Robust Binary Optimization using a Safe Tractable Approximation

Ruken Düzgün* Aurélie Thiele†

November 2014

Abstract

We present a robust optimization approach to 0-1 linear programming with uncertain objective coefficients based on a safe tractable approximation of chance constraints, when only the first two moments and the support of the random parameters is known. We obtain nonlinear problems with only one additional (continuous) variable. The resulting robust optimization problems can be interpreted as a nominal problem with modified coefficients. We compare our approach with Bertsimas and Sim [1]. In numerical experiments, we obtain solutions of similar quality in faster time.

Keywords: Robust optimization, binary problems, chance constraints.

1 Introduction

We consider binary optimization problems with uncertain objective coefficients and investigate the models that arise from enforcing probabilistic constraints on the objective in the context of robust optimization. Robust optimization is a worst-case optimization approach where the worst case is computed over a given uncertainty set. Providing an intuitive interpretation of uncertainty sets has always been of importance to operations researchers: for instance, Bertsimas and Sim [1] connect the choice of a key parameter in their approach – called the budget of uncertainty – with a probability of constraint violation. More recently, Ben-Tal et. al. [2] describe a process where a "safe tractable approximation" of probabilistic constraints leads to a robust optimization problem where the uncertainty set is determined by the chosen approximation and the probability level. Safe tractable approximations, the most famous of which is the Bernstein approximation, are

*Operations Research Scientist, Marriott International, Bethesda MD, USA, rud207@lehigh.edu

†Visiting Associate Professor, M.I.T., Cambridge MA, USA, aurelie@mit.edu Corresponding author

motivated by the fact that incorporating a chance constraint to a problem usually creates significant computational difficulties if the random variables do not obey a jointly Gaussian distribution. In Ben-Tal et. al. [2], the probabilistic constraint is replaced by a more tractable constraint that, when satisfied, guarantees that the original constraint is satisfied too. Our goal in the present paper is to investigate the theoretical and algorithmic insights we gain from this approach in the special case where decision variables are binary, and to compare this approach with that obtained in Bertsimas and Sim [1].

We make the following contributions to the literature.

- We provide robust formulations when we only know the first two moments and the support of the distributions of the uncertain parameters.
- We show that the safe tractable approximation (Bernstein approximation) in our setting can be interpreted as a deterministic problem with modified cost coefficients that only depend on problem data and one extra coefficient.
- We compare our approach in numerical experiments with that in Bertsimas and Sim [1] for the same problem setting but for a different modeling of uncertainty in robust optimization and argue that, while solution quality is comparable, the solution times in our approach are substantially smaller.

2 The Safe Tractable Approximation

We consider the following problem with uncertain objective coefficients.

$$\begin{aligned} \max \quad & \mathbf{c}'\mathbf{x} \\ \text{s.t.} \quad & x \in \mathcal{X} \subseteq \{0,1\}^n, \end{aligned} \tag{1}$$

Because the vector $\mathbf{c}$ is not known precisely, our goal here will be to maximize the greatest parameter A such that:

$$P\left(\mathbf{c}'\mathbf{x} < A\right) \leq \epsilon, \tag{2}$$

for $\epsilon > 0$ given (small). We first assume that the random coefficients are independent, and will relax this assumption in Section 4. When distributions are continuous, A can be interpreted as the ϵ-quantile of $\mathbf{c}'\mathbf{x}$.

We are interested in deriving a deterministic tractable counterpart to our problem when only a limited amount of information is known: the mean, variance and support of each uncertain parameter. Knowledge of the first two moments is a common assumption in distributional robustness (see Gabrel et. al. [3] and the references therein), while knowledge of the support is the foundation of the polyhedral uncertainty sets in Bertsimas and Sim [4].

