Safe Approximations of Chance Constraints Using Historical Data*

İhsan Yanıkoğlu
Department of Econometrics and Operations Research, Tilburg University, 5000 LE, Netherlands,
{i.yanikoglu@uvt.nl}

Dick den Hertog
Department of Econometrics and Operations Research, Tilburg University, 5000 LE, Netherlands,
{d.denhertog@uvt.nl}

This paper proposes a new way to construct uncertainty sets for robust optimization. Our approach uses the available historical data for the uncertain parameters and is based on goodness-of-fit statistics. It guarantees that the probability that the uncertain constraint holds is at least the prescribed value. Compared to existing safe approximation methods for chance constraints, our approach directly uses the historical-data information and leads to tighter uncertainty sets and therefore to better objective values. This improvement is significant especially when the number of uncertain parameters is low. Other advantages of our approach are that it can handle joint chance constraints easily, it can deal with uncertain parameters that are dependent, and it can be extended to nonlinear inequalities. Several numerical examples illustrate the validity of our approach.

Key words: robust optimization; chance constraint; phi-divergence; goodness-of-fit statistics
JEL classification: C44, C61, C63

1. Introduction

The objective of robust optimization (RO) is to find solutions that are immune to the uncertainty of the parameters in a mathematical optimization problem. It requires that the constraints of a given problem should be satisfied for all realizations of the uncertain parameters in a so-called uncertainty set. The robust version of a mathematical optimization problem is generally referred to as the robust counterpart (RC) problem. RO is popular because of the tractability of the RC for many classes of uncertainty sets. For example, the RC of an uncertain linear optimization problem with data varying in a polyhedral uncertainty set can be reformulated as a linear optimization (LO) problem [4]. Additionally, the RC of

*This version published in **INFORMS Journal on Computing 25(4): 666–681, 2013**

an uncertain LO problem with an ellipsoidal uncertainty set can be reformulated as a second-order cone problem (SOCP) that can be solved efficiently by existing solvers. The choice of the uncertainty set is important for two reasons. First, it plays a critical role in the tractability of the RC problem. Second, it should represent the actual uncertainty set in a meaningful way.

One way to define uncertainty sets is by the safe approximation of the chance constraint [18, 5, 4]. Tractable safe approximations of chance constrained programs have recently been proposed since the original stochastic counterparts introduced by Charnes et al. [8, 9], Miller and Wagner [16] and Prèkopa [20] are computationally intractable. The seminal work of Shapiro and Nemirovski [18] is based on building a computationally tractable approximation of a chance constrained problem. The authors assume that the constraints are affine and entries of the perturbation vector, so-called uncertain parameters, are independent with known support. Later, they extend their approximation approach to the ambiguous chance constraint, where random perturbations belong to a collection of distributions in a given convex compact set. Ben-Tal et al. [4] (pp 27-60) propose safe convex approximations of scalar chance constraints. The authors translate the existing stochastic uncertainties to "uncertain-but-bounded" sets assuming that the uncertain parameters are mutually independent with zero mean. The obtained approximations in [18, 4] are computationally tractable and perform good when the number of uncertain parameters is relatively high. In addition, Ben-Tal and Nemirovski [5] elaborate a safe tractable approximation of the chance constrained version of an affinely perturbed linear matrix inequality (LMI) constraint, assuming that the primitive uncertain parameters are independent with light-tail distributions (e.g., bounded or Gaussian). More generally, Calafiore and Campi [6] consider a 'randomized' approach such that the resulting randomized solution for an uncertain constraints fails to satisfy the constraint for a small proportion of the perturbation sample, provided that a sufficient number of samples is drawn.

The tractability is even more scarce for joint chance constraints, i.e., when we have a group of randomly perturbed constraints rather than a single one. Prèkopa [20] shows that joint chance constraints are convex only when the right-hand side coefficients are uncertain and follow a log-concave distribution. A commonly followed approach to simplify a joint chance constraint is to apply a Bonferroni inequality, but it is known that this approach is over-conservative. Chen et al. [10] propose an alternative conservative approximation of a joint chance constraints in terms of a worst-case conditional value-at-risk (CVaR) constraint.

The resulting approximation outperforms the Bonferroni approximation. Zymler et al. [23] develop new tools and models for approximating joint chance constraints under the assumption that first- and second-order moments together with the support of the perturbation vector are known. The authors propose an efficient sequential semidefinite programming (SDP) algorithm to solve distributionally robust chance constraint program.

In this paper, we propose an alternative safe approximation of joint chance constraints that does not require the assumption that certain moments are known. Moreover, it uses the full historical-data information and is based on goodness-of-fit statistics known in the statistics literature as ϕ-divergence. The new approach is appropriate when the number of uncertain parameters is low. Numerical results show that it leads to tighter uncertainty sets compared to the existing safe approximation methods, and therefore yields better objective values for the uncertain problem under consideration. The new approach is suitable for dependent and independent uncertain parameters, and can be extended to nonlinear inequalities. The disadvantage of our approach is that it requires extensive data when the number of uncertain parameters is high.

Research that is related to ϕ-divergence includes the following. Klabjan et al. [15] and Calafiore [7] use two special cases of ϕ-divergence to construct uncertainty regions from historical data. The former derives the robust stochastic lot-sizing problem and uses χ^2-statistics; the latter formulates the robust portfolio selection problem and considers Kullback-Leibler divergence. In both papers, the uncertain parameters are probability vectors, and the goal is to find robust solutions that are feasible for all allowable distributions of the uncertain parameters with bounded support. Ben-Tal et al. [3] take up the topic under the more general title of ϕ-divergence and focus on robust optimization problems with uncertainty regions defined by ϕ-divergence distance measures. They provide tractable formulations of robust optimization problems for ϕ-divergence-based uncertainty regions. Their results show that uncertainty sets based on ϕ-divergence are good alternatives for the uncertainty sets such as ellipsoidal, box, and their variations that are well studied in the literature. In this paper, we go one step further and use ϕ-divergence-based uncertainty sets not only for uncertain probability vectors but also for general uncertain parameters.

The remainder of the paper is organized as follows. In §2, we give an introduction to ϕ-divergence and confidence sets. In §3, we discuss the new safe approximation method. Then, in §4, we present the results of several numerical experiments. Finally, we provide concluding remarks in §5.

2. Introduction to ϕ-Divergence and Confidence Sets for Probability Vectors

In this section we define ϕ-divergence and some of the properties taken from [3, 19, 14] that are used in later sections.

2.1 Confidence Sets Based on ϕ-Divergence

The ϕ-divergence ("distance") between two vectors $p = (p_1, ..., p_m) \geq 0$ and $q = (q_1, ..., q_m) \geq 0$ in $\mathbb{R}^m$ is defined by

$$I_\phi(p, q) := \sum_{i=1}^{m} q_i \phi\left(\frac{p_i}{q_i}\right), \tag{1}$$

where $\phi(t)$ is convex for $t \geq 0$, $\phi(1) = 0$, $\phi(a/0) := a \lim_{t\to\infty} \phi(t)/t$ for $a > 0$, and $\phi(0/0) = 0$.

We consider p in (1) to be the unknown true probability vector of an uncertain parameter $\zeta \in \mathbb{R}^\ell$. Given N historical observations on ζ, the support of ζ is divided into m cells such that the number of observations o_i in cell $i \in \{1, \ldots, m\}$ is at least five:

$$\sum_{i=1}^{m} o_i = N \text{ such that } o_i \geq 5, \forall i \in \{1, \ldots, m\}.$$

Then, the historical data on ζ are translated into frequencies $q = (q_1, ..., q_m)$ such that $e^T q = 1$, where e is the all-one vector and q_i is the observed frequency of cell $i \in \{1, \ldots, m\}$ given by

$$q_i = \frac{o_i}{N}.$$

We construct a confidence set for p using the empirical estimate q and goodness-of-fit. A general goodness of fit test, based on ϕ-divergence, is used. If we assume that ϕ is twice continuously differentiable in the neighborhood of 1 and $\phi''(1) > 0$, then the test statistic

$$\frac{2N}{\phi''(1)} I_\phi(p, q)$$

asymptotically follows a χ^2_{m-1}-distribution with $(m-1)$ degrees of freedom. Using this test statistic, an approximate $(1-\alpha)$-confidence set for p is

$$\left\{p \in \mathbb{R}^m : p \geq 0, p^T e = 1, I_\phi(p, q) \leq \rho\right\}, \tag{2}$$

where

$$\rho := \frac{\phi''(1)}{2N}\chi^2_{m-1,1-\alpha}. \tag{3}$$

Different choices of $\phi(.)$ have been studied in the literature. See [19, 14, 3, 13] for an overview; Table 1 taken from [3] presents the most common choices of $\phi(.)$ together with the conjugate function that is defined as follows:

$$\phi^*(s) := \sup_{t\geq 0}\{st - \phi(t)\}.$$

In this paper, we work with ϕ-divergence distances for which the closed-form conjugates are available; see Table 1.

Table 1: ϕ-Divergence Examples

Divergence	$\phi(t), t > 0$	$I_\phi(p,q)$	$\phi^*(s)$
Kullback-Leibler	$t\log t$	$\sum_i p_i \log\left(\frac{p_i}{q_i}\right)$	e^{s-1}
Burg entropy	$-\log t$	$\sum_i q_i \log\left(\frac{p_i}{q_i}\right)$	$-1-\log(-s), s\leq 0$
χ^2-distance	$\frac{1}{t}(t-1)^2$	$\sum_i \frac{(p_i-q_i)^2}{p_i}$	$2-2\sqrt{1-s}, s\leq 1$
Pearson χ^2-distance	$(t-1)^2$	$\sum_i \frac{(p_i-q_i)^2}{q_i}$	$s+s^2/4, s\geq -2$ $-1, s<-2$
Hellinger distance	$(1-\sqrt{t})^2$	$\sum_i(\sqrt{p_i}-\sqrt{q_i})^2$	$\frac{s}{1-s}, s\leq 1$

2.2 Probability Bound for Subset of Cells

Let $V = \{1,\ldots,m\}$ be the set of cells and $S \subseteq V$, and $C(S)$ be the uncertainty region determined by S. In our approach, we choose S such that $\Pr_\zeta(\zeta \in C(S)) \geq \beta$, where ζ is the primitive uncertain parameter and β is the prescribed probability in a given chance constraint. How we find S will be clarified in §3.2; in this subsection, we determine a probability guarantee for a given S. To do this we calculate the minimal value of $\sum_{i\in S} p_i$

such that p is in the $(1-\alpha)$-confidence set (2):

$$\text{(P)} \quad \gamma(S,\alpha) = \min \sum_{i\in S} p_i \tag{4}$$

$$\text{s.t. } I_\phi(p,q) \le \frac{\phi''(1)}{2N}\chi^2_{m-1,1-\alpha}(=\rho) \tag{5}$$

$$\sum_{i\in V} p_i = 1 \tag{6}$$

$$p \ge 0. \tag{7}$$

Note that (P) is a convex optimization problem in $p \in \mathbb{R}^{|V|}$ since ϕ-divergence functions are convex. Constraints (5) to (7) define a $(1-\alpha)$-confidence set, and the probability that the uncertain parameter is in the region defined by S, is at least $\gamma(S,\alpha)$ with a $(1-\alpha)$ confidence level.

