Distributionally robust chance-constrained Markov decision processes with random payoff

Hoang Nam Nguyen, Abdel Lisser, Vikas Vikram Singh

Distributionally robust chance-constrained Markov decision processes with random payoff

Hoang Nam Nguyen[1†], Abdel Lisser[1*†], Vikas Vikram Singh[2†]

[1]Universite Paris-Saclay, CNRS, CentraleSupelec, Laboratory of signals and systems, 3 rue Joliot-Curie, Gif-sur-Yvette, 91190, France.
[2]Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, 110016, India.

*Corresponding author(s). E-mail(s): abdel.lisser@centralesupelec.fr;
Contributing authors: hoang-nam.nguyen3@centralesupelec.fr;
vikassingh@maths.iitd.ac.in;
†These authors contributed equally to this work.

Abstract

A Markov Decision Process (MDP) is a framework used for decision-making. In an MDP problem, the decision maker's goal is to maximize the expected discounted value of future rewards while navigating through different states controlled by a Markov chain. In this paper, we focus on the case where the transition probabilities vector is deterministic, while the reward vector is uncertain and follow a partially known distribution. We employ a distributionally robust chance constraints approach to model the MDP. This approach entails the construction of potential distributions of reward vector, characterized by moments or statistical metrics. We explore two situations for these uncertainty sets: one where the reward vector has a real support and another where it is constrained to be nonnegative. In the case of a real support, we demonstrate that solving the distributionally robust chance-constrained Markov decision process is mathematically equivalent to a second-order cone programming problem for moments and ϕ-divergence uncertainty sets. For Wasserstein distance uncertainty sets, it becomes a mixed-integer second-order cone programming problem. In contrast, when dealing with nonnegative reward vector, the equivalent optimization problems are different. Moments-based uncertainty sets lead to a copositive optimization problem, while Wasserstein distance-based uncertainty sets result in a biconvex optimization problem. To illustrate the practical application of these

methods, we examine a machine replacement problem and present results conducted on randomly generated instances to showcase the effectiveness of our proposed methods.

Keywords: Markov decision process, Distributionally robust chance constraints, Second-order cone programming, Copositive optimization, Mix-integer second-order cone programming, Biconvex optimization

1 Introduction

An MDP serves as a foundational framework for decision-making in modeling the dynamic of a randomized system that changes as time progresses under the influence of a managed Markov chain. We assume that the system comprises a finite set of states. At the initial time step, designated as $t = 0$, the system commences from an initial state s_0, drawn from a finite state space S, in accordance with a predetermined initial distribution ϱ. Subsequently, a decision maker selects an action, denoted as a_0, from the set of feasible actions at state s_0, represented as $A(s_0)$. This choice of action yields a reward, denoted as $R(s_0, a_0)$. Following this, at $t = 1$, the system does a transition to a new state, s_1, with a probability $p(s_0, a_0, s_1)$. This process perpetuates indefinitely into the future.

The decisions made at each time step, t, whether deterministic or stochastic, may depend on the historical record available up to that point, denoted as h_t. This historical record, h_t, encompasses information regarding the observed states and actions up to time t, represented as $h_t = (s_0, \ldots, a_{t-1}, s_t)$. We denote the set of all histories at time t as H_t. A decision rule at time t, denoted as f_t, is characterized by $f_t(h_t) \in \wp(A(s_t))$, where $\wp(A(s_t))$ is the set of all probability measures over the action set $A(s_t)$.

A sequence of decision rules contingent upon history, referred to as a history-dependent policy, is symbolized as $f^h = (f_0, f_1, \ldots)$. A policy is stationary if there is a single decision policy, represented as f, such that $f = f_t$ for any t. In this context, we informally denote a stationary policy as f and stipulate $f = (f(s))s \in S$, with $f(s) \in \wp(A(s))$ for every state $s \in S$. Under a stationary policy f, at each state s, the decision maker selects an action a with a probability determined by $f(s, a)$.

