RESEARCH ARTICLE

Performance evaluation for distributionally robust optimization with uncertain binary entries

Shunichi Ohmori[*] and Kazuho Yoshimoto

Department of Business Design and Management, Waseda University, Japan
ohmori0406@aoni.waseda.jp, kazuho@waseda.jp

ARTICLE INFO	ABSTRACT
	We consider the data-driven stochastic programming problem with binary entries where the probability of existence of each entry is not known, instead realization of data is provided. We applied the distributionally robust optimization technique to minimize the worst-case expected cost taken over the ambiguity set based on the Kullback-Leibler divergence. We investigate the out-of-sample performance of the resulting optimal decision and analyze its dependence on the sparsity of the problem.

1. Introduction

The impact of uncertainties on planning in decision-making become even more vital than ever, due to rapidly changing and highly volatile business environment. The availability of data is creating new opportunities for better decision making under uncertainty. The traditional decision-making under uncertain data in operations research and management science is the stochastic-programming, where objective function $f : \mathbb{R}^n \to \mathbb{R}$ depends on decision variable x and uncertain parameter u. If we knew full joint-distribution of u, say $p(u)$, we could incorporate uncertainty of u. However, it is often difficult in practice to assume we know full knowledge of the joint distribution. Birtsimas et al. [1] pointed out that probability distribution is imaginary, based on human-assumption, and never be observed in the practice. Data is ever what really exists and is observable. Therefore, they claim that a data-driven decision-making without explicitly considering joint distribution should be appropriate, as it enables us to decide on the basis of evidence rather than assumption.

Another problem of assuming distribution is that, in today's data-rich environment, the prediction needs high-dimensional statistical analysis, as the number of data and variables has been increasing rapidly. This is typically computationally intractable. Further even if it fits well with data set $\{u_1, \cdots, u_N\}$, it is time consuming, because one needs to careful statistical analysis. In the context of inventory management, there are thousands of SKU (Stock-Keeping-Units), each of which have different demand characteristics and shorter life cycles. It is impossible to analyze all these SKUs deeply. Given a need of quicker decision-making and shortage of data-scientists, an autonomous data-driven decision-making has been of great practical interest.

The data-driven stochastic optimization is an alternative paradigm, in which the probability distribution of uncertain parameter is not known, instead realization of data is provided. The data is composed of training-data and validation-data. In the training phase, the data-driven decision

---

*Corresponding Author

and the certificate is output, and in the validation phase, the data-driven decision of the out-of-sample performance is evaluated. A natural approach to solve this problem is sample average approximation (SAA), in which the probability distribution is approximated by the empirical discrete distribution. This approach, however, often leads to the poor out-of-sample performance, because of the optimization-based selection bias, called the optimizer's curse [2].

The distributionally robust optimization (DRO) is a recent emerging topic in data-driven stochastic programming, in which the objective function is the worst-case expectation taken over an ambiguity set. DRO has several good properties [3]. First, the theoretical guarantee of the out-of-sample performance has been derived. Second, asymptotic optimality is proven, that is, as the number of the samples goes to infinity, the data-driven solution converges to the true optimal solution of the problem. Third, the distributionally robust problem can be transformed to the convex programming problem, and thus can be solved very efficiently by the off-the-shelf solver. Finally, a worst-case approach mitigates the optimizer's curse and often leads to the better out-of-sample performance than SAA.

Despite these theoretical advances, the number of applications is still limited. Especially, it is not fully clear on how well the DRO performs for which type of problems. This paper study the performance evaluation of DRO with phi-divergence, applied to the linear programming with partial observed entries. Applications in many settings contains massive data with binary entries, including incidence matrix of the network optimization problem. These data are recorded from observations or measurements, which changes over time in practice. An important problem that arises in applications is to understand how the decision should be made based on the uncertainty of binary entries, using past observation of data. We generate problem instances randomly and apply the DRO and compare the results with SAA. We investigate the out-of-sample performance of the resulting optimal decisions experimentally and analyze its dependence on the number of training samples and the sparsity.

The remainder of the paper is as follows: In section 2, we review related research. In section 3, we outline the conventional stochastic programming, robust optimization, distributionally robust optimization. In section 4, we present the result of numerical experiments. In section 5, we discuss potential extensions and variations.

2. Literature review

2.1. Stochastic programming and robust optimization

Stochastic Programming is the traditional approach for the decision under uncertainty. A fundamental input of the stochastic programming and the chance constrained programming is the probability distribution. However, in practice, it is often difficult to assume the probability distribution.