Lemma 2.1. *If $E[\exp\{-\theta c_i\}]$ can be computed efficiently for all i and any $\theta > 0$, a safe tractable approximation to Problem (1) is:*

$$\begin{aligned} \max_{\theta,\, x} \quad & \frac{\ln \epsilon}{\theta} - \frac{1}{\theta} \sum_{i=1}^{n} \ln E[\exp\{-\theta c_i\}] \cdot x_i \\ s.t. \quad & x \in \mathcal{X} \subseteq \{0,1\}^n, \;\; \theta > 0. \end{aligned} \tag{3}$$

Proof. From Ben-Tal et. al. [2], Eq. (2) can be written as, with $\theta > 0$:

$$P\left(-\theta \sum_{i=1}^{n} c_i x_i > -\theta A\right) = P\left(\exp\left\{-\theta \sum_{i=1}^{n} c_i x_i\right\} > \exp\{-\theta A\}\right) \leq \epsilon.$$

Since the exponential function is nonnegative and nondecreasing, we invoke Markov's Inequality and inject that the coefficients are independent:

$$P\left(\mathbf{c}'\mathbf{x} < A\right) \leq \frac{E[\exp\{-\theta \sum_{i=1}^{n} c_i x_i\}]}{\exp\{-\theta A\}} = \frac{\prod_{i=1}^{n} E[\exp\{-\theta c_i x_i\}]}{\exp\{-\theta A\}}.$$

The safe tractable approximation replaces $P\left(\mathbf{c}'\mathbf{x} < A\right) \leq \epsilon$ with $\frac{\prod_{i=1}^{n} E[\exp\{-\theta c_i x_i\}]}{\exp\{-\theta A\}} \leq \epsilon$, thus guaranteeing that the original probabilistic constraint is satisfied. Taking the logarithm of the left and right-hand sides leads to the greatest possible value of A being $\frac{\ln \epsilon}{\theta} - \frac{1}{\theta} \sum_{i=1}^{n} \ln E[\exp\{-\theta c_i x_i\}]$. We conclude by using that the x_i are binary, so $\ln E[\exp\{-\theta c_i x_i\}] = \ln E[\exp\{-\theta c_i\}] \cdot x_i$ for all i. ☐

However, in our problem setup, the $E[\exp\{-\theta c_i\}]$'s are not known exactly; instead, only the first two moments and the support of the random parameters are known. Therefore, we seek tight upper bounds of the $E[\exp\{-\theta c_i\}]$'s given this information by adapting the linear semi-infinite optimization approach of Bertsimas and Popescu ([5], [6]).

Lemma 2.2. *Assume that the random parameter c, with known mean μ and standard deviation σ, has for support $[\bar{c} - \hat{c}, \bar{c} + \hat{c}]$ with $\bar{c} = \mu$ and let m be a positive number such that $\hat{c} = m\,\sigma$ for*

$m \geq 1$. *Let* π *be the set of such possible distributions. Then,*

$$\max_{f\in\pi} E_f[\exp\{-\theta c\}] = \frac{\exp\{-\theta\bar{c}\}}{m^2+1} \cdot \left(\exp\{\theta\hat{c}\} + m^2 \cdot \exp\left\{-\frac{\theta\hat{c}}{m^2}\right\}\right). \tag{4}$$

Proof. We have

$$E[\exp\{-\theta c\}] = \int_{c^-}^{c^+} \exp\{-\theta c\} f(c)dc$$

The problem $\max_{f\in\pi} E_f[\exp\{-\theta c\}]$ can be formulated as, with μ and $\mu^2+\sigma^2$ the first two moments of the distribution:

$$\begin{array}{ll} \max & \int_{c^-}^{c^+} \exp\{-\theta c\} f(c)dc \\ \text{s.t.} & \int_{c^-}^{c^+} f(c)dc = 1 \\ & \int_{c^-}^{c^+} cf(c)dc = \mu \\ & \int_{c^-}^{c^+} c^2 f(c)dc = \mu^2+\sigma^2 \\ & f(c) \geq 0, \quad \forall c \in [c^-, c^+] \end{array} \tag{5}$$

Isii [7] shows that strong duality holds for Problem (5) and its dual is:

$$\begin{array}{ll} \min & \alpha + \mu\beta + (\mu^2+\sigma^2)\gamma \\ \text{s.t.} & \alpha + c\beta + c^2\gamma \geq \exp\{-\theta c\} \quad \forall c \in [c^-, c^+] \end{array} \tag{6}$$

Problem (6) has three decision variables and the function in the right-hand side is non-increasing. Hence, we look for a feasible solution that will make our bound as tight as possible and the constraint must be binding twice, with the left-hand and right-hand-side expressions being tangent at least once and the other intersect being at an extremity of $[c^-, c^+]$. Therefore, we have two possible two-point distributions to consider, both with mean $\bar{c}$ and standard deviation $\sigma = \hat{c}/m$.