The following theorem shows an alternative way of calculating $\gamma(S,\alpha)$ by using the dual problem of (P).

Theorem 1 *Suppose $\phi(.)$ is convex and $\alpha < 1$, then the optimal objective value of problem (P) is equal to the optimal objective value of the following lagrangian dual (LD) problem:*

$$(LD) \quad \max_{\eta\ge 0,\,\lambda} \left\{ -\eta\rho - \lambda - \eta\left[\phi^*\left(-\frac{\lambda+1}{\eta}\right)\sum_{i\in S} q_i + \phi^*\left(-\frac{\lambda}{\eta}\right)\sum_{i\in V\setminus S} q_i\right]\right\}$$

in which $\phi^*(s) = \sup_{t\ge 0}\{st - \phi(t)\}$.

Proof. See Appendix A.1. □

The dual problem is an optimization problem with two variables (η,λ) and a simple constraint $\eta \ge 0$. Furthermore, the convexity of $\phi^*(\lambda)$ implies that $\eta\phi^*\left(\frac{\lambda}{\eta}\right)$ is jointly convex in λ and η. Hence, (LD) is a convex optimization problem with only two variables that can be solved efficiently.

Independent uncertain parameters. In some cases it may be known that the uncertain parameters ζ_j are independent for $j \in \{1,\ldots,\ell\}$. Let V_j denote the set of cells for the j^{th} uncertain parameter and $m_j = |V_j|$. Since the uncertain parameters are independent, we may have historical data for each uncertain parameter separately, and N_j denotes the sample size of the data for the j^{th} parameter. In addition, the probability that the j^{th} uncertain parameter is in cell $i \in V_j$ is denoted by $p_i^{(j)}$. Similarly, $q_i^{(j)}$ denotes the frequency of cell

$i \in V_j$ for the j^{th} uncertain parameter. An aggregate cell is indexed by $(i_1, i_2 \ldots, i_\ell)$, where $i_j \in V_j$ for all $j \in \{1, \ldots, \ell\}$. Because of the independence, the probability that the uncertain parameters are in cell $(i_1, i_2 \ldots, i_\ell)$ is equivalent to

$$p_{i_1,i_2\ldots,i_\ell} = \prod_{j=1}^{\ell} p_{i_j}^{(j)},$$

and the frequency of cell $(i_1, i_2 \ldots, i_\ell)$ is given by

$$q_{i_1,i_2\ldots,i_\ell} = \prod_{j=1}^{\ell} q_{i_j}^{(j)}.$$

All elements $p_{i_1,i_2\ldots,i_\ell}$ are collected in vector $p \in \mathbb{R}^{m_1 m_2 \ldots m_\ell}$, according to the order of the indices. Similarly, $q_{i_1,i_2\ldots,i_\ell}$ are collected in vector $q \in \mathbb{R}^{m1m2\ldots m_\ell}$. Then, the following mathematical optimization problem is a special case of (P), in the case of ℓ independent parameters:

$$\begin{aligned}
\text{(IP)} \quad \min \quad & \sum_{(i_1,i_2,\ldots,i_\ell)\in S} p_{i_1,i_2\ldots,i_\ell} \\
\text{s.t.} \quad & I_\phi(p,q) \le \frac{\phi''(1)}{2N_1N_2\ldots N_\ell}\chi^2_{(m_1-1)(m_2-1)\ldots(m_\ell-1),1-\alpha} && (8)\\
& \sum_{i\in V_j} p_i^{(j)} = 1 \quad \forall j \in \{1,\ldots,\ell\} && (9)\\
& p_{i_1,i_2\ldots,i_\ell} = \prod_{j=1}^{\ell} p_{i_j}^{(j)} \quad \forall i_j \in V_j, \forall j \in \{1,\ldots,\ell\} && (10)\\
& p_i^{(j)} \ge 0 \quad \forall i \in V_j, \forall j \in \{1,\ldots,\ell\}, && (11)
\end{aligned}$$

where $(m_1-1)(m_2-1)\ldots(m_\ell-1)$ denotes the degrees of freedom when we have ℓ independent parameters. It is easy to see that (IP) has highly nonlinear terms in constraint (10) and is nonconvex. Fortunately, the following theorem relaxes the nonlinear structure of (IP) and provides a lower bound for the objective function.

Theorem 2 *Let $V = V_1 \times V_2 \ldots \times V_\ell$, $m-1 = (m_1-1)(m_2-1)\ldots(m_\ell-1)$, $N = N_1N_2\ldots N_\ell$, and $S \subseteq V$, then (P) is a relaxation of (IP).*

Proof. See Appendix A.2. □

Note that the optimal solution $\hat{p}$ of (P) does not necessarily satisfy (10) for the individual probabilities $\hat{p}_{i_j}^{(j)}$ given by $[\hat{p}_{i_j}^{(j)} = \sum_{k\neq j}^{\ell} \sum_{i_k \in V_k} \hat{p}_{i_1,.,i_j,.,i_\ell}]$, and hence the elements of $\hat{p}$ may not be independent. However, we are looking for a good lower-bound probability for S that

can be computed efficiently. This is why we use (P), or equivalently (LD), that yields a tight probability bound $\gamma(S, \alpha)$ for any given parameter structure that can be dependent or independent. Nevertheless, working with independent uncertain parameters has some advantages compared to the dependent case. First, we obtain tighter $(1-\alpha)$-confidence sets for p. This is because we have fewer degrees of freedom for the same number of cells, so the ρ value gets smaller in (5). Second, the sample size becomes the product of the individual sample sizes; see Theorem 2, and we require fewer data.

3. Safe Approximation Method

In this section, we provide our method to derive safe approximations for chance constrained problems. We first describe the general setup of our approach and then explain the details of each step in our algorithm. Finally, we mention possible extensions of the algorithm to joint chance constraints and nonlinear inequalities.

3.1 General Setup

For the sake of simplicity, we explain our safe approximation method for linear optimization. Later in §3.3.2, it is shown how the method is extended to nonlinear inequalities.

We consider the following chance constrained linear optimization problem:

$$\begin{aligned} \text{(ULO)} \quad & \max \; c^{\mathrm{T}} x \\ & \text{s.t. } \Pr_{\zeta}\{\zeta \in [-1,1]^{\ell} : a_k(\zeta)^{\mathrm{T}} x \leq b_k, \forall k \in \{1,...,K\}\} \geq \beta, \end{aligned} \tag{12}$$

where $x \in \mathbb{R}^n$ is a vector of decision variables, $c \in \mathbb{R}^n$ is a vector of objective coefficients, $b \in \mathbb{R}^K$ is a vector of right-hand side values, β is the given probability bound and $a_k(\zeta) \in \mathbb{R}^n$ is linear in the primitive uncertain parameter $\zeta \in [-1,1]^{\ell}$, i.e.,

$$a_k(\zeta) = a_k^0 + \sum_{j=1}^{\ell} \zeta_j a_k^j \quad \forall k \in \{1,...,K\}, \tag{13}$$

where $a_k^j \in \mathbb{R}^n$, $j \in \{0,\ldots,\ell\}$. We may assume w.l.o.g. that $\zeta \in [-1,1]^{\ell}$, since scaling for different intervals can be done by adjusting all the a_k^j. We may also assume w.lo.g that the right-hand side vector b is certain, since the uncertain right-hand side can easily be reformulated in RO. Moreover, we assume that the number of uncertain parameters, ℓ, is much smaller than n. This is motivated by the fact that in many cases a few primitive sources

of uncertainty affect many other parameters of a given system. For example, engineering design problems [22], portfolio optimizations problems [1, 12], etc., often have only a few primitive uncertain parameters and regression or factor models are used to obtain (13). For the sake of simplicity, we focus below on an individual chance constraint, so subindex k is omitted, but in §3.3.1 we show how the method is extended to joint chance constraints.

Eventually, our objective is to find the tightest uncertainty set $\mathcal{Z}$ such that for any feasible solution $x \in \mathbb{R}^n$ of

$$a(\zeta)^T x \leq b \quad \forall \zeta \in \mathcal{Z} \tag{14}$$

the chance constraint

$$\Pr_\zeta\{\zeta : a\,(\zeta)^T\, x \leq b\} \geq \beta \tag{15}$$

is satisfied. Constraint (14) is called a safe approximation of chance constraint (15). Furthermore, (14) is also the RC of the uncertain constraint with the uncertainty set $\mathcal{Z}$.

To determine $\mathcal{Z}$ and the corresponding probability bound, we first divide the domain of ζ into cells such that in each cell there are sufficient historical data. Then, using these data, we calculate the frequency q_i of each cell $i \in V$. The true probability of a cell is denoted by p_i, and the true probability vector p is in the $(1-\alpha)$-confidence set (2). Furthermore, we denote the uncertainty region of cell $i \in V$ by C_i (e.g., C_i is a cube in a three-dimensional uncertainty space). The uncertainty region for all the cells in $S \subseteq V$ is given by

$$C(S) = \bigcup_{i \in S} C_i.$$

Let $\mathcal{Z}$ be $C(S)$ and $x \in \mathbb{R}^n$ be any feasible solution for the safe approximation (14). Then from §2.2 we have

$$\Pr_\zeta\{\zeta : a\,(\zeta)^T\, x \leq b\} \geq \gamma(S, \alpha) \tag{16}$$

with confidence level $(1-\alpha)$, where $\gamma(S,\alpha)$ is given by (4). The aim is to find a tight S that approximates the uncertainty region by $C(S)$ such that $\gamma(S,\alpha) \geq \beta$, and hence (15) holds with a $(1-\alpha)$ confidence level.

In the following section, we present an algorithm that finds such a tight uncertainty set for a given probability bound β.

3.2 Algorithm

In this section, we present an algorithm that iteratively constructs an uncertainty set $\mathcal{Z}$ that satisfies the probability bound β given in (15). We illustrate our approach using the following toy problem:

$$\max_{\boldsymbol{x}\geq 0} \ \left\{x_1 + x_2 \ : \ \Pr_\zeta \left\{\zeta \in [-1,1]^2 : \zeta_1 x_1 + \zeta_2 x_2 \leq 1\right\} \geq \beta\right\}, \tag{17}$$

where β is the prescribed probability, and $\zeta_1 \in [-1,1]$ and $\zeta_2 \in [-1,1]$ are the primitive uncertain parameters that we have historical data on. Later in this section, we adopt the general notation in §3.1, i.e., also for the toy problem, and the steps of the algorithm are explained in detail below.