Robust optimization and chance-constrained programming are alternative popular approach to optimization under uncertainty. Charnes and Cooper [4] first proposed chance constrained programming. Soyster [5] proposed the concepts of uncertainty sets of parameters and find the solution for worst-case value. Ben-Tal and Nemirovski [6–8] and El-Ghaoui et al [9,10] built theoretical foundation for modern robust optimization, with a focus on deriving tractable robust counterpart for the linear programming under ellipsoidal parameter uncertainty. Bertsimas and Sim [11] proposed the concept "price of robustness", which flexibly adjust the level of conservatism of the robust solutions in terms of probabilistic bounds of constraint violations. Ben-Tal et al. [12] propose a systematic way to construct the robust counterpart of a nonlinear uncertain inequality that is concave in the uncertain parameters using support functions, conjugate functions, Fenchel duality. There are extensive review papers. See Ben-Tal et al. [13], Ben-Tal and Nemirovski [14], Gorissen et al. [15], Gabrel, Murat and Thiele [16], Sozuer Thiele [17], Delage and Iancu [18] and the references therein.

Recent papers studies connection with stochastic optimization. Bandi and Bertsimas [19] propose a new approach to analyze stochastic systems based on robust optimization, to overcome the computational intractability with high-dimension. Nemirovski [20] presents several simulation-based and simulation-free computationally tractable approximations of chance constrained convex programs, primarily, those of chance constrained linear, conic quadratic and semidefinite programming.

The applications of robust optimization arises in many field such as finance [21–23], control [24] and supply chain management [25]. Ozmen et al. [21] proposed the methodology that generated the future scenarios through Robust Conic Multivariate Adaptive Regression Splines (RCMARS) that mitigates data uncertainty. They applied their RCMAS to the real-world stock exchange

market financial data. Savku & Weber [22] studied stochastic process model where the underlying regime is switched over time. They proposed a stochastic maximum principle and applied the principle to finance. Kara et al. [23] applied parallelepiped uncertainty set methodology to the portfolio optimization problem and form CVaR. They showed the stability advances in robust portfolio optimization under parallelepiped uncertainty. Baltas et al. [24] studied the optimal control problem in which the model it self is uncertain. They model the problem as the stochastic differential game proposed a robust control of parabolic stochastic partial differential equations under model uncertainty. Sangaiah et al. [25] applied robust optimization and mixed-integer linear programming model for LNG supply chain planning problem.

The key idea of the robust optimization is to define an uncertainty set of possible realizations of the uncertain parameters and then optimize against worst-case realizations within this set [14, 26]. Defining uncertainty set relies on the human-assumption or the probability distribution, which is often difficult. Therefore, the data-driven decision-making approach is of practical importance.

2.2. Distributionally robust optimization

Distributionally robust optimization is a paradigm for decision making under uncertainty where the uncertain problem data are governed by a probability distribution that is itself subject to uncertainty.

Delge and Ye [27] proposed the DRO model with the moment based ambiguity set. Ben-Tal et al. [28] studies the problem of constructing robust classifiers when the training is plagued with uncertainty. They employ Bernstein bounding schemes to relax the chance constrained problem as a convex second order cone program whose solution is guaranteed to satisfy the probabilistic constraint. Dupacova and Kopa [29] studies the robustness for stochastic programs whose set of feasible solutions depends on the unknown probability distribution P and they derived local bounds using contamination technique. Xu et al. [30] investigate probabilistic interpretations of robust optimization. They establish a connection between robust optimization and distributionally robust optimization, showing that the solution to any optimization problem is also a solution to a distributionally robust optimization problem. They consider the case where multiple uncertain parameters belong to the same fixed dimensional space and find the set of distributions