1. Distribution 1 takes value c^- w.p. $\dfrac{1}{m^2+1}$ and value $\bar{c} + \dfrac{\hat{c}}{m^2}$ w.p. $\dfrac{m^2}{m^2+1}$
2. Distribution 2 takes value $\bar{c} - \dfrac{\hat{c}}{m^2}$ w.p. $\dfrac{m^2}{m^2+1}$ and value c^+ w.p. $\dfrac{1}{m^2+1}$.

We then compute $E_f[\exp\{-\theta c\}]$ for both candidate distributions and keep the higher value.

$$\max_{f\in\pi} E_f[\exp\{-\theta c\}] =$$
$$\frac{\exp\{-\theta\bar{c}\}}{m^2+1} \cdot \max\left(\exp\{\theta\hat{c}\} + m^2 \cdot \exp\left\{-\frac{\theta\hat{c}}{m^2}\right\}, \exp\{-\theta\hat{c}\} + m^2 \cdot \exp\left\{\frac{\theta\hat{c}}{m^2}\right\}\right).$$

Studying the function $f(\theta) = e^{\theta\hat{c}} + m^2 \cdot e^{-\frac{\theta\hat{c}}{m^2}} - \left(e^{-\theta\hat{c}} + m^2 \cdot e^{\frac{\theta\hat{c}}{m^2}}\right)$, we have $f(0) = 0$ and $f'(\theta) =$

$2\hat{c}\left[cosh(\theta\hat{c}) - cosh(\theta\hat{c}/m^2)\right] \geq 0$ for all $\theta \geq 0$, where we use that $cosh$ is non-decreasing over $[0, \infty)$ and $m \geq 1$ so $\theta\hat{c} \geq \theta\hat{c}/m^2$. Hence, $f(\theta) \geq f(0) = 0$ for all $\theta \geq 0$ and the maximum between $e^{\theta\hat{c}} + m^2 \cdot e^{-\frac{\theta\hat{c}}{m^2}}$ and $e^{-\theta\hat{c}} + m^2 \cdot e^{\frac{\theta\hat{c}}{m^2}}$ is always achieved at the first term. □

Let us define the function F as:

$$F_\theta(m, \widehat{c}) = \frac{1}{m^2+1} \cdot \left(\exp\{\theta\hat{c}\} + m^2 \cdot \exp\left\{-\frac{\theta\hat{c}}{m^2}\right\}\right), \tag{7}$$

so that: $\max_{f\in\pi} E_f[\exp\{-\theta c\}] = \exp\{-\theta\bar{c}\}\, F_\theta(m, \widehat{c})$. Note that, due to the convexity of the function $\theta \mapsto \exp\{-\theta c\}$, we have $E_f[\exp\{-\theta c\}] \geq \exp\{-\theta\bar{c}\}$ so $F_\theta(m, \widehat{c}) \geq 1$ for all m, $\hat{c}$ and θ.

Theorem 2.3 (Robust Problem with Bernstein Approximation, independent case)**.** *The robust counterpart of Problem (1) is:*

$$\begin{aligned} \max_{\theta\geq 0} \frac{\ln \epsilon}{\theta} + \quad \max_{x} \quad & \sum_{i=1}^{n} \left(\bar{c}_i - \frac{1}{\theta} \ln\, F_\theta(m_i, \widehat{c}_i)\right) x_i \\ s.t. \quad & x \in \mathcal{X} \subseteq \{0,1\}^n. \end{aligned} \tag{8}$$

where F is defined in Eq. (7).

Proof. Follows from injecting Eq. (4) of Lemma 2.2 into Problem (3). □

Comments: (i) The robust model (10) can be interpreted as a deterministic problem with modified cost coefficients with only one extra parameter, θ, and no new constraints. In particular, at θ given, it remains linear. (ii) The modified coefficients are all smaller than or equal to their nominal counterparts because $F_\theta(m, \widehat{c}) \geq 1$ for all m, $\hat{c}$ and θ. .

3 Comparison with the Bertsimas-Sim model

3.1 Setup

Our robust optimization model (10) has the appealing feature of being, at θ given, a deterministic linear problem with modified objective coefficients; however, Bertsimas and Sim [1] provide a different modeling of the uncertainty for the same problem and show that robust binary optimization problems with uncertain cost coefficients can also be solved either as a mixed-integer linear problem with new constraints and variables, or as a series of binary optimization problems with modified cost coefficients. Therefore, it is natural to investigate the relative performance of our approach compared to theirs.