Step 0. We scale ζ to $[-1,1]^\ell$, where the uncertain parameter is equivalent to the following vector:

$$a(\zeta) = a^0 + \sum_{j=1}^{\ell} \zeta_j a^j. \tag{18}$$

For the toy problem, ℓ is equivalent to 2, $a^0 = 0$, and $a^j \in \mathbb{R}^2$ equals the unit vector e_j; hence, $a(\zeta) = \zeta$.

Then, we calculate the frequency q_i of each cell $i \in V$ as described in §2.1. Figure 1 shows the historical data on ζ for the toy problem, as well as the cells that include the data. The size of the cells is such that each cell includes "enough" data, i.e., at least five observations according to the rule of thumb.

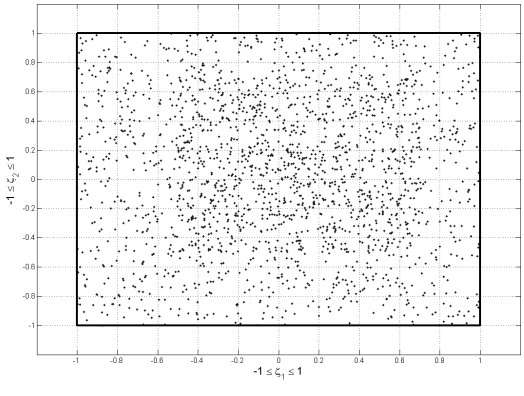

Figure 1: Historical Data for ζ_1 and ζ_2

Remark 1 *Cells with low frequencies can be combined to get "enough" data. Figure 1 presents the standard situation where all cells have the same geometry.*

Step 1. The robust counterpart problem given by

$$\begin{aligned} &\max \ c^{\mathrm{T}}x \\ &\text{s.t. } a(\zeta)^{\mathrm{T}}x \leq b \quad \forall \zeta \in \mathcal{Z} \end{aligned} \tag{19}$$

is solved, where $\mathcal{Z}$ equals the ball-box uncertainty set:

$$BB_{\Omega} := \left\{\zeta \in \mathbb{R}^{\ell} : ||\zeta||_2 \leq \Omega, ||\zeta||_{\infty} \leq 1\right\}. \tag{20}$$

The exact formulation of (19) for $\mathcal{Z}$ equals BB_{Ω}, is equivalent to:

$$z_j + w_j = -[a^j]^T x, \quad \forall j \in \{1, ..., \ell\} \tag{21}$$

$$\sum_{j=1}^{\ell} |z_j| + \Omega \sqrt{\sum_{j=1}^{\ell} w_j^2} \leq b - [a^0]^T x, \tag{22}$$

where z and w are the additional variables. Note that the above formulation can easily be reformulated as an SOCP. In Figure 2, we illustrate the uncertain constraint in the toy problem, when x is fixed to the robust optimal solution x^* and $BB_{0.5}$ is the uncertainty set used in the robust counterpart.

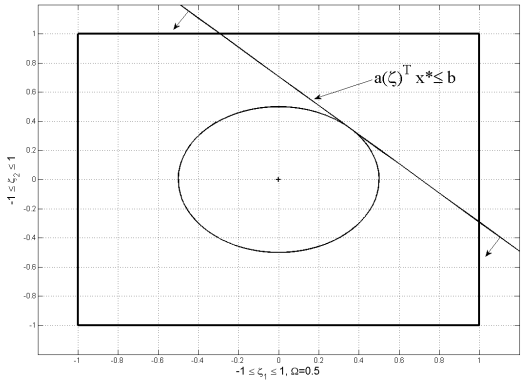

Figure 2: Uncertain Constraint and Ball-Box Uncertainty Set

Remark 2 *Instead of an ellipsoidal uncertainty set, we can also use other uncertainty sets such as the box. In §3.3.3, we discuss that in detail.*

Step 2. We calculate the set of cells

$$S = \left\{ i \in V : a(c^i)^T x^* \leq b \right\}, \tag{23}$$

where $c^i = (c_1^i, c_2^i \ldots, c_\ell^i)$ is the center point of cell $i \in V$. If the center point of a cell satisfies the constraint in (23) for a given x^*, then we assume that all the realizations in the associated cell are feasible for the uncertain constraint. Conversely, if the center point of a cell does not satisfy (23) for a given x^*, then we assume that all the realizations in this cell are infeasible for the uncertain constraint. This assumption is referred as the *center point assumption* in later sections. For the toy problem, the region determined by S is presented in Figure 3.

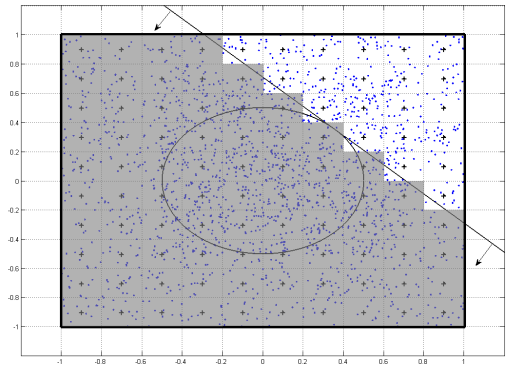

Figure 3: Uncertainty Region $C(S)$

Let I be the intersection of the support, i.e., the box $[-1, 1]^\ell$, and the region determined by the constraint $[a(\zeta)^T x^* \leq b]$. Then, an important observation is that solution x^* is also robust to the uncertainty set I. In addition, the probability that ζ is an element of I is at least the probability that ζ is an element of BB_Ω or equivalently

$$\Pr_\zeta\{\zeta \in I\} \;\geq\; \Pr_\zeta\{\zeta \in BB_\Omega\},$$

since BB_Ω is a subset of I. Hence, using I instead of BB_Ω provides a better probability bound for the optimal solution x^*. To calculate the probability bound, I is approximated by $C(S)$.

Step 3. We calculate $\gamma(S, \alpha)$ as in (4). If $\gamma(S, \alpha) \geq \beta$ then the region determined by I is selected as the uncertainty set and the algorithm is terminated. Otherwise, we go to Step 4.

Step 4. We increase Ω by the step size ω and go to Step 1.

Algorithm 1 (Constraint-wise algorithm)	
Inputs:	LO problem, set of cells V, frequency vector q, step size ω, confidence level $(1-\alpha)$, probability bound β, and $\Omega = 0$.
Outputs:	Uncertainty set $\mathcal{Z}$, robust optimal solution x^*, and radius Ω.
Step 1:	Solve the robust counterpart of the given problem according to the uncertainty set BB_Ω and find the optimal solution x^*.
Step 2:	Calculate $S = \{i \in V : a(c^i)^T x^* \leq b\}$.
Step 3:	Calculate $\gamma(S, \alpha)$ **if** $\gamma(S, \alpha) \geq \beta$ **then** $\mathcal{Z} = \{\zeta \in [-1, 1]^\ell : a(\zeta)^T x^* \leq b\}$ and **terminate** the algorithm **else go to** Step 4.
Step 4:	Set $\Omega = \Omega + \omega$ and **go to** Step 1.

Remark 3 *Notice that $\gamma(S, \alpha)$ is not necessarily increasing in Ω.*

Complexity. In an ℓ-dimensional uncertainty space, Ω can be at most $\sqrt{\ell}$ since BB_Ω is equivalent to the support, $[-1, 1]^\ell$, when Ω is at least $\sqrt{\ell}$. Hence, Ω is changed in at most $O(\omega^{-1}\sqrt{\ell})$ iterations of the algorithm.

3.3 Extensions

3.3.1 Safe Approximation of Joint Chance Constraint

Our approach can also be used to approximate a joint chance constraint:

$$\Pr_\zeta \left\{ \zeta : a_k (\zeta)^T x \leq b_k \ \ \forall k \in \{1, \ldots, K\} \right\} \geq \beta, \tag{24}$$

where $x \in \mathbb{R}^n$ and k denotes the constraint index. The only difference is that we work with multiple constraints rather than a single one. We can use the same algorithm for the joint version by applying the following slight change in Step 2 of Algorithm 1:

$$\textbf{Step } \ \mathbf{2'}: \quad \text{Calculate } S := \{i \in V : a_k(c^i)^T x^* \leq b_k \ \ \forall k \in \{1, ..., K\}\} \ . \tag{25}$$

Note that $C(S)$ now coincides with the region determined by all K constraints and the probability $\gamma(S, \alpha)$ calculated by the algorithm is a joint probability bound satisfied by the approximation of the given joint chance constraint. In Figure 4, we illustrate a $C(S)$ that is determined by multiple constraints including nonlinear ones.

Remark 4 *If we have separate chance constraints rather than a joint one, then in this case the uncertainty set of each constraint must be considered separately. Our approach can also*

be adapted to this case, however we do not consider that in the context of this paper. We see the joint chance constraint as a practically and theoretically more interesting topic to look at.

Remark 5 *Applying the Bonferroni approach to a joint chance constraint, is known to be too pessimistic.*

3.3.2 Extension to Nonlinear Inequalities

Our approach can be extended to nonlinear inequalities. We can focus w.l.o.g. on a single nonlinear constraint, and the robust counterpart of the uncertain constraint with the uncertainty set $\mathcal{Z}$ is given by

$$f(a(\zeta), x) \leq b \quad \forall \zeta \in \mathcal{Z}, \tag{26}$$

where function $f(a(\zeta), x)$ denotes the uncertain nonlinear left-hand side of the constraint. Nonlinearity may be in terms of the decision variables $x \in \mathbb{R}^n$ and/or the uncertain parameters $\zeta \in \mathbb{R}^\ell$. We have no assumption on the decision variables x as long as (26) is tractable; the tractable formulations of such problems are studied in [2], but we assume f is convex in the uncertain parameters ζ for any x. If this assumption holds, then only Step 2 of Algorithm 1 changes slightly as follows:

$$\textbf{Step } \mathbf{2''}: \quad \text{Calculate } S := \{i \in V : f(a(c^i), x^*) \leq b\}, \tag{27}$$

where x^* is the optimal solution of the robust counterpart problem with constraint (26). Note that the algorithm can be extended to joint nonlinear constraints with the following change:

$$\textbf{Step } \mathbf{2'''}: \quad \text{Calculate } S := \{i \in V : f_k(a_k(c^i), x^*) \leq b_k \ \ \forall k \in \{1, ..., K\}\},$$

where k denotes the constraint index. In Figure 4, we illustrate an iteration of the algorithm for a problem that has one linear and two nonlinear constraints in a two-dimensional uncertainty space. Note that the dark region denotes $C(S)$ and the linear constraint is presented by the dashed line.

