$$

Further, for all J and $i \in [m]$, $\Pr_\Theta\left(\mathbb{X} \in s'_{i-1}\big(j^J(i-1), c(i-1)\big)\Big|\mathbb{X} \in s'_i(j^J(i), c(i))\right)$ is:

$$\begin{aligned}
&\frac{\Pr_\Theta\big(\mathbb{X} \in s'_{i-1}(j^J(i-1), c(i-1))\big)}{\Pr_\Theta\big(\mathbb{X} \in s'_i(j^J(i), c(i))\big)} \\
&= \frac{\Pr_\Theta\big(\mathbb{X} \in s_{i-1}(j^J(i-1), c(i-1))\big)}{\Pr_\Theta\big(\mathbb{X} \in s_i(j^J(i), c(i))\big)} \Pr_\Theta(\mathbb{X}_i = \widetilde{x}_i) \\
&= \frac{|s_{i-1}(j^J(i-1), c(i-1))|_\Theta}{|s_i(j^J(i), c(i))|_\Theta} \Pr_\Theta(\mathbb{X}_i = \widetilde{x}_i) n_i = p_i^J \delta_{\widetilde{x}_i=0} + (1-p_i^J)\delta_{\widetilde{x}_i=1},
\end{aligned}$$

where $p_i^J = \frac{|s_{i-1}(j^J(i),c(i))|_\Theta}{|s_i(j^J(i),c(i))|_\Theta}(n_i - a_i)$ and $\delta_{\widetilde{x}=\mathfrak{a}}$ is 1 if $\widetilde{x} = \mathfrak{a}$ and 0 otherwise. The first equality is because the event $\mathbb{X} \in s_i(j,c)$ is independent of $\{\mathbb{X}_{i'}\}_{i'=i+1}^m$ and, therefore, $\Pr_\Theta\left(\mathbb{X} \in s_i'(j^J(i), c(i))\right) = \Pr_\Theta\left(\mathbb{X} \in s_i(j^J(i), c(i))\right) \prod_{i'=i+1}^m \Pr_\Theta(\mathbb{X}_{i'} = \widetilde{x}_{i'})$. The second equality follows since $|s_i(j^J(i), c(i))|_\Theta = \Pr_\Theta\left(\mathbb{X} \in s_i(j^J(i), c(i))\right) \prod_{i'=1}^i n_{i'}$ and the last equality is because $|s_{i-1}(j^J(i), c(i))|_\Theta (n_i - a_i) + |s_{i-1}(j^J(i) - \Omega_{.,i}, c(i) - w_i)|_\Theta a_i = |s_i(j^J(i), c(i))|_\Theta$. Now, we compute $\Pr(\widetilde{\mathbb{X}} = \widetilde{x})$, where $\widetilde{\mathbb{X}}$ is the generated random variable. We write $\Pr(\widetilde{\mathbb{X}} = \widetilde{x}) = \sum_J \Pr\left(\widetilde{\mathbb{X}} \in S(J)\right) \prod_{i=1}^m \Pr(\widetilde{\mathbb{X}}_i = \widetilde{x}_i \mid \widetilde{\mathbb{X}}_k = \widetilde{x}_k \forall k > i \text{ and } \widetilde{\mathbb{X}}_i \in S(J))$. Let $\tilde{p}_i^J = \Pr(\widetilde{\mathbb{X}}_i = 0 | \widetilde{\mathbb{X}}_k = \widetilde{x}_k \forall k > i \text{ and } \widetilde{\mathbb{X}}_i \in S(J))$. At the $(m+1-i)^{\text{th}}$ iteration, the algorithm chooses the value for $\widetilde{\mathbb{X}}_i$. Assume that $\widetilde{\mathbb{X}}_k$ was chosen to be $\widetilde{x}_k$ for $k > i$. Then,

$$\widetilde{p}_i^J = \frac{\tilde{s}^{\geq}((i-1, j^J(i)), c(i))(n_i - a_i)}{\tilde{s}^{\geq}((i-1, j^J(i)), c(i))(n_i - a_i) + \tilde{s}^{\geq}((i-1, j^J(i) - \Omega_{.,i}), c(i) - w_i)a_i}.$$