Bertsimas and Sim [1] model the uncertain objective coefficients as uncertain parameters in the range $[\bar{c}_i - \widehat{c}_i, \bar{c}_i + \widehat{c}_i]$ for each i and define $\Gamma \in \{0, \ldots, n\}$, also called *budget of uncertainty*, as a measure of the decision-maker's conservatism. Specifically, Γ is the number of uncertain coefficients that can take their worst-case values simultaneously. The robust counterpart of our maximization problem in the Bertsimas-Sim framework is:

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} \bar{c}_i \, x_i \; - \Gamma z_0 - \sum_{i=1}^{n} z_i \\ \text{s.t.} \quad & \sum_{i=1}^{n} cd_i \, x_i \leq B \\ & \mathbf{x} \in \mathcal{X} \\ & z_i + z_0 \geq \hat{c}_i \, x_i, \qquad \forall i, \\ & z_i, z_0 \geq 0 \qquad \forall i. \end{aligned} \tag{9}$$

We tested both models with 10 sets of randomly generated data using the Uniform distribution. The size of each data set is provided in Table 1.

3.2 Results

We ran the Bertsimas-Sim model (9) for the Γ value provided in Eq. (21) of Bertsimas and Sim [4]. We ran our robust model (10) for values of θ starting from 0.0001 and increasing it by a step size equal to $-\ln \epsilon / (\sum_{i=1}^{n} x_i)$ until the objective function value decreased. (Note that in our approach, we seek a "good" θ to have a good approximation, but we do not necessarily need θ to be optimal. The decision-maker trades off computational speed for optimality.) We formulated the problems in GAMS and used IBM ILOG CPLEX 24.2.3 as our solver. Table 1 shows the number of binary variables in the original problem, the average solution times of each problem and the percent difference in expected revenue between each solution, i.e., $(D\&T - B\&S)/B\&S$. While the expected revenue between solutions changes very slightly, we observe a significant difference between solution times. Problem (10) – which we call the Duzgun & Thiele or D&T model – takes shorter time to solve the problem as the size of the data set increases. This is not surprising since Problem (9) – which we call the B&S model for Bertsimas & Sim – increases the problem size whereas Problem (10) does not. 1000 seconds was set as a resource limit for this problem, which was reached by the B&S model as the problem size increases. D&T model never hits the time limit. Solving Problem (9) as a series of binary problems instead of a single MIP, as presented in Bertsimas and Sim [1], increases the solution time even further and those numbers are omitted here.

	n	**B&S**	**D&T**	**% Difference**
Data Set 1	100	0.125	0.265	0
Data Set 2	100	0.172	0.204	0.93
Data Set 3	100	0.469	0.218	0.82
Data Set 4	500	1.437	0.844	1.20
Data Set 5	500	0.656	0.578	1.10
Data Set 6	1000	2.266	0.719	2.09
Data Set 7	1000	2.625	0.922	1.03
Data Set 8	5000	Time limit	0.953	"∞"
Data Set 9	5000	Time limit	1.344	"∞"
Data Set 10	10000	Time limit	1.906	"∞"

Table 1: Average solution time of Models (9)–(B&S) and (10)–(D&T) with respect to different data sets . Time limit = 1000 seconds.

We now seek to compare the quality of the two solutions by simulating random scenarios. We simulate 10,000 instances where each objective coefficient is generated randomly using a Normal distribution. Its mean is the nominal value of the uncertain parameter and the standard deviation is randomly generated between 30% to 100% of the nominal value. Figure 1 displays the cumulative probability distribution for a representative test case. We observe that although the B&S and D&T solutions perform very similarly, in our example the D&T solution slightly dominates the B&S solution.

Our numerical results show that the solution is not very sensitive to the changes in θ. Specifically, we only obtained 2 or 3 different solutions (depending on the data set) in 500 iterations, including the optimal solution for the nominal model, which were all recovered for very small values of θ. Figure 2 provides more insights in the sensitivity of the objective function in θ, for the decision maker seeking to select θ optimally.