Remark 6 *In Figure 4, the linear uncertain constraint is not tangent to the ellipsoidal uncertainty set $BB_{0.5}$ for the robust optimal solution. This is because the associated constraint in the RC is not binding at optimality.*

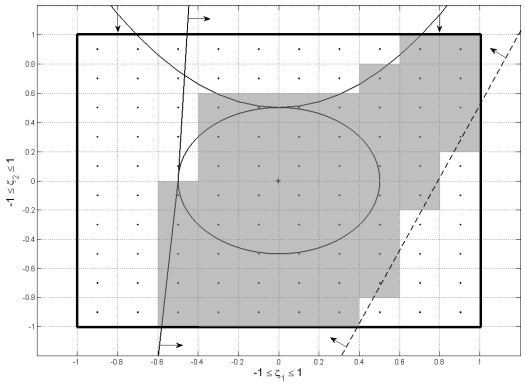

Figure 4: Iteration with Nonlinear Constraints

3.3.3 Extension to Box Uncertainty Set

So far, we have constructed a tight uncertainty set for a given uncertain optimization problem by using ellipsoids. In this subsection, we discuss how we can apply the same method starting from an uncertainty set different than the ellipsoid.

To begin with, we want the tractability of the RC to be as good as that with the ellipsoid. This is why we consider the box uncertainty set as a good choice, and in the sequel of this section we consider

$$\text{Box}_\Omega := \left\{\zeta \in \mathbb{R}^\ell : ||\zeta||_\infty \leq \Omega\right\}$$

as the starting uncertainty set at each iteration of our approach. The RC is no longer an SOCP given by constraints (21) and (22) in Algorithm 1, but it is equivalent to

$$(a^0)^T x + \Omega \sum_{j=1}^{\ell} \left|(a^j)^T x\right| \leq b, \tag{28}$$

that can easily be reformulated as a LO problem. Hence Step 1 of Algorithm 1 changes slightly as follows:

Step 1′ : Solve the RC with constraint (28) for given Ω and find the optimal x^*.

The numerical results in §4 show that using the box or the ellipsoid yields similar results in the safe approximation method. However, we have the impression that, especially for joint constraints, the ellipsoidal uncertainty set has more flexibility than the box in finding the final tight uncertainty set. This is because of the special geometry of the ellipsoid that avoids that the worst case realizations of the uncertain parameters are in the corners of the box.

4. Experimental Results

In this section, we provide the results of the experiments we have conducted for the algorithm presented in §3.2 and its extension in §3.3. All computations are done on a 32-bit Windows machine equipped with a 2.66 GHz Intel Core 2 Quad processor with 3.2 GB of RAM. To solve the mathematical programs, we have used KNITRO 7.0 embedded in MATLAB 2011b. We have conducted four different experiments. In the first experiment, we solve a simple uncertain linear optimization problem with a single constraint. The performance of our algorithm is compared with the approximation of the chance constraint presented in [4]. The optimal objective value is considered as the main performance measure in this experiment. In the second experiment, we apply our algorithm to a multi-period work scheduling (MWS) problem, and the aim is to find the uncertainty region that satisfies a given probability bound for the joint constraints. In this experiment, we also consider the extension presented in §3.3.3. We report the specifications of the uncertainty region found by the algorithm and the optimal objective value of the related robust counterpart problem. In the third experiment, we focus on a robust response model of a cross-docking distribution center in China. The main difference of this experiment from the second one is that the related robust counterpart is a nonlinear optimization problem (NLP). Furthermore, we have also used dependent data in this experiment. Finally, in the last experiment, we apply our approach to another real-life problem originated by the need of a Dutch based electronics company. The related problem, *TV tube problem*, has six uncertain parameters and many uncertain constraints. Numerical results show that our approach provides significant improvements to the nominal case of the associated problem.

4.1 Illustrative Example

Similar to the example in §3.2, we focus on a simple linear uncertain optimization problem with an individual chance constraint. The problem is as follows:

$$\begin{aligned} \text{(M)} \quad \max \quad & x_1 + x_2 \\ \text{s.t.} \quad & \Pr_\zeta\{\zeta \in [-1,1]^2 : (1+\zeta_1)x_1 + (1+\zeta_2)x_2 \leq 10\} \geq \beta \\ & x_1, x_2 \geq 0, \end{aligned} \tag{29}$$

where $\zeta_1 \in [-1,1]$ and $\zeta_2 \in [-1,1]$ are the independent uncertain parameters, and β is the prescribed probability bound. In addition, we have historical data for both of the uncertain

parameters separately, and each data set has a sample size of 100. To obtain the frequencies, we divide the domain of each parameter into ten equal intervals of size 0.2 such that we have enough data points in each interval. The frequencies of the parameters according to the given data are presented in Table 2.

Table 2: Frequencies of ζ_1 and ζ_2

ζ_1, ζ_2	[-1 -0.8]	[-0.8 -0.6]	[-0.6 -0.4]	[-0.4 -0.2]	[-0.2 0]	[0 0.2]	[0.2 0.4]	[0.4 0.6]	[0.6 0.8]	[0.8 1]
freq.(ζ_1)	0.05	0.05	0.1	0.1	0.15	0.15	0.15	0.15	0.05	0.05
freq.(ζ_2)	0.025	0.075	0.2	0.15	0.05	0.125	0.175	0.1	0.075	0.025

The joint uncertainty set of ζ_1 and ζ_2 has 100 (10×10) cells and the frequency of a cell is found by multiplying the frequencies of the associated intervals for ζ_1 and ζ_2. Note that this may be done since ζ_1 and ζ_2 are independent.

The aim of the experiment is to compare the optimal objective values of our safe approximation method to those provided by the safe approximation of the chance constraint (ACC) presented in §2 of [4]. The individual chance constraint (29) is approximated by both approaches for different values of the probability bound β and numerical results are listed in Table 3.

Table 3: Results for Example 4.1

β	$\gamma(S,\alpha)$	Ω	BB_Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.	Ω_{acc}	Obj_{acc}.	%Improv.
Nom.	0.5	0	0	45	10	-	-	-
0.6	0.6	0.15	0.03	36	9.04	1.35	5.11	80.8
0.7	0.7	0.29	0.15	28	8.29	>1.41	5	65.9
0.8	0.87	0.57	0.35	15	7.12	>1.41	5	42.5
0.9	0.92	0.71	0.64	10	6.65	>1.41	5	33.1
0.91	0.92	0.71	0.64	10	6.65	>1.41	5	33.1
0.92	0.92	0.71	0.64	10	6.65	>1.41	5	33.1
0.93	0.96	0.85	0.72	6	6.24	>1.41	5	24.9
0.94	0.96	0.85	0.72	6	6.24	>1.41	5	24.9
0.95	0.96	0.85	0.72	6	6.24	>1.41	5	24.9
0.96	0.976	0.99	0.89	3	5.88	>1.41	5	17.6
0.97	0.976	0.99	0.89	3	5.88	>1.41	5	17.6
0.98	0.984	1.14	0.95	1	5.53	>1.41	5	10.7
0.99	1	1.28	1	0	5	>1.41	5	0
FRC	1	>1.41*	1	0	5	-	-	-

$* \sqrt{2} \approx 1.41$

In this experiment, we have used Algorithm 1 for the case of independent uncertain

parameters. We use χ^2-distance as the ϕ-divergence function when $\alpha = 0.001$; see subproblem (P) in §2.2. The first column in Table 3 presents the probability bounds β and the second column gives the bound satisfied by the algorithm, where $\gamma(S, \alpha)$ represents the optimal objective value of subproblem (P), or equivalently (LD). The third column presents the radius of the minimal ball in the tight uncertainty region calculated by the algorithm. The fourth column gives the probability bound provided by the algorithm, if we would have used the ball-box as the final uncertainty set. The fifth column gives the number of cells removed from the uncertainty space to obtain S, and the sixth column presents the optimal objective value provided by our algorithm. The seventh column corresponds to the radius of the ball, which is equivalent to $\sqrt{2|\ln(1-\beta)|}$ by ACC [4], and the eighth column lists the associated optimal objective value. Finally, the ninth column gives the percentage of improvement in the optimal objective value of ACC when our algorithm is used, or equivalently $((\text{Obj} - \text{Obj}_{acc})/\text{Obj}_{acc}) \times 100$. ACC yields the same optimal solution when Ω is higher than $\sqrt{2}$ since the ball becomes larger than the box uncertainty set in the two-dimensional space. Hence, the uncertainty set BB_Ω in (20) coincides with $[-1, 1]^2$ that results in the worst-case objective value of 5 for (M).

The first row in Table 3 is the nominal problem. We provide the tightest uncertainty set and the probability bound satisfied by the nominal solution. The last row corresponds to the worst-case solution with respect to the full space of uncertainty (FRC). The results in Table 3 reveal that our approach outperforms ACC with respect to the optimal objective value for the given probability bounds. For instance, when the probability bound is 0.8, the improvement in the objective value is 42.5%. Even for high probability bounds such as 0.97 our algorithm yields a 17.6% improvement in the objective. It is clear that both the improvement in the objective value and the number of cells removed from the initial uncertainty set increase as the probability bound β decreases. Furthermore, if we compare the values in the second and the fourth columns, it is easy to see that the final uncertainty set $\mathcal{Z}$ yields significantly better probability bounds then the starting ball-box uncertainty set, BB_Ω, especially when Ω is low. As a concluding remark, we have also conducted the same experiment when different ϕ-divergence functions such as Hellinger and Kullback-Leibler distances are used in subproblem (P); see numerical results in Appendix B.1.

4.2 Multi-Period Work Scheduling Problem

In this experiment, we solve a modified version of multi-period work scheduling (MWS) problem. MWS is a linear optimization problem used to schedule employees for a multi-period time horizon where the demand changes over time.

Computer Service Store. CLS is a computer service store that requires the following skilled-repair times in the next five months: 3000, 3500, 4000, 4500, and 5500. The repair work is done by skilled technicians and these technicians can each work up to 160 hours per month. Furthermore, the technicians may train apprentices to meet future demand. It takes an average of 50 hours to train an apprentice, and new technicians start serving CLS in the month following their training session. In addition, the training sessions have 100% efficiency, so an apprentice always becomes a technician at the end of the training period. The hiring of technicians is done only in the first period and the start-up cost of hiring a technician is \$8000. In addition, each technician is paid \$2000 and each apprentice costs \$1000 per month. On the other hand, 5% of the technicians quit at the end of each month. Finally, the objective of CLS is to minimize the total labor cost incurred to meet the demand in the next five months. The mathematical model of this problem is presented below:

$$
\begin{aligned}
\text{(NMWS)} \quad \min \ & \sum_{i=1}^{5} 1000x_i + \sum_{i=1}^{5} 2000y_i + 8000y_1 & & \\
\text{s.t.} \ & 160y_i - 50x_i \geq d_i & i \in \{1,\ldots,5\} \quad & (30)\\
& 0.95y_i + x_i = y_{i+1} & i \in \{1,\ldots,4\} \quad & (31)\\
& x_i, y_i \geq 0 & i \in \{1,\ldots,5\}, \quad & (32)
\end{aligned}
$$

where y_i represents the number of technicians, x_i corresponds to the number of apprentices in training, and d_i is the repair time demanded in period $i \in \{1,\ldots,5\}$. In practice, the average working and training hours usually deviate from the estimated values because of overtime, illness, vacations, and other factors. We have historical data for 120 months giving the average working and training hours spent per technician each month. These data are used to derive the frequencies in Table 4.