Since $\tilde{s}^{\geq}((i, j^J(i)), c(i))$ is a $(1+\delta_s)^{i-1}$ approximation of $|s_i(j^J(i), c(i))|_\Theta$:

$$\frac{p_i^J}{(1+\delta_s)^{i-2}} \leq \widetilde{p}_i^J \leq (1+\delta_s)^{i-2} p_i^J \tag{J.2a}$$

$$\frac{1 - p_i^J}{(1+\delta_s)^{i-2}} \leq 1 - \widetilde{p}_i^J \leq (1+\delta_s)^{i-2}(1 - p_i^J), \tag{J.2b}$$

where the left hand side inequality in (J.2a) (respectively (J.2b)) is obtained by realizing that $\widetilde{s}^{\geq}((i-1, j^J(i)), c(i)) \geq |s_{i-1}(j^J(i), c(i))|_\Theta$ and $\widetilde{s}^{\geq}((i-1, j^J(i) - \Omega_{.,i}), c(i) - w_i) \leq (1+\delta_s)^{i-2} |s_{i-1}(j^J(i) - \Omega_{.,i}, c(i) - w_i)|_\Theta$, (respectively $\widetilde{s}^{\geq}((i-1, j^J(i) - \Omega_{.,i}), c(i) - w_i) \geq |s_{i-1}(j^J(i) - \Omega_{.,i}, c(i) - w_i)|_\Theta$ and $\widetilde{s}^{\geq}((i-1, j^J(i)), c(i)) \leq (1+\delta)^{i-2}|s_{i-1}(j^J(i), c(i))|_\Theta$). The right hand side of (J.2a), (J.2b) can be obtained in a similar way. For $\delta_s \in (0,1)$, we have $1/(1+\delta_s)^i > (1-\delta_s)^i$. Thus, $(1-\delta_s)^{i-2} p_i^J \leq \widetilde{p}_i^J \leq (1+\delta_s)^{i-2} p_i^J$. We let $\Pr\left(\widetilde{\mathbb{X}} \in S(J)\right) = \frac{\tilde{s}^{\geq}((m,J),C)}{\sum_{J'} \tilde{s}^{\geq}((m,J'),C)}$, and observe that:

$$\frac{\Pr(\mathbb{X} \in S(J) \mid \mathbb{X} \in S_\Omega)}{(1+\delta)^{m-1}} \leq \Pr\left(\widetilde{\mathbb{X}} \in S(J)\right) \leq (1+\delta)^{m-1} \Pr(\mathbb{X} \in S(J) \mid \mathbb{X} \in S_\Omega).$$

Therefore, each term in the summation on the right hand side of (J.1) is approximated within a relative error of $(1+\delta_s)^\eta$ where $\eta = m(m-1)/2$. It follows that

$$(1-\delta_s)^\eta \Pr_\Theta(\mathbb{X} = \widetilde{x} | \mathbb{X} \in S_\Omega) \leq \Pr(\widetilde{\mathbb{X}} = \widetilde{x}) \leq (1+\delta_s)^\eta \Pr_\Theta(\mathbb{X} = \widetilde{x} | \mathbb{X} \in S_\Omega). \tag{J.3}$$

Now, we obtain a $(1 \pm \epsilon_s)$ approximation if $\delta_s \leq 1 - (1-\epsilon_s)^{1/m^2}$ and $\delta_s \leq (1+\epsilon_s)^{1/m^2} - 1$. Since $(\cdot)^{\frac{1}{m^2}}$ is concave, it follows that $\frac{1}{2}(1+\epsilon_s)^{1/m^2} + \frac{1}{2}(1-\epsilon_s)^{1/m^2} \leq 1$. Therefore, it suffices to choose $\delta_s = (1+\epsilon_s)^{1/m^2} - 1$ in J.3. The desired complexity follows from Theorem 3. □

K Proof of Proposition 4

We assume wlog that $x_{ij} = 0$ for all $\langle i, j\rangle \in E$ and $U = 1$. Given a solution to Slack-MLU (x), we construct a feasible solution to MLU(x). If $\underline{y}$ is a solution to MLU(x) with demand $\underline{d}$ and $\underline{d} = \underline{d}' + \underline{d}''$ where $\underline{d}', \underline{d}'' \geq 0$, then, using augmenting paths, $\underline{y}$ can be decomposed into $\underline{y}', \underline{y}'' \geq 0$, where $\underline{y}'$ services $\underline{d}'$, $\underline{y}''$ services $\underline{d}''$ and y'' does not contain cycles. Now, let (y^a, a), be the given solution to Slack-MLU(x) where, y^a, is a routing of d'. Then, we decompose y^a into y^1 and y^2, where y^1 routes d, y^2 routes a, and y^2 does not contain cycles. Assume wlog that the support of a is a pair (i, j), and, so:

$$\sum_{t\in V} y^1_{klt} + y^2_{klj} \leq c_{kl} + a_{ij}\delta_{(\langle i,j\rangle=\langle k,l\rangle)} \tag{K.1}$$

where $\delta_{(\langle i,j\rangle=\langle k,l\rangle)} = 1$ if $\langle i, j\rangle = \langle k, l\rangle$ and 0 otherwise. Clearly, $a_{ij} \geq y^2_{ijj}$ because y^2 does not contain cycles. We define

$$Z_t = \begin{cases} \frac{y^1_{ijt}}{c_{ij}+a_{ij}-y^2_{ijj}} & \text{if } c_{ij} + a_{ij} - y^2_{ijj} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{K.2}$$

Since $0 \leq \sum_{t\in V} y^1_{ijt} \leq c_{ij} + a_{ij} - y^2_{ijj}$, we get $0 \leq \frac{\sum_{t\in V} y^1_{ijt}}{c_{ij}+a_{ij}-y^2_{ijj}} = \sum_{t\in V} Z_t \leq 1$. We argue that the flow y'', defined as

$$y''_{klt} = y^1_{klt} + Z_t y^2_{klj} - Z_t a_{ij}\delta_{(\langle i,j\rangle=\langle k,l\rangle)} \tag{K.3}$$

is feasible to MLU (x). First, we show feasibility to the capacity constraint.

C.1 Consider $\langle i, j\rangle = \langle k, l\rangle$ and observe that: $\sum_{t\in V} Z_t c_{ij} - \sum_{t\in V} y''_{ijt} = \sum_{t\in V}\left(Z_t c_{ij} + Z_t a_{ij} - Z_t y^2_{ijj} - y^1_{ijt}\right) = 0$, where the first equality is by (K.3). If $c_{ij} + a_{ij} - y^2_{ijj} > 0$, the second equality is from (K.2). Otherwise, it follows from $0 \leq \sum_{t\in V} y^1_{ijt} \leq \sum_{t\in V} Z_t(c_{ij} + a_{ij} - y^2_{ijj}) = 0$. Then, $\sum_{t\in V} y''_{ijt} \leq c_{ij}$ because $0 \leq \sum_{t\in V} Z_t c_{ij} \leq c_{ij}$, where the second inequality holds because $\sum_{t\in V} Z_t \leq 1$.
C.2 Now, consider $\langle k, l\rangle \neq \langle i, j\rangle$. We have $0 \leq \sum_{t\in V} y''_{klt} = \sum_{t\in V}(y^1_{klt} + Z_t y^2_{klj}) \leq \sum_{t\in V} y^1_{klt} + y^2_{klj} \leq c_{kl}$, where, the first equality is from (K.3), the first inequality is because Z_t, y^1, and y^2 are non-negative, the second inequality is because $y^2_{klj} \geq 0$ and $\sum_{t\in V} Z_t \leq 1$, and the last inequality follows from (K.1).

Finally, y'' satisfies flow balance equations in MLU(x) because it is defined in (K.3) by adding a circulation to y^1 which services d. □

L Formulation of Gen-R3

For a directed arc e from i to j, we write tail(e) to represent i and head(e) to represent j. For a node j and commodity t, we write $ex'(r, j, t)$ to represent $\sum_{e\in E:\text{tail}(e)=j} r_{et} - \sum_{e\in E:\text{head}(e)=j} r_{et}$. Then Gen-R3 is: [9]:

$$\text{Gen-R3:} \min_{r,p,a} \quad U \tag{L.1a}$$

$$\sum_{t\in V} r_{et} + \sum_{l\in E} p_{el} x_l \le U c_e(1 - x_e) + a_e x_e \qquad \forall e \in E, \forall x \in \mathscr{X}_{\mathfrak{b}} \tag{L.1b}$$

$$ex'(r, j, t) = d_{jt} - \sum_{i\in V} d_{it}\delta_{j=t} \qquad \forall j, t \in V \tag{L.1c}$$

$$ex'(p, j, l) = a_l \delta_{\text{tail}(l)=i} - a_l \delta_{\text{head}(l)=j} \qquad j \in V, l \in E \tag{L.1d}$$

$$r_{et}, p_{el} \ge 0 \qquad e, l \in E, t \in V. \tag{L.1e}$$

Here, r_{et} is the traffic on link e destined to t and p_{el} is the the amount of traffic on link l that is bypassed on e when l fails, and a_e is the reservation to bypass traffic on link e.