4 Extension to Correlated Data

So far we have assumed that the uncertain data are independent. We now relax this assumption. Let the parameter c_i be defined as $c_i = \bar{c}_i + \sum_{j=1}^{n} d_{i,j} z_j$ where the z_j's are independent with mean 0, standard deviation 1 and support $[-b_j, b_j]$, and d represents the Cholesky decomposition of the

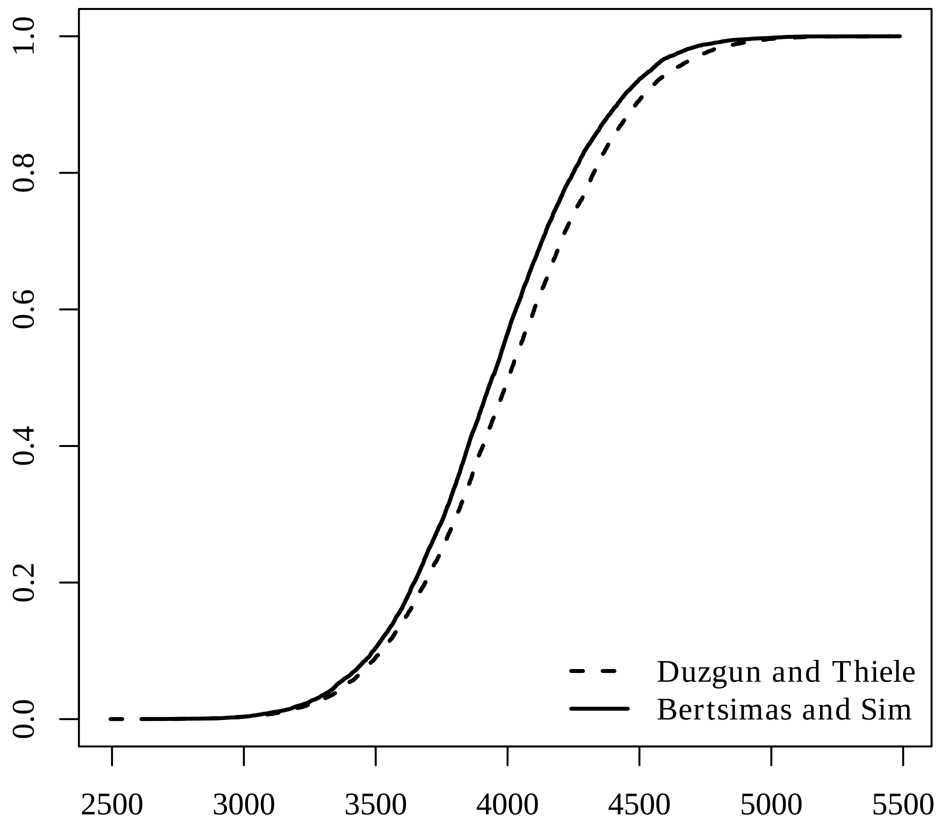

Figure 1: Cumulative Probability Distributions

covariance matrix of c.

Theorem 4.1 (Robust Problem with Bernstein Approximation, correlated case)**.** *The robust counterpart of Problem (1) in the correlated case is:*

$$\begin{aligned} \max_{\theta \geq 0} \frac{\ln \epsilon}{\theta} + \quad \max_{x} \quad & \sum_{i=1}^{n} \left(\bar{c}_i - \frac{1}{\theta} \sum_{j=1}^{n} \ln\, F_\theta(b_j,\, d_{i,j} b_j) \right) x_i \\ s.t. \quad & x \in \mathcal{X} \subseteq \{0,1\}^n. \end{aligned} \tag{10}$$

where F is defined in Eq. (7).

Proof. We have: $\sum_{i=1}^{n} c_i x_i = \sum_{i=1}^{n} \bar{c}_i x_i + \sum_{j=1}^{n} \left(\sum_{i=1}^{n} d_{i,j}\, x_i\right) z_j$. Then:

$$\frac{E[\exp\{-\theta \sum_{i=1}^{n} c_i x_i\}]}{\exp\{-\theta A\}} = \exp\{\theta A\} \prod_{i=1}^{n} \exp\{-\theta x_i \bar{c}_i\} \prod_{j=1}^{n} E[\exp(-\theta x_i\, d_{i,j}\, z_j)]$$

We seek the highest A for which the right-hand term is at most ϵ. This is $(\ln \epsilon)/\theta + \sum_{i=1}^{n} \bar{c}_i\, x_i -$