Note that the working hours range from 120 to 200, so the mean is 160 and the half-length of the data range is 40. Similarly, for the training hours the mean is 50 and the half-length of the data range is 20. Using this information from the historical data, we

Table 4: Frequencies for Working (W.H.) and Training (T.H.) Hours

ζ_1, ζ_2	[-1 -0.8]	[-0.8 -0.6]	[-0.6 -0.4]	[-0.4 -0.2]	[-0.2 0]	[0 0.2]	[0.2 0.4]	[0.4 0.6]	[0.6 0.8]	[0.8 1]
W.H.	[120 128]	[128 136]	[136 144]	[144 152]	[152 160]	[160 168]	[168 176]	[176 184]	[184 192]	[192 200]
freq.(ζ_1)	0.02	0.04	0.1	0.1	0.2	0.3	0.1	0.1	0.02	0.02
T.H.	[30 34]	[34 38]	[38 42]	[42 46]	[46 50]	[50 54]	[54 58]	[58 62]	[62 66]	[66 70]
freq.(ζ_2)	0.015	0.07	0.1	0.15	0.15	0.17	0.15	0.11	0.07	0.015

introduce uncertainty to constraint (30) as follows:

$$\begin{aligned} \min \quad & \sum_{i=1}^{5} 1000x_i + \sum_{i=1}^{5} 2000y_i + 8000y_1 \\ \text{s.t.} \quad & \Pr_{\zeta}\{\zeta \in [-1,1]^2 : (160 + 40\zeta_1)\, y_i - (50 + 20\zeta_2)\, x_i \geq d_i, \forall i \in \{1,\ldots,5\}\} \geq \beta \quad (33) \\ & (31), (32), \end{aligned}$$

where $\zeta_1 \in [-1,1]$ and $\zeta_2 \in [-1,1]$ are the uncertain parameters, and β is the prescribed probability bound. The frequencies of the working and training hours are scaled into the frequencies of ζ_1 and ζ_2 in Table 4. Furthermore, the uncertain parameters are independent; therefore, the joint frequencies can be derived similarly to the first experiment. The joint uncertainty region is again divided into 100 (10×10) cells. Eventually, using our safe approximation method, we find the tightest uncertainty set $\mathcal{Z}$ such that for any feasible solution (x, y) of the RC:

$$(160 + 40\zeta_1)\, y_i - (50 + 20\zeta_2)\, x_i \geq d_i, \forall i \in \{1,\ldots,5\}, \forall \zeta \in \mathcal{Z}, \quad (34)$$

the joint chance constraint (33) is satisfied for the given probability bound β. In this experiment we use the extension in §3.3.1, i.e., the safe approximation of the joint chance constraint. The results are reported in Table 5.

The meanings of the columns in Table 5 are the same as for the first experiment. Note that the optimal objective values for the nominal problem (NMWS) and the robust counterpart for the full space of uncertainty (FRC) are 448105 and 621356, respectively (see the first and last row of Table 5). The results show that when the probability bound is as low as 0.6, the optimal objective value calculated by the algorithm is 3% higher than that of the nominal solution. Moreover, with respect to the nominal solution, we see a 14% increase in the immunity to uncertainty in constraint (33), which is a considerable improvement for a 3% sacrifice in terms of the objective value. For the higher probability bounds of 0.92 and 0.94, the improvement in the optimal objective value of FRC is 12% and 11%, respectively.

Table 5: Results for CLS Example

β	$\gamma(S, \alpha^*)$	Ω	BB_Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom.	0.49	0	0	50	448105
0.6	0.63	0.12	0^{**}	44	462691
0.65	0.66	0.14	0^{**}	43	465214
0.7	0.75	0.3	0.12	35	486434
0.75	0.77	0.32	0.12	34	489222
0.8	0.8	0.38	0.33	31	497782
0.85	0.86	0.5	0.38	25	515830
0.9	0.9	0.56	0.55	22	525351
0.91	0.91	0.58	0.55	21	528603
0.92	0.94	0.7	0.66	15	548988
0.93	0.94	0.7	0.66	15	548987
0.94	0.94	0.72	0.8	14	552538
0.95	0.95	0.74	0.8	13	556134
0.96	0.96	0.78	0.85	11	563468
0.97	0.98	0.9	0.9	5	586672
0.98	0.98	0.94	0.93	3	594834
0.99	0.99	0.98	0.95	1	603225
FRC	1	$>1.41^{***}$	1	0	621356

$* \ \alpha = 0.001$, $** \ 1.9\times10^{-8} \approx 0$, $*** \ \sqrt{2} \approx 1.41$

Furthermore, for the probability bound of 0.98, the algorithm improves the objective value of FRC by 4% and the solution is robust to at least 98.3% of the uncertainty. It is clear that when the probability bound β increases, we remove fewer cells from the initial uncertainty region and the radius Ω of BB_Ω gets larger. Ultimately, the decision maker must make the decision by looking at the results in Table 5 and choosing the best option for CLS.

Later in this example, we consider the extension in §3.3.3, i.e., using the box instead of the ellipsoid as the starting uncertainty set of our algorithm. Note that we apply this extension to the same problem. In addition, we test the new approach for the same data set and when the inputs of the algorithm such as the number of cells and step size ω are held constant. The numerical results are presented in Table 6. The symbol $(*)$ denotes an instance where using the box yields a better optimal objective value than using the ellipsoid for a given probability bound β.

The numerical results reveal that using the box or the ellipsoid as the starting uncertainty set yields similar optimal objective values, e.g, the highest difference between optimal objective values of two approaches is around 1%. Nevertheless, using the ellipsoid is more

Table 6: Results for CLS Example (Box)

β	$\gamma(S, \alpha^*)$	Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom.	0.49	0	50	448105
0.6	0.63	0.1	44	461039(∗)
0.65	0.66	0.12	43	463715(∗)
0.7	0.78	0.3	33	489236
0.75	0.78	0.3	33	489236
0.8	0.80	0.34	31	495286(∗)
0.85	0.90	0.5	22	521022
0.9	0.90	0.5	22	521022
0.91	0.91	0.52	21	524423(∗)
0.92	0.96	0.7	11	557107
0.93	0.96	0.7	11	557107
0.94	0.96	0.7	11	557107
0.95	0.96	0.7	11	557107
0.96	0.96	0.7	11	557107(∗)
0.97	1	0.9	0	598402
0.98	1	0.9	0	598402
0.99	1	0.9	0	598402(∗)
FRC	1	>1.41**	0	621356

$* \ \alpha = 0.001$, $** \ \sqrt{2} \approx 1.41$

flexible in finding the final tight uncertainty sets for the CLS problem. For instance, if the probability bound is in between 0.92 and 0.96 or higher than 0.96, then the safe approximation method using the box finds only one uncertainty set for each of the cases; whereas, the results in Table 5 show that the safe approximation method using the ellipsoid finds a unique tight uncertainty set for each of the probability bounds (except 0.92 and 0.93).

4.3 Optimization of Cross-Docking Distribution Center

Our method can also be applied to the area of robust optimization via (computer) experiments. For a detailed treatment, see [17]. The problem is to find settings for a number of design variables ($x \in R^n$) such that a given objective is optimized and the performance constraints are met with a prescribed probability. One has to work with probabilities since uncontrollable noise factors ($\zeta \in R^m$) influence the performance. Using (computer) experiments in which both the design variables and the noise factors are varied, response functions (or metamodels), $\hat{y}_i(x, \zeta)$, can be developed. The constraint now becomes

$$\Pr_\zeta\{\zeta \in [-1, 1]^2 : \hat{y}_i(x, \zeta) \leq \theta_i, \ \forall i\} \geq \beta.$$

One commonly followed approach is to replace each constraint by

$$\mathbb{E}_\zeta[\hat{y}_i(x,\zeta)] + \kappa\sqrt{\mathbb{V}_\zeta[\hat{y}_i(x,\zeta)]} \leq \theta_i,$$

where κ is such that $\Pr(X \leq \kappa) \geq \beta$, where X is a standard normally distributed variable. For a recent real-life application see [22].

A disadvantage of this approach is that one has to assume that ζ is normally distributed with a known mean and variance. Second, one has to assume that $\hat{y}_i(x,\zeta)$ is normally distributed, which is probably not the case when $\hat{y}_i(x,\zeta)$ is nonlinear in ζ. Moreover, to guarantee the joint probability constraint, one has to apply Bonferroni to get probability bounds for each constraint separately. The resulting constraints are conservative. Our method offers an alternative way to deal with the uncertain noise factors. It does not require a normality assumption and does not use Bonferroni. Moreover, our method explicitly uses historical observations of ζ. Observe that the number of noise factors in practice is often low (usually up to 5), which is ideal for our approach.

In this example we focus on the robust response model of a cross-docking distribution center (CDDC); see [21]. The associated research is motivated by the desire of a third-party logistics (TPL) company to improve its supply chain management. As background information, TPL distributes units from part suppliers (PSs) to an assembly plant (AP) that manufactures automobiles. There are five decision factors (DFs) and two environmental factors (EFs) affecting the system. The EFs are primitive sources of the uncertainty; they are the quantity variability and the suppliers' production interruption probability. The DFs are the number of receiving doors, shipping doors, forklifts, conveyors, and threshold parts; these factors are under the control of the users. Note that the DFs are denoted by the coded variables $x_i \in [-1,1]$, $i \in \{1,\ldots,5\}$; the EFs are denoted by $\zeta_j \in [-1,1]$ where $j \in \{1,2\}$.

Because of an estimated demand growth rate of 10% to 15%, a new AP will be established. When the two APs operate simultaneously, the CDDC will not be able to maintain a steady distribution to the APs. Therefore, the CDDC's internal operations must be optimized to satisfy the AP demand under supply uncertainty. Based on simulation results, [21] derives response functions of the performance measures to be used in the mathematical optimization problem. These measures are the dwelling time (DT) in the temporary storage area, the total throughput (TT) of the CDDC, and the quantities that exceed the threshold time (ET) in

the temporary storage area. We focus on the following chance constrained problem:

$$\begin{aligned}
\text{(CTPL)} \quad \max \ & \mathbb{E}_\zeta \left[TT(x,\zeta) \right] \\
\text{s.t.} \ & \Pr_\zeta \{ \zeta \in [-1,1]^2 : \hat{y}_{DT}(x,\zeta) \leq 20, \ \hat{y}_{ET}(x,\zeta) \leq 40000 \} \geq \beta \qquad (35) \\
& -1 \leq x_i \leq 1 \ \ \forall i \in \{1,\ldots,5\}, \qquad (36)
\end{aligned}$$

where $\hat{y}_{DT}$ and $\hat{y}_{ET}$ are the response functions of DT and ET, respectively, and $\mathcal{Z}$ is the uncertainty set. The response functions are polynomials in x_i but linear in terms of the uncertain parameters ζ_i. For complete formulas of the response functions see Appendix B.2.1. The objective of (CTPL) is to minimize the expected TT denoted by $\mathbb{E}_\zeta[TT(x,\zeta)]$, and (CTPL) is an NLP since the response functions are nonlinear in x. We apply our safe approximation method to find an uncertainty set $\mathcal{Z}$ such that for any feasible solution $x \in \mathbb{R}^5$ of the RC:

$$\hat{y}_{DT}(x,\zeta) \leq 20, \ \ \forall \zeta \in Z \qquad (37)$$

$$\hat{y}_{ET}(x,\zeta) \leq 40000, \ \ \forall \zeta \in Z, \qquad (38)$$

the joint chance constraint (35) is satisfied for the given probability bound β.