of the equivalent distributionally robust optimization problem. Zymler et al. [31] develop tractable semidefinite programming based approximations for distributionally robust individual and joint chance constraints, assuming that only the first- and second-order moments as well as the support of the uncertain parameters are given. Sun et al. [32] develops a distributionally robust joint chance constrained optimization model for a dynamic network design problem (NDP) under demand uncertainty. Wiesemann et al. [33] introduce standardized ambiguity sets that contain all distributions with prescribed conic representable confidence sets and with mean values residing on an affine manifold. They derived conditions under which distributionally robust optimization problems based on our standardized ambiguity sets are computationally tractable. Ben-Tal et al. [34] proposed distributionally robust optimization with the ambiguity set defined by ϕ-divergences. Bayraksan and Love [35] studies the value of additonal data in the DRO with ϕ-divergence-based ambiguity set. Bertsimas et al. [26] proposed the framework named "data-driven robust optimization" where ambiguity set is formed by goodness-to-fit test. Bertsimas and Kallus [1] proposed the concept "predictive prescription". In this framework, the objective is to minimize conditional expected cost wherein a decision is chosen in an optimal manner to minimize an uncertain cost that depends on a random variable on the basis of an observation of auxiliary covariates. Bertsimas and van Parys [36] proposed the framework named "bootstrap robust analytics", that integrate distributionally robust optimization and statistical bootstrap that are designed to produce out- of-samples guarantees by exploiting the use of a confidence region, derived from ϕ-divergence. Esfahani and Kuhn [3] proposed an ambiguity set derived from the Wasserstein distance.

2.3. Contribution

Our research makes the following contributions over the above cited literature. In theory, the DRO problem mitigate the optimizer's curse and is supposed to derive the better solution than SAA. However, the applications of the DRO are still limited. Therefore, it is important to understand with which types of problems the DRO works well.

Most of the existing DRO research are applied to the variance of data under the condition where entries of uncertain parameters u are full or dense. On the other hand, this research examines the performance of the DRO under the condition where the entries of uncertain parameters u are

sparse. Further, we analyze the dependence performance of the DRO on the sparsity. This result provides the managerial implication of how valuable the observation of the data.

3. Distributionally robust optimization

The following section describes the modeling framework. Section 3.1 provides the data-driven two-stage stochastic programming.

3.1. Problem description

We consider the stochastic linear programming problem (1)

$$
\begin{aligned}
\text{minimize} \quad & \mathbb{E}f(x,u) \\
\text{subject to} \quad & \mathbb{E}g_i(x,u) \le 0, \quad i = 1, \cdots, m
\end{aligned} \tag{1}
$$

where $x \in \mathbb{R}^n$ is the decision variable $u = [c^T, a_1^T, \cdots, a_m^T, b^T]^T$ is the random variable, $f(x,u) = c^T x$ is the objective function and $g_i(x,u) = -a_i^T x + b$ is the constraints. The problem can be expressed as (2).

$$
\begin{aligned}
\text{minimize} \quad & \mathbb{E}[c^T x] \\
\text{subject to} \quad & \mathbb{E}[a_i^T x - b] \ge 0, \quad i = 1 \cdots, m.
\end{aligned} \tag{2}
$$

We assume the k-th entries of u take 1 with probability π_k and 0 with probability $1 - \pi_k$. In practice the distribution $p(u)$ is not known, and therefore must be inferred from data. This is called *data-driven settings*. In the data-driven settings, $p(u)$ is partially observable through a finite set of M independent samples, *e.g.* past realization of the random variable

$$
\mathcal{U}_T := \{u_1, \cdots, u_M\},
$$

called the *training dataset*. In the training phase, we seek a decision $\hat{x}_T$ by minimizing the training problem (3).

$$
\begin{aligned}
\text{minimize} \quad & \mathbb{E}[f(x,u)|\mathcal{U}_T] \\
\text{subject to} \quad & \mathbb{E}[g_i(x,u)|\mathcal{U}_T] \ge 0, \quad i = 1, \cdots, m
\end{aligned} \tag{3}
$$

The solution of the training problem $\hat{x}_T$ is called the *data-driven solution* and the objective function value of the training problem $\hat{z}_T = \mathbb{E}[f(\hat{x}_T, u)|\mathcal{U}_T]$ is called *certificate*.

The goal of a data-driven problem is to minimize *out-of-sample* performance of a data-driven solution $\hat{x}_T$ is defined as (4).

$$
z_V = \mathbb{E}f(\hat{x}_T, u) \tag{4}
$$

As $p(u)$ is unknown, however, the exact out-of-sample performance cannot be evaluated in practice, therefore, it is evaluated by the *validation dataset* $\mathcal{U}_V = \{\hat{u}_1, \cdots, \hat{u}_N\}$ as (5).

$$
\mathbb{E}f(\hat{x}_T, u) \simeq \frac{1}{N} \sum_{j=1}^{N} f(\hat{x}_T, \hat{u}_j) \tag{5}
$$

A natural approach to generate data-driven solutions $\hat{x}_T$ is the sample approximate approximation (SAA) formulation that approximate p with $p_j = (1/M), \forall j \in \mathcal{U}_T$. SAA formulation with training samples u_j can be written as (6).