We designate the set of all stationary policies as PO_S. A stationary policy $f \in PO_S$, combined with an initial distribution ϱ, gives rise to a probability distribution P_ϱ^f governing the trajectories of states and actions. The expectation value associated with this probability measure is denominated as E_ϱ^f. The expected discounted reward with a discount factor $\alpha \in (0, 1)$ is defined as [1, 21]

$$V_\alpha(\varrho, f) = (1 - \alpha)\mathbb{E}_\varrho^f \left(\sum_{t=0}^{\infty} \alpha^t R(X_t, A_t) \right). \tag{1}$$

Here, we introduce the notations X_t and A_t to represent the state and action at time t. When the running rewards and the transition probabilities are stationary, we can restrict to stationary policies without loss of optimality [1, 21]. It follows from Theorem

3.2 of [1] that a stationary optimal policy of the MDP problem can be obtained by solving following linear programming problem

$$\max_{\chi \in \mathcal{Q}_\alpha(\varrho)} \chi^{\mathrm{T}} R, \tag{2}$$

where $R = (R(s,a))_{s\in S, a\in A(s)}$ is a running reward vector and T denotes the transposition. The set $\mathcal{Q}_\alpha(\varrho)$ is defined as

$$\mathcal{Q}_\alpha(\varrho) = \Bigg\{ \chi \in \mathbb{R}^{|\mathcal{W}|} \mid \sum_{(s,a)\in\mathcal{W}} \chi(s,a)\Big(-\alpha p(s,a,s') + \delta(s',s)\Big) = (1-\alpha)\varrho(s'),\ s' \in S,$$
$$\chi(s,a) \geq 0,\ \forall\ s \in S,\ a \in A(s) \Bigg\},$$

where $\delta(s',s)$ is the Kronecker delta and $\mathcal{W} = \{(s,a) \mid s \in S, a \in A(s)\}$. When χ^* stands as an optimal solution to (2), the stationary optimal policy f^* can be formulated as follows

$$f^*(s,a) = \frac{\chi^*(s,a)}{\sum_{a\in A(s)} \chi^*(s,a)},\ \forall\ s \in S,\ a \in A(s),$$

This formulation holds true as long as the denominator is non-zero. If not, we arbitrarily select $f^*(s)$ from the set of probability distributions over $A(s)$ [1]. In practical applications, the parameters of the MDP model, $R(\cdot)$ and $p(\cdot)$, are typically not known in advance. Instead, they are estimated based on historical data. However, this process introduces inaccuracies into the optimal policies [15]. Consequently, addressing this uncertainty has been an important point in the study of robust MDPs. In these robust MDPs, we focus on scenarios where either the transition probabilities or the rewards are assumed to belong to predefined uncertainty sets, as discussed in the literature [12, 18, 25, 29, 30]. However, Delage and Mannor [5] made a significant observation that the robust MDP often results in overly conservative policies. To address this issue, they introduced the concept of a chance-constrained Markov Decision Process (CCMDP) in their work, wherein the controller seeks to achieve an expected discounted reward with a certain level of confidence. In their research [5], they delved into two distinct cases, random transition probabilities random rewards are studied independently. They stated that a CCMDP can be effectively reformulated as an SOCP problem if the reward vector follows a multivariate Gaussian distribution and the transition probabilities are known. However, in situation where the transition probabilities follow a Dirichlet distribution while the reward vector is known, solving the CCMDP problem is computationally challenging, necessitating the use of approximation methods to obtain optimal policies. In a related development, Varagapriya et al. [26] examined a CMDP problem with a joint chance constraint. In this scenario, the running cost vectors are treated as random vectors, while the transition probabilities are known. They proposed lower and upper bounds for the CMDP problem,

assuming that the cost vectors follow multivariate elliptical distributions and that the dependence between constraints is governed by an Archimedean copula.

In numerous practical scenarios, we often encounter situations where we possess only partial information about the underlying distribution, typically derived from historical data. In such cases, a distributionally robust approach serves as a tool to model uncertainties. This methodology has found extensive application in modeling uncertainties in various optimization and game-related problems [13, 14, 23].