Similarly to the earlier experiments, the uncertainty space is divided into 100 (10×10) cells. Furthermore, the uncertain parameters ζ_1 and ζ_2 are assumed to be independent and normally distributed in [21]. These assumptions are not essential for our approach, but we have used them for the sake of comparison. Thus, random data for ζ_1 and ζ_2 are obtained from $N(20,5)$ and $N(0.02,0.01)$ with a sample size of 1000, respectively, and scaled to the interval $[-1,1]$. Table 7 presents the results of the experiment.

The optimal objective value of the nominal problem is 496597. Moreover, the probability bound satisfied by this solution is 0.49. In other words, the joint uncertain constraint will not be satisfied with 51% probability, when x is fixed to the nominal solution in (35).

The target expected total throughput (TT) of the TPL Company is 480000 [21]. Our results in Table 7 show that this target can be satisfied for a probability bound as high as 0.81. In addition, the immunity to 81% of the uncertainty is significantly better than that provided by the nominal solution. Between the nominal solution and the solution satisfying a bound of 0.8, the optimal objective value decreases by 3%, while there is a 32% increase in the immunity to uncertainty. On the other hand, for probability bounds above 0.9, we can no longer satisfy the target. For instance, our optimal solution can not satisfy 5% of 480000, when the prescribed probability is 0.99.

Table 7: Results for CDDC Example

β	$\gamma(S,\alpha^*)$	Ω	BB_Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom.	0.49	0	0	50	496597
0.6	0.62	0.1	0^{**}	44	491096
0.7	0.72	0.18	0.21	40	486534
0.8	0.81	0.27	0.21	36	481507
0.9	0.9	0.44	0.61	31	472870
0.91	0.92	0.46	0.61	29	471925
0.92	0.92	0.46	0.61	29	471925
0.93	0.94	0.51	0.74	23	469386
0.94	0.94	0.51	0.74	23	469386
0.95	0.95	0.56	0.74	20	467129
0.96	0.96	0.62	0.83	18	464540
0.97	0.97	0.66	0.83	15	462884
0.975	0.98	0.71	0.9	12	460890
0.98	0.98	0.76	0.91	10	458977
0.99	0.99	0.86	0.94	6	455381
FRC	1	$>1.41^{***}$	1	0	445172

$*\ \alpha = 0.001$, $**\ 2 \times 10^{-8} \approx 0$, $***\ \sqrt{2} \approx 1.41$

The trade-off between the probability guarantee and the optimal objective value is clear in the reported results. Using the solutions in Table 7, the decision maker can select the best strategy for the new distribution system. This could involve accepting a small reduction from the expected target for the sake of a higher probability guarantee, or satisfying the target with a lower guarantee.

Dependent Data. Later in this example, we use the dependent data that is presented in Table 13; see Appendix B.2.2. The data is obtained using a bivariate normal distribution by post-processing the "tail" cells that have less observations. The values in Table 13 correspond to the number of observations in the associated cells and the sample size is 3033, hence the frequency of a cell can be calculated by dividing the number of observations in the associated cell to the sample size. The total number of cells is again 100.

According to the given data, we apply our safe approximation method to the CDDC problem and the numerical results are reported in Table 8. The uncertainty sets that are reported in Table 8 are larger than the ones provided in Table 7. This is because of three reasons: First is the data structure, e.g., extensive data locate on the corners of the uncertainty region, namely, the top-left and the bottom-right corners in Table 13. Second, the ρ value in constraint (5) increases, since the degrees of freedom increases. Note that the

Table 8: Results for Dependent Data

β	$\gamma(S,\alpha^*)$	Ω	BB_Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom.	0.36	0	0	50	496597
0.6	0.62	0.24	0.06	37	483161
0.7	0.71	0.45	0.26	30	472397
0.8	0.81	0.51	0.36	23	469387
0.9	0.90	0.71	0.52	12	460890
0.91	0.91	0.75	0.53	11	459354
0.92	0.93	0.81	0.57	8	457143
0.93	0.94	0.86	0.57	6	455382
0.94	0.95	0.99	0.74	3	451126
0.95	0.95	1.04	0.8	1	449603
0.96	1	1.09	0.81	0	448139
0.97	1	1.09	0.81	0	448139
0.98	1	1.09	0.81	0	448139
0.99	1	1.09	0.81	0	448139
FRC	1	$>1.41^{**}$	1	0	445172

$* \ \alpha = 0.001$, $** \ \sqrt{2} \approx 1.41$

degrees of freedom is 99 for the dependent case; whereas it is 81 for the independent case. Third, the sample size of the dependent data is smaller than that of the independent data. As a result, to satisfy the same probability guarantees we require larger uncertainty sets. Note that a larger uncertainty sets implies a conservative RC and this is why the optimal objective values in Table 8 are lower than the ones in Table 7. Nevertheless, we still have significant improvements to the nominal solution. For instance, the solution satisfying a bound of 0.6 has 26% higher immunity to uncertainty than that of the nominal solution and it is a considerable improvement for a 2.7% loss in the optimal objective value. For probability bounds that are higher than 0.95, the safe approximation method finds the same tight uncertainty set yielding the probability bound of one (using the discretization and the *center point assumption* of the safe approximation method). Furthermore, the optimal objective value of the RC with $BB_{1.09}$ is 0.6% higher than the worst-case optimal 445172 given by FRC.

To conclude, it is clear that using the safe approximation method yields significant improvements to the immunity to uncertainty, provided by the nominal solution, for relatively small losses in terms of the optimal objective value.

4.4 Optimizing Color Picture Tube

In the manufacturing process of a standard television, the *color picture tube* is assembled to the other components using a manufacturing oven. The oven temperature causes thermal stresses on different points of the tube and if the temperature is too high, it will scrap the tube due to implosions. Figure 5 taken from [11] gives an example of a temperature profile on a tube.

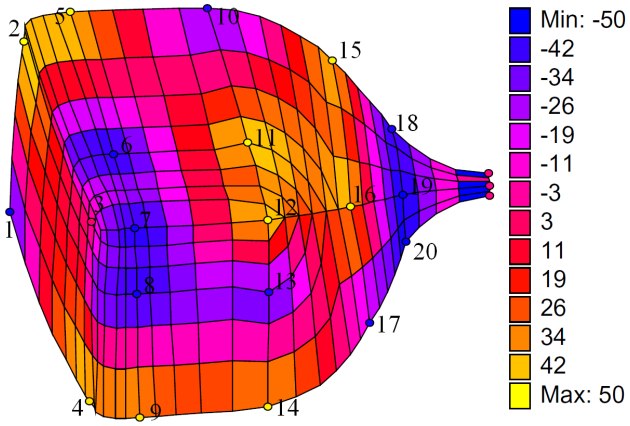

Figure 5: Temperature Profile

To minimize the cost and hence the number of scraps, the manufacturer would like to make an optimal temperature profile such that the temperatures are in the specified range, the temperature differences between near locations are not too high and the maximal stress for the TV tube is minimal. Den Hertog and Stehouwer [11] formulated the associated problem as follows:

$$
\begin{aligned}
\min \quad & s_{\max} \\
\text{s.t.} \quad & a_k + b_k^T x - s_{\max} \leq 0 && \forall k \in \{1, \ldots, K\} && (39) \\
& -\triangle T_{\max} \leq Ax \leq \triangle T_{\max} && && (40) \\
& l \leq x \leq u, && && (41)
\end{aligned}
$$

where $s_{\max} \in \mathbb{R}$ is the maximal stress, $a_k + b_k^T x \in \mathbb{R}$ is the stress at location k, i.e., linear in x, and $x \in \mathbb{R}^n$ represents the vector of temperatures. The vectors $l \in \mathbb{R}^n$ and $u \in \mathbb{R}^n$ are the lower and upper bounds of the decision variables, respectively. The parameter $\triangle T_{\max} \in \mathbb{R}^d$ represents the maximal allowed temperature on d location combinations. $A \in \mathbb{R}^{d \times n}$ coincides with the coefficients in the linear constraints that enforce the specified temperatures do not

differ more than $\triangle T_{\max}$. There are 20 temperature points on the TV tube and hence $n = 20$; see Figure 5. Furthermore, these temperatures result in 216 thermal stresses on different parts of the tube so $K = 216$. The response functions of the thermal stresses, $a_k + b_k^T x$, are derived by using FEM simulator and regression in [11]. In this example, we use the same response functions, but the decision variable x_i is replaced by $x_i(1 + \zeta_j)$, where ζ_j is the multiplicative uncertain parameter, i.e., commonly referred as the implementation error (e.g., $\zeta_j = 0.2$ means 20% implementation error in x_i).

According to the proximity of the temperature points, we form the following six subgroups:

j	1	2	3	4	5	6
$T(j)$	{1}	{2,5,10}	{3,6,7,8}	{4,9,14}	{11,12,13}	{15,16,17,18,19,20}

$T(j)$ denotes the set of indices of the decision variable(s) that are assumed to be affected by the same uncertain parameter ζ_j. This is a valid assumption since closer points in the TV tube have similar temperatures in practice. Eventually, using the safe approximation method, our objective is to find the tightest uncertainty set $\mathcal{Z}$ for the RC:

$$\begin{aligned} \min \ & s_{\max} \\ \text{s.t.} \ & a_k + \sum_{j=1}^{6} \sum_{i \in T(j)} b_{ik} x_i (1 + \zeta_j) - s_{\max} \leq 0 \quad \forall k \in \{1, \ldots, K\}, \forall \zeta \in \mathcal{Z} \\ & (40), (41), \end{aligned} \tag{42}$$

such that the joint chance constraint:

$$\Pr_{\zeta}\big\{\zeta \in [-1, 1]^6 : a_k + \sum_{j=1}^{6} \sum_{i \in T(j)} b_{ik} x_i (1 + \zeta_j) \leq s_{\max}, \ \forall k \in \{1, \ldots, K\}\big\} \geq \beta \tag{43}$$

is satisfied for any feasible RC solution $(x, s_{\max})$, where β is the given probability bound. The RC problem has 21 decision variables including $s_{\max}$, six primitive uncertain parameters, 216 linear uncertain constraints (i.e., given by constraint (42)) and 56 linear constraints (i.e., given by constraints (40) and (41)).