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{M} \sum_{j=1}^{M} f(x, u_j) \\
\text{subject to} \quad & x \in X
\end{aligned} \tag{6}
$$

This formulation, however, often leads to the poor out-of-sample performance.

3.2. Distributionally robust optimization

In this section, we present distributionally robust optimization (DRO) problem. The DRO has a form (7)

$$
\begin{aligned}
\text{minimize} \quad & \sup_{p \in \mathcal{P}} \mathbb{E}_{p(u)} f(x,u) \\
\text{subject to} \quad & \sup_{p \in \mathcal{P}} \mathbb{E}_{p(u)} g_i(x,u) \le 0 \\
& \quad i = 1, \cdots, m,
\end{aligned} \tag{7}
$$

where $\mathcal{P}$ is *ambiguity set* of probability distribution, that is a family of probability distribution. In the DRO, the worst case expected cost is minimized in which the expectation is taken over the ambiguity set.

The ambiguity set $\mathcal{P}$ is the fundamental input of the DRO, and desired to have the following properties. $\mathcal{P}$ should be rich enough to contain the true data-generating distribution with high confidence. $\mathcal{P}$ should be small enough to exclude pathological distributions, which would incentivize overly conservative decisions. $\mathcal{P}$ should also be easy to parameterize from data. $\mathcal{P}$ should facilitate a tractable reformulation of the distributionally robust optimization problem.

There are several ways to form the ambiguity set $\mathcal{P}$ from the data, *e.g.*, moment ambiguity sets, confidence regions of goodness-of-fit tests., a ball in the space of probability distributions by using a probability distance function such as the Prohorov metric, Wasserstein metric, the ϕ-divergence. This paper examines the properties of DRO problems, where the distributional uncertainty is handled via ϕ-divergences (Ben-Tal et al. [34]). ϕ-divergences measure distances between distributions.

ϕ-divergences measure the distance between two nonnegative vectors $p = (p_1, \cdots, p_M)^T$ and $q = (q_1, \cdots, q_M)^T$, where p and q satisfy $\sum_{j=1}^M p_j = \sum_{j=1}^M q_j = 1$. The ϕ-divergence is defined by (8)

$$D(p, q) = \sum_{j=1}^M q_j \phi(\frac{p_j}{q_j}) \qquad (8)$$

where $\phi(t)$, called the phi-divergence function, which is a convex function on $t \geq 0$. The ϕ divergence satisfies $D(p, q) \geq 0$, and $D(p, q) = 0$ if and only if $p = q$, and so can be used as a measure of deviation between two positive vectors.

There are several ϕ-divergence functions, that hold these properties, each of which performs well. In this research, we use the the *Kullback-Leibler* (KL) divergence between p, q, given by (9).

$$D_{kl}(p, q) = \sum_{j=1}^M (p_j \log(p_j/q_j) - p_j + q_j) \qquad (9)$$

Using the KL divergence, the ambiguity set $\mathcal{P}$ can be expressed as (10).

$$\mathcal{P} = \{p : \begin{array}{l} \sum_{j=1}^M D_{kl}(p, q) \leq \rho, \\ \sum_{j=1}^M p_j = 1, \\ p_j \geq 0, \forall j \}, \end{array} \qquad (10)$$

where q is the nominal value with $q_j = 1/M$, and ρ is the target distance. By restricting the probability distribution in the ambiguity set, i.e., $p \in \mathcal{P}$, the optimization model hedge against the distributional uncertainty. By setting ρ properly, the decision-maker controls the risk preferences between risk-neutral and risk-averse approach.

3.3. Formulation

The formulation of the DRO with KL divergence is presented as formulation (11).

$$\begin{array}{ll} \text{minimize} & \sup_p \sum_{j=1}^M p_j f(x, u_j) \\ \text{subject to} & \sup_p \sum_{j=1}^M p_j g_i(x, u_j) \leq 0, \\ & \qquad\qquad i = 1, \cdots, m \\ & D_{kl}(p, q) \leq \rho, \\ & \sum_{j=1}^M p_j = 1 \\ & p \geq 0 \end{array} \qquad (11)$$

To derive the closed-form of the inner maximization, we present the dual formulation. For a given x, the inner maximization is a convex optimizaion problem. The inner problem is formulated as (12).

$$\begin{array}{ll} \text{maximize}_p & \sum_{j=1}^M p_j f(x, u_j) \\ \text{subject to} & \sum_{j=1}^M q_j \phi(\frac{p_j}{q_j}) \leq \rho, \\ & \sum_{j=1}^M p_j = 1, \\ & p_j \geq 0 \end{array} \qquad (12)$$