We solve Gen-R3 with $\mathfrak{b} = 1$ in $\mathscr{X}_{\mathfrak{b}}$ *i.e.,*for $\mathscr{X}_1$ in (L.1b). Then using the obtained (r^*, p^*, a^*), the G-cuts are the negation of constraint (L.1b) with U fixed to one *i.e.,*

$$\sum_{t\in V} r^*_{et} + \sum_{l\in E} p^*_{el} x_l > c_e(1 - x_e) + a^*_e x_e \text{ for } e \in E \text{ and } x \in \mathscr{X}_{\mathfrak{b}}. \tag{L.2}$$

Constraint (L.2) can be used to outer-approximate the set of scenarios in $\mathscr{X}_{\mathfrak{b}}$ where MLU exceeds 1.

References

1. ApS, M.: Mosek modeling cookbook (2020)
2. Bangla, A.K., Ghaffarkhah, A., Preskill, B., Koley, B., Albrecht, C., Danna, E., Jiang, J., Zhao, X.: Capacity planning for the google backbone network. https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45385.pdf (2015)
3. Bao, X., Sahinidis, N.V., Tawarmalani, M.: Multiterm polyhedral relaxations for nonconvex, quadratically constrained quadratic programs. Optim. Methods Softw. **24**(4–5), 485–504 (2009)
4. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust Optim., vol. 28. Princeton University Press, Princeton (2009)
5. Ben-Tal, A., Nemirovski, A.: Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications, vol. 2. SIAM, Philadelphia (2001)
6. Ben-Tal, A., Nemirovski, A.: On safe tractable approximations of chance-constrained linear matrix inequalities. Math. Oper. Res. **34**(1), 1–25 (2009)
7. Bertsimas, D., Goyal, V.: On the power and limitations of affine policies in two-stage adaptive optimization. Math. Program. **134**(2), 491–531 (2012)