Data. The data for implementation errors are invented by us and the data range is divided to five equal intervals. The frequencies of the associated intervals are shown in Table 14 and 15; see Appendix B.3, we have the same frequencies in two different data ranges that are: [-.1, .1] and [-.2, .2]. These ranges correspond to 10% and 20% implementation errors, respectively. In addition, we assume the uncertain parameters are independent and

hence the frequency of a cell may be found by multiplying the frequencies of the associated intervals for $(\zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5, \zeta_6)$. The total number of cells in the joint uncertainty space is 15625 (5^6).

The numerical results are shown in Table 9 and Table 10. In both tables, it is easy to see

Table 9: TV Tube Example (10% Imp. Err.)

β	$\gamma(S, \alpha^*)$	Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom	0^{**}	0	15625	14.14
0.3	0.3	0.56	12642	14.4
0.35	0.35	0.6	11967	14.42
0.4	0.41	0.64	11245	14.44
0.45	0.46	0.68	10447	14.45
0.5	0.53	0.72	9547	14.47
0.6	0.63	0.78	8031	14.5
0.7	0.71	0.84	6753	14.52
0.75	0.76	0.88	5938	14.54
0.8	0.82	0.94	4772	14.56
0.85	0.85	0.98	4046	14.58
0.9	0.91	1.06	2833	14.61
0.92	0.93	1.1	2321	14.63
0.95	0.95	1.16	1648	14.65
0.96	0.97	1.2	1300	14.67
0.97	0.97	1.22	1140	14.68
0.98	0.98	1.28	736	14.7
0.99	0.99	1.34	455	14.73
FRC	1	$>2.45^{***}$	0	14.91

$* \ \alpha = 0.001$, $** \ 2{\times}10^{-8} \approx 0$, $*** \ \sqrt{6} \approx 2.45$

that the nominal solution is not immune to the implementation errors. To be more precise, if the decision variables $(x, s_{\max})$ are fixed to the nominal solution in the joint chance constraint (43), then the left-hand side probability is almost zero (i.e., 2×10^{-8}). This means that the ζ values that are feasible for the joint constraint, are realized with almost zero probability. This is why implementing the nominal solution can be a risky decision in practice, but using the safe approximation method we can find significantly better solutions.

Numerical results in Table 9 show: Between the nominal solution and the solution satisfying a bound of 0.3, the optimal objective value increases by 1.8%, while there is a 30% increase in the immunity to uncertainty. In addition, the solution satisfying a bound of 0.85 has an optimal objective value that is 3% higher than that of the nominal solution. These

are significant improvements in the immunity to uncertainty for the losses in the optimal objective value.

Table 10: TV Tube Example (20% Imp. Err.)

β	$\gamma(S,\alpha^*)$	Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.
Nom.	0^{**}	0	15625	14.14
0.3	0.31	0.56	12626	14.64
0.35	0.37	0.6	11857	14.67
0.4	0.43	0.64	11015	14.7
0.45	0.46	0.66	10593	14.72
0.5	0.53	0.7	9680	14.75
0.6	0.63	0.76	8078	14.8
0.7	0.71	0.82	6640	14.85
0.75	0.75	0.86	5906	14.87
0.8	0.82	0.92	4594	14.92
0.85	0.86	0.96	3776	14.95
0.9	0.9	1.02	2827	14.99
0.92	0.92	1.06	2286	15.02
0.95	0.95	1.14	1463	15.07
0.96	0.97	1.18	1114	15.1
0.97	0.97	1.2	950	15.11
0.98	0.98	1.24	676	15.13
0.99	0.99	1.3	372	15.17
FRC	1	$> 2.45^{***}$	0	15.44

$* \ \alpha = 0.001$, $** \ 2{\times}10^{-8} \approx 0$, $*** \ \sqrt{6} \approx 2.45$

Note that when we increase the implementation errors from %10 to %20, then the variance from the nominal case increases and we require larger tight uncertainty sets to satisfy the same probability bounds. This is why the number of cells removed from the full space of uncertainty is fewer when the implementation errors are higher; see the fourth columns of Table 9 and 10. A larger tight uncertainty set implies a more restrictive RC and hence the optimal objective values shown in Table 10 are on average 2.2% higher than those provided in Table 9. The lowest difference between two optimal objectives, i.e., 1.7%, is obtained at 0.3 probability bound, and the highest, i.e., 2.9%, is obtained at 0.99, and there is a gradual increase in between.

5. Conclusions

In this paper, we have proposed new safe approximations for joint chance constraints. Using historical data and goodness-of-fit statistics based on ϕ-divergence, we constructed the uncertainty sets that are used in safe approximations. The numerical results show that our approach yields tighter uncertainty sets, and therefore better objective values than the existing method, for the same probability guarantees, especially when the number of uncertain parameters is low. In addition, we do not impose the assumptions that the uncertain parameters are independent or certain moments are known. Last but not least, the new approach can also handle nonlinear inequalities.

It is important to observe that the computational performance of our approach is highly dependent on the number of uncertain parameters. Furthermore, we may require many data points, especially when the uncertain parameters are dependent and the number of uncertain parameters is high, and this data requirement may be hard to manage in practice. In future research, we will investigate the improvement of our approach in such situations. The extension of our approach to simulation based optimization and nonlinear problems will also be further analyzed in future research.

Appendix

A. Proofs of Theorems

A.1 Proof of Theorem 1

The objective function of (P) can be rewritten as $\sum_{i\in V} a_i p_i$, where

$$a_i = \begin{cases} 1, & i \in S \\ 0, & \text{otherwise.} \end{cases}$$

Then we can derive the Lagrangian function as follows:

$$\begin{aligned} L(p,\eta,\lambda) &= \sum_{i\in V} a_i p_i + \lambda\left(\sum_{i\in V} p_i - 1\right) + \eta\left(I_\phi(p,q) - \rho\right) \\ &= -\eta\rho - \lambda + \sum_{i\in V}\left[(\lambda + a_i)p_i\right] + \eta I_\phi(p,q). \end{aligned}$$

The corresponding Lagrangian objective function is as follows:

$$\begin{aligned} g(\lambda,\eta) &= \min_{p\geq 0} L(p,\eta,\lambda) \\ &= -\eta\rho - \lambda + \min_{p\geq 0}\left\{\sum_{i\in V}\left[(\lambda+a_i)p_i + \eta q_i \phi\left(\frac{p_i}{q_i}\right)\right]\right\} \\ &= -\eta\rho - \lambda + \min_{p\geq 0}\left\{\sum_{i\in V} -\eta q_i\left[-\frac{(\lambda+a_i)}{\eta}\frac{p_i}{q_i} - \phi\left(\frac{p_i}{q_i}\right)\right]\right\}. \end{aligned}$$

In the last term of the above formulation we have used (1). Then the Lagrangian objective is equivalent to the following:

$$\begin{aligned} g(\lambda,\eta) &= -\eta\rho - \lambda - \max_{p\geq 0}\left\{\sum_{i\in V}\eta q_i\left[\left(\frac{-\lambda - a_i}{\eta}\right)\frac{p_i}{q_i} - \phi\left(\frac{p_i}{q_i}\right)\right]\right\} \\ &= -\eta\rho - \lambda - \sum_{i\in V}\eta q_i \max_{p\geq 0}\left\{\left[\left(\frac{-\lambda - a_i}{\eta}\right)\frac{p_i}{q_i} - \phi\left(\frac{p_i}{q_i}\right)\right]\right\} \\ &= -\eta\rho - \lambda - \eta\sum_{i\in V}\left[q_i\phi^*\left(-\frac{\lambda+a_i}{\eta}\right)\right] \\ &= -\eta\rho - \lambda - \eta\left[\sum_{i\in S} q_i\phi^*\left(-\frac{\lambda+1}{\eta}\right) + \sum_{i\in V\setminus S} q_i\phi^*\left(-\frac{\lambda}{\eta}\right)\right], \end{aligned}$$

where

$$\phi^*(s) := \sup_{t\geq 0}\left\{st - \phi(t)\right\}.$$

Finally, the Lagrangian Dual Problem is the maximization problem presented below:

$$\begin{aligned}(\text{LD}) &\max_{\eta\geq 0,\,\lambda}\{g(\lambda,\eta)\}\\ &=\max_{\eta\geq 0,\,\lambda}\left\{-\eta\rho-\lambda-\eta\left[\sum_{i\in S}q_i\phi^*\left(-\frac{\lambda+1}{\eta}\right)+\sum_{i\in V\setminus S}q_i\phi^*\left(-\frac{\lambda}{\eta}\right)\right]\right\}.\end{aligned}$$

□

A.2 Proof of Theorem 2

Let $(\hat{p},\hat{p}^{(1)},\ldots,\hat{p}^{(\ell)})$ be a feasible solution of (IP). If we prove that $\hat{p}\in\mathbb{R}^{m_1m_2\cdots m_\ell}$ of $(\hat{p},\hat{p}^{(1)},\ldots,\hat{p}^{(\ell)})$ are feasible for (P), then we can conclude that (P) is a reduced relaxation of (IP).

To begin with, let $V=V_1\times V_2...\times V_\ell$, $m-1=(m_1-1)(m_2-1)\ldots(m_\ell-1)$ and $N=N_1N_2\ldots N_\ell$. Then, constraint (8) in (IP) coincides with constraint (5) in (P). In addition, constraints (9) and (10) imply that the $\hat{p}$ values sum up to 1. Moreover, from constraints (10) and (11) in (IP), it is easy to verify that $\hat{p}_{i_1,i_2,\ldots,i_\ell}\geq 0$ for all $i=(i_1,\ldots,i_\ell)\in V$. As a result, $\hat{p}$ satisfy all the constraints in (P).