Let λ and μ denote the Lagrangian multipliers. Multiplying the first and second constraints by λ and μ and eliminating constraints we have Lagrangian as (13).

$$\begin{array}{ll} \text{maximize}_{p \geq 0} & \sum_{j=1}^M p_j f(x, u_j) \\ & +\lambda\rho - \lambda \sum_{j=1}^M q_j \phi(\frac{p_j}{q_j}) \\ & +\mu - \mu \sum_{j=1}^M p_j \end{array} \qquad (13)$$

For simplicity of exposition, we use s_j to denote as (14).

$$s_j = \frac{f(x, u_j) - \mu}{\lambda} \Leftrightarrow f(x, u_j) = \lambda s_j + \mu \qquad (14)$$

Using this expression, we have the following reformulation.

$$\begin{array}{l} \text{maximize}_{p \geq 0} \quad \sum_{j=1}^M p_j(\lambda s_j - \mu) \\ \qquad +\lambda\rho - \lambda \sum_{j=1}^M q_j \phi(\frac{p_j}{q_j}) \\ \qquad +\mu - \mu \sum_{j=1}^M p_j \\ = \lambda\rho + \mu \\ \qquad +\lambda \sum_{j=1}^M (p_j s_j - q_j \phi(\frac{p_j}{q_j})) \\ = \lambda\rho + \mu \\ \qquad +\lambda \sum_{j=1}^M q_j(s_j \frac{p_j}{q_j} - \phi(\frac{p_j}{q_j})) \end{array}$$

By definition of conjugate function (15)

$$\phi^\star(s) = \sup_{t \geq 0} \{st - \phi(t)\}, \qquad (15)$$

we have the following reformulation (16).

$$\begin{array}{l} \lambda\rho + \mu + \sup_{p \geq 0} \lambda \sum_{j=1}^M q_j(s_j \frac{p_j}{q_j} - \phi(\frac{p_j}{q_j})) \\ = \lambda\rho + \mu + \lambda \sum_{j=1}^M q_j \phi^\star(s_j) \end{array} \qquad (16)$$

Applying the same discussion to the constraint function $\sup_{p \in \mathcal{P}} g_i(x, u)$, we have the following closed form.

$$\sup_{p \in \mathcal{P}} g_i(x, u) = \lambda\rho + \mu + \lambda \sum_{j=1}^M q_j \phi^\star(t_{ij}),$$

where t_{ij} is defined as (17).

$$t_{ij} = \frac{g_i(u_j, x) - \mu}{\lambda} \qquad (17)$$

The conjugate function of KL divergence is defined as (18).

$$\phi^\star(s) = e^s - 1 \qquad (18)$$

Combining the dual problem with the outer minimization results in the dual formulation (19).

$$
\begin{aligned}
\text{minimize} \quad & \mu + \rho\lambda + \lambda \sum_{j=1}^{M} q_j (e^{s_j} - 1) \\
\text{subject to} \quad & \mu + \rho\lambda + \lambda \sum_{j=1}^{M} q_j (e^{t_{ij}} - 1), \\
& \qquad\qquad i = 1, \cdots, m \\
& \lambda \geq 0
\end{aligned}
\qquad (19)
$$

Reader should refer to Ben-Tal et al. [34] and Bayraksan and Love [35] for further details, such as the DRO formulations for other ϕ-divergence functions and the confidence region on the true distribution.

4. Performance evaluation

In this section, we examine the performance of out-of-sample performance of the DRO, presented in previous section. The primal intent of this section is to examine if the DRO performs well with the application to the linear programming with partially observed data, compared to the SAA. We also investigate how much impact the density and sparsity of the data have on the performance of the DRO and SAA solutions. We generate the data set, each of which is a randomly generated sample drawn from the know distribution. Each of the data has different sparsity pattern. We applied both the DRO and the SAA and compared the out-of-sample performance.

We have M training samples and N validation samples. We repeat this same experiment where the decision-maker sees M samples and solves the problem 5000 times. Each time, we got an optimal decision $\hat{x}_T$ which is random variable that depends on training samples. Each of these decisions, we evaluated the objective of the optimization problem by using another N test samples to compute the out-of-samples performance. Each time, we record the optimal value of the optimization problem $\hat{z}_V$.

Experimental conditions are Intel(R) CoreTM i7-8700 (3.20GHz, 3.19GHz) with 32.0 GB memory. Program was coded in Julia with Gurobi optimizer called from Convex.jl.