8. Bertsimas, D., Popescu, I.: Optimal inequalities in probability theory: a convex optimization approach. SIAM J. Optim. **15**(3), 780–804 (2005)
9. Chang, Y., Jiang, C., Chandra, A., Rao, S., Tawarmalani, M.: Lancet: better network resilience by designing for pruned failure sets. Proc. ACM Meas. Anal. Comput. Syst. **3**(3), 1–26 (2019)
10. Chang, Y., Rao, S., Tawarmalani, M.: Robust validation of network designs under uncertain demands and failures. In: Proceedings of the 14th USENIX Conference on Networked Systems Design and Implementation, NSDI'17, pp. 347–362. USENIX Association, USA (2017)
11. Doerr, B.: Probabilistic tools for the analysis of randomized optimization heuristics. In: Theory of Evolutionary Computation, pp. 1–87. Springer (2020)
12. Dworkin, L., Kearns, M., Xia, L.: Efficient inference for complex queries on complex distributions. In: Kaski, S., Corander, J. (eds.) Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, vol. 33, pp. 211–219. PMLR, Reykjavik (2014)
13. Gawrychowski, P., Markin, L., Weimann, O.: A faster FPTAS for #knapsack. In: ICALP, LIPIcs, vol. 107, pp. 64:1–64:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2018)
14. Ghanem, R., Higdon, D., Owhadi, H.: Handbook of Uncertainty Quantification, vol. 6. Springer, Berlin (2017)
15. Gill, P., Jain, N., Nagappan, N.: Understanding network failures in data centers: measurement, analysis, and implications. In: Proceedings of the ACM SIGCOMM 2011 Conference, pp. 350–361 (2011)
16. Gopalan, P., Klivans, A., Meka, R., Stefankovic, D., Vempala, S., Vigoda, E.: An FPTAS for #knapsack and related counting problems. In: Proceedings of the 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS'11, pp. 817–826. IEEE Computer Society, USA (2011)
17. Gurobi Optimization Incorporated: Gurobi optimizer reference manual. http://www.gurobi.com (2019)
18. Han, S., Tao, M., Topcu, U., Owhadi, H., Murray, R.M.: Convex optimal uncertainty quantification. SIAM J. Optim. **25**(3), 1368–1387 (2015)
19. Hanasusanto, G.A., Kuhn, D.: Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. Oper. Res. **66**(3), 849–869 (2018)
20. Karp, R.M., Luby, M., Madras, N.: Monte-Carlo approximation algorithms for enumeration problems. J. Algorithms **10**(3), 429–448 (1989)
21. Knight, S., Nguyen, H., Falkner, N., Bowden, R., Roughan, M.: The internet topology zoo. IEEE J. Select. Areas Commun. **29**(9), 1765–1775 (2011)
22. Lasserre, J.B.: A semidefinite programming approach to the generalized problem of moments. Math. Program. **112**(1), 65–92 (2008)
23. Lasserre, J.B.: The moment-sos hierarchy. In: Proceedings of the International Congress of Mathematicians (ICM 2018) (2019)
24. Laurent, M.: Sums of squares, moment matrices and optimization over polynomials, The IMA Volumes in Mathematics and its Applications Series, vol. 149, pp. 155–270. Springer, Germany (2009)
25. Liu, H.H., Kandula, S., Mahajan, R., Zhang, M., Gelernter, D.: Traffic engineering with forward fault correction. In: Proceedings of the 2014 ACM Conference on SIGCOMM, pp. 527–538 (2014)
26. Luedtke, J., Ahmed, S.: A sample approximation approach for optimization with probabilistic constraints. SIAM J. Optim. **19**(2), 674–699 (2008)
27. Markopoulou, A., Iannaccone, G., Bhattacharyya, S., Chuah, C.N., Ganjali, Y., Diot, C.: Characterization of failures in an operational IP backbone network. IEEE/ACM Trans. Netw. **16**(4), 749–762 (2008)
28. Matthews, L.R., Gounaris, C.E., Kevrekidis, I.G.: Designing networks with resiliency to edge failures using two-stage robust optimization. Eur. J. Oper. Res. **279**(3), 704–720 (2019)
29. Mihalák, M., Šrámek, R., Widmayer, P.: Counting approximately-shortest paths in directed acyclic graphs. In: Kaklamanis, C., Pruhs, K. (eds.) Approximation and Online Algorithms, pp. 156–167. Springer, Cham (2014)
30. Mühlpfordt, T., Roald, L., Hagenmeyer, V., Faulwasser, T., Misra, S.: Chance-constrained ac optimal power flow: a polynomial chaos approach. IEEE Trans. Power Syst. **34**(6), 4806–4816 (2019)
31. Nemirovski, A., Shapiro, A.: Convex approximations of chance constrained programs. SIAM J. Optim. **17**(4), 969–996 (2007)
32. Owhadi, H., Scovel, C., Sullivan, T.J., McKerns, M., Ortiz, M.: Optimal uncertainty quantification. SIAM Rev. **55**(2), 271–345 (2013)
33. Rizzi, R., Tomescu, A.I.: Faster FPTASes for counting and random generation of knapsack solutions. Inf. Comput. **267**, 135–144 (2019)

34. Rockafellar, R.T.: Convex analysis 28. Princeton University Press (1970)
35. Rockafellar, R.T., Uryasev, S., et al.: Optimization of conditional value-at-risk. J. Risk **2**, 21–42 (2000)
36. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis, vol. 317. Springer, Berlin (2009)
37. Sherali, H.D., Adams, W.P.: A Reformulation–Linearization Technique for Solving Discrete and Continuous Nonconvex Problems, vol. 31. Springer, Berlin (2013)
38. Tawarmalani, M., Richard, J.P.P., Xiong, C.: Explicit convex and concave envelopes through polyhedral subdivisions. Math. Program. **138**(1–2), 531–577 (2013)
39. Wang, Y., Wang, H., Mahimkar, A., Alimi, R., Zhang, Y., Qiu, L., Yang, Y.R.: R3: resilient routing reconfiguration. In: Proceedings of the ACM SIGCOMM 2010 Conference, pp. 291–302 (2010)
40. Wets, R.J.B.: Stochastic programs with fixed recourse: the equivalent deterministic program. SIAM Rev. **16**(3), 309–339 (1974)
41. Wood, R.K.: Deterministic network interdiction. Math. Comput. Model. **17**(2), 1–18 (1993)
42. Xu, G., Burer, S.: A copositive approach for two-stage adjustable robust optimization with uncertain right-hand sides. Comput. Optim. Appl. **70**(1), 33–59 (2018)
43. Zhang, Y., Ge, Z., Greenberg, A., Roughan, M.: Network anomography. In: Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement, pp. 30–30 (2005)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.