□

B. Data and Additional Results

B.1 Extra Results for Example 4.1

Table 11: Kullback-Leibler Distance

β	$\gamma(S,\alpha^*)$	Ω	BB_Ω	$\lvert V\rvert-\lvert S\rvert$	Obj.	Ω_{acc}	Obj$_{acc.}$	%Improv.
Nom.	0.5	0	0	45	10	-	-	-
0.60	0.69	0.29	0.03	28	8.30	1.35	5.11	62
0.70	0.78	0.43	0.22	21	7.67	>1.41	5	53.4
0.80	0.86	0.57	0.35	15	7.13	>1.41	5	42.5
0.90	0.92	0.71	0.64	10	6.66	>1.41	5	33.2
0.91	0.92	0.71	0.64	10	6.66	>1.41	5	33.2
0.92	0.96	0.85	0.72	6	6.25	>1.41	5	24.9
0.93	0.96	0.85	0.72	6	6.25	>1.41	5	24.9
0.94	0.96	0.85	0.72	6	6.25	>1.41	5	24.9
0.95	0.96	0.85	0.72	6	6.25	>1.41	5	24.9
0.96	0.98	0.99	0.89	3	5.88	>1.41	5	17.6
0.97	0.98	0.99	0.89	3	5.88	>1.41	5	17.6
0.98	0.98	0.99	0.89	3	5.88	>1.41	5	17.6
0.98	0.99	1.14	0.95	1	5.54	>1.41	5	10.7
0.99	1	1.28	0.98	0	5	>1.41	5	0
1	-	-	-	5	-	-	-	

$*$ α=0.001

Table 12: Hellinger Distance

β	$\gamma(S,\alpha^*)$	Ω	BB_Ω	$\lvert V\rvert - \lvert S\rvert$	Obj.	Ω_{acc}	$\text{Obj}_{acc.}$	%Improv.
Nom.	0.5	0	0	45	10	-	-	-
0.60	0.66	0.29	0.03	28	8.30	1.35	5.11	62
0.70	0.75	0.43	0.22	21	7.67	>1.41	5	53.4
0.80	0.83	0.57	0.35	15	7.13	>1.41	5	42.5
0.90	0.94	0.85	0.64	6	6.25	>1.41	5	24.9
0.91	0.94	0.85	0.64	6	6.25	>1.41	5	24.9
0.92	0.94	0.85	0.64	6	6.25	>1.41	5	24.9
0.93	0.94	0.85	0.64	6	6.25	>1.41	5	24.9
0.94	0.96	0.99	0.89	3	5.88	>1.41	5	17.6
0.95	0.96	0.99	0.89	3	5.88	>1.41	5	17.6
0.96	0.96	0.99	0.89	3	5.88	>1.41	5	17.6
0.97	0.98	1.14	0.95	1	5.54	>1.41	5	10.7
0.98	0.98	1.14	0.95	1	5.54	>1.41	5	10.7
0.99	1	1.27	0.98	0	5	>1.41	5	0
1	-	-	-	-	5	-	-	-

$*$ α=0.001

B.2 Example 4.3

B.2.1 Response Functions

$$\begin{aligned}\mathbb{E}_\zeta[TT(x,\zeta)] = &-479700 - 39819.17 * x(1) - 20253.25 * x(2) + 312.12 * x(3) - 7339.86 * x(4) - 339.78 * x(5) - \\ &7895.49 * x(1) * x(2) - 121.06 * x(1) * x(3) - 33.75 * x(1) * x(4) + 21.24 * x(1) * x(5) - \\ &7.36 * x(2) * x(3) + 649.55 * x(2) * x(4) + 1136.31 * x(2) * x(5) + 788.4 * x(3) * x(4) + \\ &407.64 * x(3) * x(5) - 1101.55 * x(4) * x(5) + 34063.49 * x(1)^2 + 17810.89 * x(2)^2 + \\ &108.13 * x(3)^2 + 10333.23 * x(4)^2 - 1107.72 * x(5)^2.\end{aligned}$$

$$\begin{aligned}\hat{y}_{DT}(x,\zeta) = &-8.57 + 1.2 * x(1) + 2.04 * x(2) - 0.17 * x(3) + 0.78 * x(4) + 3.30 * x(5) - 0.44 * x(1) * x(2) + \\ &0.29 * x(1) * x(3) - 0.26 * x(1) * x(4) + 0.33 * x(1) * x(5) + 0.21 * x(2) * x(3) - 0.45 * x(2) * x(4) + \\ &0.55 * x(2) * x(5) - 0.061 * x(3) * x(4) + 0.062 * x(3) * x(5) + 0.35 * x(4) * x(5) - 0.63 * x(1)^2 - \\ &1.27 * x(2)^2 + 0.19 * x(3)^2 - 0.25 * x(4)^2 - 0.11 * x(5)^2 + \\ &\{7.11 + 0.78 * x(1) + 1.63 * x(2) - 0.081 * x(3) + 0.57 * x(4) + 2.72 * x(5)\} * \zeta_1 + \\ &\{3.21 + 0.46 * x(1) + 0.49 * x(2) - 0.073 * x(3) + 0.16 * x(4) + 1.17 * x(5)\} * \zeta_2 \leq 0.\end{aligned}$$

$$\begin{aligned}\hat{y}_{ET}(x,\zeta) = &-7517.8 + 10256.36 * x(1) + 13753.61 * x(2) - 300.42 * x(3) + 4379.24 * x(4) + 52.43 * x(5) + \\ &5415.96 * x(1) * x(2) + 437.38 * x(1) * x(3) + 214.75 * x(1) * x(4) + 597.11 * x(1) * x(5) - \\ &97.79 * x(2) * x(3) - 1618.36 * x(2) * x(4) - 724.67 * x(2) * x(5) - 1639.28 * x(3) * x(4) - \\ &1243.25 * x(3) * x(5) + 1728.59 * x(4) * x(5) - \\ &1118.43 * x(1)^2 - 1072.35 * x(2)^2 + 226.71 * x(3)^2 - 372.2 * x(4)^2 + 148.92 * x(5)^2 + \\ &\{36087.44 + 13066.74 * x(1) + 17605.17 * x(2) - 739.11 * x(3) + 5944.33 * x(4) + 446.33 * x(5)\} * \zeta_1 + \\ &\{-10868 - 3824.22 * x(1) - 5975.83 * x(2) + 209.48 * x(3) - 2506.4 * x(4) - 579.61 * x(5)\} * \zeta_2 \leq 0.\end{aligned}$$

B.2.2 Dependent Data Set

Table 13: Dependent Data used in Example 4.3

		ζ_2									
	Cells	1	2	3	4	5	6	7	8	9	10
ζ_1	1	66	30	27	23	9	5	5	5	5	5
	2	53	35	38	23	8	13	5	5	5	5
	3	42	28	53	35	44	28	17	6	5	5
	4	32	35	44	79	81	49	26	28	9	5
	5	28	23	53	85	86	70	69	31	14	13
	6	15	15	41	46	83	102	83	67	23	20
	7	5	6	12	36	46	75	73	51	38	30
	8	5	8	5	14	30	49	65	43	38	42
	9	5	5	5	7	15	20	25	34	28	56
	10	5	5	5	5	5	12	15	28	15	79

$(*)$ $N = 3033$

B.3 Data Set of Example 4.4

Table 14: Data Set 1 (10% Imp. Err.)

	[-.1 -.06]	[-.06 -.02]	[-.02 .02]	[.02 .06]	[.06 .1]
ζ_1	0.1	0.21	0.29	0.22	0.18
ζ_2	0.09	0.18	0.38	0.23	0.12
ζ_3	0.13	0.23	0.3	0.17	0.17
ζ_4	0.11	0.22	0.31	0.24	0.12
ζ_5	0.09	0.2	0.28	0.23	0.2
ζ_6	0.17	0.22	0.23	0.2	0.18

$(*)$ $N_j = 100 \ \forall j \in \{1, \ldots, 6\}$

Table 15: Data Set 2 (20% Imp. Err.)

	[-.2 -.12]	[-.12 -.04]	[-.04 .04]	[.04 .12]	[.12 .2]
ζ_1	0.1	0.21	0.29	0.22	0.18
ζ_2	0.09	0.18	0.38	0.23	0.12
ζ_3	0.13	0.23	0.3	0.17	0.17
ζ_4	0.11	0.22	0.31	0.24	0.12
ζ_5	0.09	0.2	0.28	0.23	0.2
ζ_6	0.17	0.22	0.23	0.2	0.18

$(*)$ $N_j = 100 \ \forall j \in \{1, \ldots, 6\}$

Acknowledgments

We thank Aharon Ben-Tal, Jack Kleijnen, Bertrand Melenberg and Goos Kant for their valuable discussions and insightful comments during the realization of this paper. We also thank Erwin Stinstra at CQM for sharing the TV tube data with us.

References

[1] C. Bemis, X. Hu, W. Lin, S. Moazeni, L. Wang, T. Wang, and J. Zhang. Robust portfolio optimization using a simple factor model. University of Minnesota. IMA Preprint Series, 2009.

[2] A. Ben-Tal, D. den Hertog, and J.-P. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. Working Paper, 2011.

[3] A. Ben-Tal, D. den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust optimization with uncertainty regions based on ϕ-divergence. CentER Discussion Paper, 2011-061.

[4] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton Press, Princeton, 2009.

[5] A. Ben-Tal and A. Nemirovski. On safe tractable approximations of chance-constrained linear matrix inequalities. *Mathematics of Operations Research*, 34:1–25, 2009.

[6] G. Calafiore and M.C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102:25–46, 2005.

[7] G. C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–57, 2007.

[8] A. Charnes and W. W. Cooper. Chance-constrained programming. *Management Science*, 6(1):73–79, 1959.

[9] A. Charnes, W. W. Cooper, and G. H. Symonds. Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Management Science*, 4(3):pp. 235–263, 1958.

[10] W. Chen, M. Sim, J. Sun, and C.P. Teo. From CVaR to uncertainty set: Implications in joint chance-constrained optimization. *Operations Research*, 58(2):470–485, 2010.

[11] D. den Hertog and P. Stehouwer. Optimizing color picture tubes by high-cost nonlinear programming. *European Journal of Operational Research*, 140(2):197 – 211, 2002.

[12] E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3 – 56, 1993.

[13] A. A. Gushchin. On an extension of the notion of f-divergence. *Theory of Probability and Its Applications*, 52(3):439–455, 2008.

[14] L. Jager and J. A. Weller. Goodness-of-fit tests via phi-divergences. *Annals of Statistics*, 35:2018–2053, 2007.

[15] D. Klabjan, D. Simchi-Levi, and M. Song. Robust stochastic lot-sizing by means of histograms. Working Paper, 2007.

[16] B. L. Miller and H. M. Wagner. Chance constrained programming with joint constraints. *Operations Research*, 13(6):930–945, 1965.

[17] R. H. Myers and D. C. Montgomery. *Response Surface Methodology: Process and Product in Optimization Using Designed Experiments*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1995.

[18] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

[19] L. Pardo. *Statistical Inference Based on Divergence Measures*. Chapman & Hall\CRC, Boca Raton, 2006.

[20] A. Prékopa. Efficient robust optimization of metal forming processes using a sequential metamodel based strategy. In *Proceedings of the Princeton Symposium on Mathematical Programming*, pages 113–138. Princeton University Press, Princeton, 1970.

[21] S. Wen. Design of pre-enhanced cross-docking distribution center under supply uncertainty: RSM robust optimization method. HUST. Working Paper, 2011.

[22] J. H. Wiebenga, G. Klaseboer, and A. H. van den Boogaard. Efficient robust optimization of metal forming processes using a sequential metamodel based strategy. In J. Chung, editor, *8th International Conference and Workshop on Numerical Simulation of 3D Sheet Metal Forming Processes, Numisheet 2011*, pages 978–985. Korean Society for Technology of Plasticity, August 2011.

[23] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*.