4.1. Data generation

We use random instances of the data-driven stochastic programming problem (2) generated in the following way. The k-th entries of u takes one with

the probability π_k, and takes zero with $1 - \pi_k$. The probabilities of the entries of the cost vector c are set as $\pi_k = \pi_c$, and the probabilities of the entries of the constraints vectors $a_1, \cdots, a_m, b$ are set as $\pi_k = \pi_{ab}$, where π_c and π_{ab} are set from the $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. Both the probability π_c and π_{ab} can be used to control the degree of density and sparsity of the parameters of the problem. The dimensions of the variable is set as $n = 100$ and the number of constraints are set as $m = 1000$. The sample sizes are set as $M = 10000, N = 5000$.

Table 1. Comparison of average out-of-sample performance.

(a)average out-of-sample performance of the SAA

$\pi_c \backslash \pi_{ab}$	0.1	0.3	0.5	0.7	0.9
0.1	97.40	69.35	9.09	0.36	0.12
0.3	99.17	32.65	12.25	0.75	0.45
0.5	20.73	30.91	26.70	1.26	0.74
0.7	46.91	79.01	92.74	1.65	1.05
0.9	175.13	24.44	78.26	2.20	1.30

(b)average out-of-sample performance of the DRO

| $\pi_c \backslash \pi_{ab}$	0.1	0.3	0.5	0.7	0.9
0.1	1.6	0.78	0.56	0.25	0.13
0.3	5.78	3.42	1.25	0.75	0.44
0.5	9.75	6.73	2.52	1.24	0.74
0.7	14.02	8.68	3.37	1.62	1.05
0.9	18.08	10.83	3.64	2.18	1.3

Table 2. Comparison of standard deviation of out-of-sample performance.

(a)standard deviation of out-of-sample performance of the SAA

$\pi_c \backslash \pi_{ab}$	0.1	0.3	0.5	0.7	0.9
0.1	10.29	61.7	51.64	0.31	0.25
0.3	52.67	28.19	5.56	0.28	0.17
0.5	4.45	34.1	8.03	0.28	0.16
0.7	6.62	1.3	26.2	0.28	0.15
0.9	42.3	7.92	9.2	0.15	0.1

(b)standard deviation of out-of-sample performance of the DRO

$\pi_c \backslash \pi_{ab}$	0.1	0.3	0.5	0.7	0.9
0.1	1.18	0.62	0.39	0.24	0.13
0.3	2.32	1.04	0.44	0.28	0.16
0.5	1.99	1.56	0.5	0.27	0.15
0.7	1.74	1.3	0.53	0.27	0.15
0.9	1.27	1.05	0.34	0.14	0.1
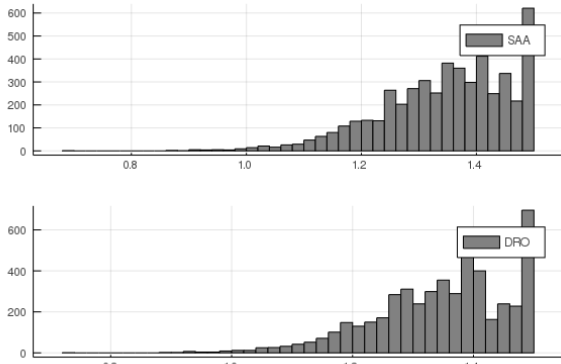
Figure 1. Histogram of out-of-sample performance($\pi_c = 0.9, \pi_{ab} = 0.9$)
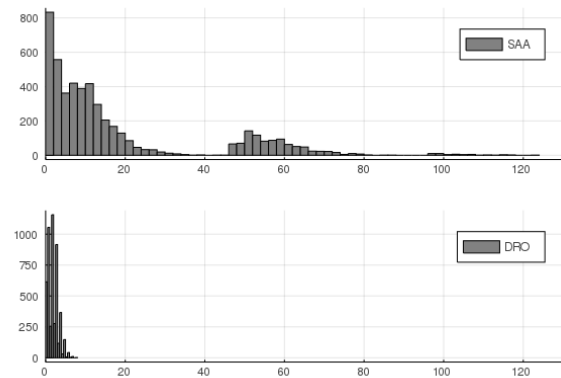
Figure 2. Histogram of out-of-sample performance($\pi_c = 0.1, \pi_{ab} = 0.3$)
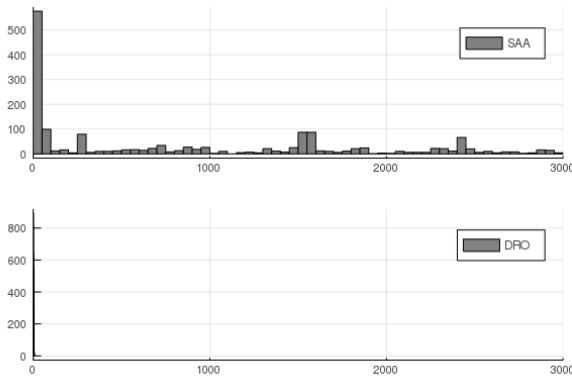
Figure 3. Histogram of out-of-sample performance($\pi_c = 0.1, \pi_{ab} = 0.9$)