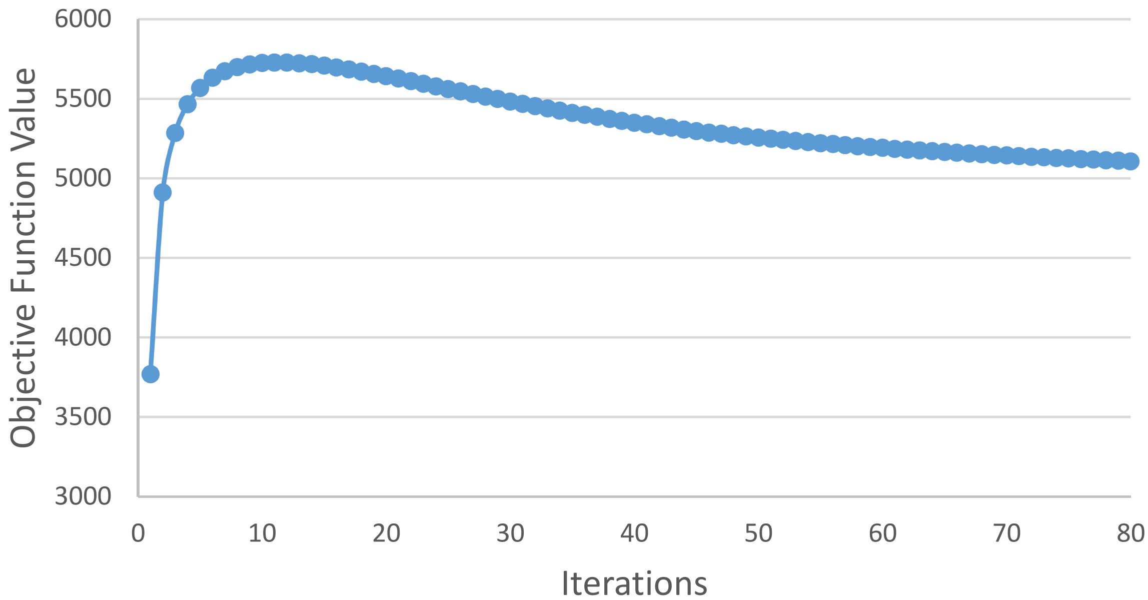

Figure 2: Objective function as a function of θ.

$(1/\theta) \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} \ln E^{UB} [\exp(-\theta d_{i,j} z_j)] \cdot x_i$, where UB refers to an upper bound based on available information on the distribution. Applying Lemma 2.2 to the random variable $d_{i,j} \, z_j$ with mean zero, standard deviation $d_{i,j}$, and support $[-b_j d_{i,j}, \, b_j d_{i,j}]$ allows us to conclude. $\square$

5 Conclusions

In this paper, we have investigated the connection between robust optimization and probabilistic models when decision variables are binary, incorporating the knowledge of the first two moments and the support of the distribution into the formulation. We have shown that the robust counterparts are deterministic problems with modified objective coefficients, which depend on a new parameter introduced in the Bernstein approximation. We have also extended our models to the case when uncertain objective parameters are correlated. The comparison between our approach and the robust discrete optimization model of Bertsimas and Sim [1] shows that our solution quality is comparable to theirs while our approach is significantly faster.

References

[1] D. Bertsimas, M. Sim, Robust discrete optimization and network flows, Mathematical Programming B 98 (2003) 49–71.

[2] A. Ben-Tal, L. El-Ghaoui, A. Nemirovski, Robust optimization, Princeton University Press, Princeton, NJ, 2009.

[3] V. Gabrel, C. Murat, A. Thiele, Recent advances in robust optimization: an overview, European Journal of Operational Research 235 (3) (2014) 471–483.

[4] D. Bertsimas, M. Sim, The price of robustness, Operations Research 52 (1) (2004) 35–53.

[5] D. Bertsimas, I. Popescu, On the relation between option and stock prices: A convex optimization approach, Operations Research 50 (2) (2002) 358–374.

[6] D. Bertsimas, I. Popescu, Optimal inequalities in probability theory: A convex optimization approach, SIAM J on Optim 15 (3) (2005) 780–804.

[7] K. Isii, The extrema of probability determined by generalized moments (i) bounded random variables, Ann Inst Stat Math 12 (1960) 119–133.