4.2. Results

The result of out-of-sample performance of the SAA and the DRO is summarized in the table 1. In all problem instances, the out-of-sample performance of the DRO was lower than the SAA. Especially, when π_c is lower, the average out-of-sample performance is lower. This is because there are less entries in the objective function, which end

up in lower objective function value. when π_{ab} is lower, the average out-of-sample performance is higher. This is because when there are less entries in the constraints, the higher values are assigned to the values of x in order to avoid violence of constraints, which end up in higher objective function value.

The standard deviation of the out-of-sample performance of the SAA and the DRO is summarized in the table 2. The histograms of the out-of-sample performance of the SAA and the DRO for the $(\pi_c = 0.9, \pi_{ab} = 0.9), (\pi_c = 0.1, \pi_{ab} = 0.3)$ and $(\pi_c = 0.3, \pi_{ab} = 0.5)$ and is shown in figure 1,2,3. The variance of the DRO is much smaller than that of the SAA.

Especially, when $\pi_{ab} = 0.1, 0.3, 0.5$, the gap between the DRO and the SAA is significant. This implies that when the sparsity of the constraint vectors has a significant impact on the out-of-sample performance. Decision makers should therefore take these factors into consideration.

All these results imply that the DRO performed consistently better than the SAA the application to the linear programming with partially observed data.

5. Conclusion

The importance of a decision making under uncertainty has never been higher than ever before. The availability of data is creating new opportunities for better decision making under uncertainty. The DRO is a recent emerging topic in optimization under uncertainty, in which the probability distribution of uncertain data is not known, instead realization of data is provided. Despite these theoretical advances, the number of applications is still limited, a decision maker is left with the question of how well the DRO performs for which type of problems.

In this study, this paper addresses the problem in which the decision is made where the probability of existence of each entry is not known, instead realization of data is provided. We present a mathematical formulation of the DRO problem using the ambiguity set of constraints based on the KL divergence. We applied the DRO to randomly generated problem instances and compare the results with SAA. We show the DRO performed consistently better than the SAA. We also discovered that the DRO outperformed the SAA when the constraint vectors have sparse entries. This result raised the importance of analyzing the data structure of the optimization problem for the decision making under the data-driven settings.

For future research, the analysis should be extended to the more general problem such as semidefinite programming (SDP) or second-order cone programming (SOCP). Another related topic is application to the specific topics such as minimum flow cost problem and shortest path problem.

Acknowledgments

References

[1] Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, 66(3), 1025-1044.

[2] Smith, J. E., & Winkler, R. L. (2006). The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3), 311-322.

[3] Esfahani, P. M., & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2), 115-166.

[4] Charnes, A., & Cooper, W. W. (1959). Chance-constrained programming. *Management science*, 6(1), 73-79.

[5] Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, 21(5), 1154-1157.

[6] Ben-Tal, A., & Nemirovski, A. (1998). Robust convex optimization. *Mathematics of operations research*, 23(4), 769-805.

[7] Ben-Tal, A., & Nemirovski, A. (1999). Robust solutions of uncertain linear programs. *Operations research letters*, 25(1), 1-13.

[8] Ben-Tal, A., & Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming*, 88(3), 411-424.

[9] El Ghaoui, L., & Lebret, H. (1997). Robust solutions to least-squares problems with uncertain data. *SIAM Journal on matrix analysis and applications*, 18(4), 1035-1064.

[10] El Ghaoui, L., Oustry, F., & Lebret, H. (1998). Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9(1), 33-52.

[11] Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations research*, 52(1), 35-53.

[12] Ben-Tal, A., Den Hertog, D., & Vial, J. P. (2015). Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical programming*, 149(1-2), 265-299.

[13] Ben-Tal, A., & Nemirovski, A. (2008). Selected topics in robust convex optimization. *Mathematical Programming*, 112(1), 125-158.

[14] Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization* (Vol. 28). Princeton University Press.

[15] Gorissen, B. L., Yanıkoglu, İ., & den Hertog, D. (2015). A practical guide to robust optimization. *Omega*, 53, 124-137.

[16] Gabrel, V., Murat, C., & Thiele, A. (2014). Recent advances in robust optimization: An overview. *European journal of operational research*, 235(3), 471-483.

[17] Sozuer, S., & Thiele, A. C. (2016). *The state of robust optimization. In Robustness Analysis in Decision Aiding, Optimization, and Analytics* (pp. 89-112). Springer, Cham.

[18] Delage, E., & Iancu, D. A. (2015). *Robust multistage decision making. In The Operations Research Revolution* (pp. 20-46). INFORMS.

[19] Bandi, C., & Bertsimas, D. (2012). Tractable stochastic analysis in high dimensions via robust optimization. *Mathematical programming*, 134(1), 23-70.

[20] Nemirovski, A. (2012). On safe tractable approximations of chance constraints. *European Journal of Operational Research*, 219(3), 707-718.

[21] Ozmen, A., Weber, G. W., & Karimov, A. (2013). A new robust optimization tool applied on financial data.

[22] Savku, E., & Weber, G. W. (2018). A stochastic maximum principle for a markov regime-switching jump-diffusion model with delay and an application to finance. *Journal of Optimization Theory and Applications*, 179(2), 696-721.

[23] Kara, G., Özmen, A., & Weber, G. W. (2019). Stability advances in robust portfolio optimization under parallelepiped uncertainty. *Central European Journal of Operations Research*, 27(1), 241-261.

[24] Baltas, I., Xepapadeas, A., & Yannacopoulos, A. N. (2019). Robust control of parabolic stochastic partial differential equations under model uncertainty. *European Journal of Control*, 46, 1-13.

[25] Sangaiah, A. K., Tirkolaee, E. B., Goli, A., & Dehnavi-Arani, S. (2019). Robust optimization and mixed-integer linear programming

model for LNG supply chain planning problem. *Soft Computing*, 1-21.

[26] Bertsimas, D., Gupta, V., & Kallus, N. (2018). Data-driven robust optimization. *Mathematical Programming*, 167(2), 235-292.

[27] Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3), 595-612.

[28] Ben-Tal, A., Bhadra, S., Bhattacharyya, C., & Nath, J. S. (2011). Chance constrained uncertain classification via robust optimization. *Mathematical programming*, 127(1), 145-173.

[29] Dupacova, J., & Kopa, M. (2012). Robustness in stochastic programs with risk constraints. *Annals of Operations Research*, 200(1), 55-74.

[30] Xu, H., Caramanis, C., & Mannor, S. (2012). A distributional interpretation of robust optimization. *Mathematics of Operations Research*, 37(1), 95-110.

[31] Zymler, S., Kuhn, D., & Rustem, B. (2013). Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2), 167-198.

[32] Sun, H., Gao, Z., Szeto, W. Y., Long, J., & Zhao, F. (2014). A distributionally robust joint chance constrained optimization model for the dynamic network design problem under demand uncertainty. *Networks and Spatial Economics*, 14(3-4), 409-433.

[33] Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, 62(6), 1358-1376.

[34] Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., & Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2), 341-357.

[35] Bayraksan, G., & Love, D. K. (2015). *Data-driven stochastic programming using phi-divergences. In The Operations Research Revolution* (pp. 1-19). INFORMS.

[36] Bertsimas, D., & Van Parys, B. (2017). Bootstrap robust prescriptive analytics. *arXiv preprint arXiv:1711.09974.*

Shunichi Ohmori (PhD) *is an associate professor at department of industrial and system engineering, Waseda University in Japan, and a researcher at institute of globalproduction and logistics at Waseda University, and a researcher at data science institute at Waseda University. He received the master and Ph.D degree in engineering at Waseda University. His research interest lies in operations research and supply chain management.*
ⓘ*D https://orcid.org/0000-0002-2635-0450*

Kazuho Yoshimoto (Dr. Engg.) *is a professor at department of industrial and system engineering at Waseda University in Japan, and a head of Institute of Global Production and Logistics at Waseda University. He received the master and Ph.D degree in engineering at Waseda University. His research interest lies in facility and logistics design.*
ⓘ*D https://orcid.org/0000-0003-2870-9612*