There are two methods for constructing an uncertainty set. The first method relies on partial information regarding the moments of the true distribution, while the second method employs statistical distances to measure the difference between the true distribution and a nominal distribution. Moments-based uncertainty sets make specific assumptions about the first two moments of the distribution [4, 6, 20]. On the other hand, statistical distance-based uncertainty sets encompass all distributions lying within a small-radius ball centered at the nominal distribution. This nominal distribution is often a Gaussian distribution or an empirical distribution [7, 13]. To quantify the distance between distributions, either a $\phi-$divergence [2, 13] or the Wasserstein distance metric is commonly employed [7, 8, 32].

We consider an infinite horizon MDP with a discounted reward criterion, as introduced in Section 1. In this setting, the reward vector is characterized as a random vector, while the transition probabilities are assumed to be known. The distribution of the reward vector is not entirely known and assumed to belong to a predefined uncertainty set. Our primary objective is to formulate the problem of random discounted rewards while incorporating a distributionally robust chance constraint. This constraint is designed to ensure that a given policy maximizes the reward with a specified level of confidence. This particular class of MDPs is referred to as a distributionally robust chance-constrained Markov Decision Process (DRCCMDP). The random reward vector has either a full support or a nonnegative support. Our primary contributions are as follows.

1. We consider three types of moments based uncertainty sets based on the full/partial information on the first two moments of the random reward vector. For the case of full support and nonnegative support, a DRCCMDP problem is equivalent to an SOCP problem and a copositive optimization problem, respectively.
2. We consider four different types of ϕ-divergences to construct statistical distance based uncertainty sets. We show that a DRCCMDP problem is equivalent to an SOCP problem when the nominal distribution is a normal distribution.
3. We consider the nominal distribution to be an empirical distribution when statistical distance based uncertainty set is defined with Wasserstein distance metric. For the case of full support and nonnegative support, we show that a DRCCMDP problem is equivalent to a mixed integer second-order cone programming (MISOCP) problem and a biconvex optimization problem, respectively.

The structure of the paper is as follows. In Section 2, we present the definition of a DRCCMDP under a discounted reward criterion. Section 3 contains a DRCCMDP under moments based uncertainty sets and their equivalent reformulations for the case of full and nonnegative supports. A DRCCMDP under statistical distance based

uncertainty sets defined using ϕ-divergence distance metric and Wasserstein distance metric and their equivalent reformulations are presented in Section 4. We present numerical results in Section 5 and conclude the paper in Section 6.

2 The model

We consider an infinite horizon MDP defined in Section 1 where the transition probabilities are exactly known and the running reward vector is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which is denoted as $\hat{R}$. Therefore, for each realization $\omega \in \Omega$, $\hat{R}(s, a, \omega)$ represents a real valued reward received at state s when an action a is taken. We assume that the random vector $\hat{R}$ does not vary with time. Since $\hat{R}$ is a random vector, for a given policy f^h and initial distribution ϱ, the expected discounted reward defined by (1) becomes a random variable. The controller wants to maximize the discounted reward which can be achieved with at least a specified confidence level $(1-\epsilon)$, where $\epsilon \in (0, 1)$. This leads to the following CCMDP problem

$$\begin{aligned} &\sup_{y \in \mathbb{R},\ f^h \in F_{HD}} \quad y \\ &\text{s.t.} \quad \mathbb{P}\left(V_\alpha(s, f^h) \geq y\right) \geq 1 - \epsilon. \end{aligned} \tag{3}$$

Since the transition probabilities are exactly known and random reward vector does not depend on time, we can represent the CCMDP problem (3) equivalently in terms of decision vector (y, χ) as follows

$$\begin{aligned} &\sup \quad y \\ &\text{s.t.} \quad \text{(i)}\ \mathbb{P}\left(\chi^{\mathrm{T}} \hat{R} \geq y\right) \geq 1 - \epsilon,\ \text{(ii)}\ \chi \in \mathcal{Q}_\alpha(\varrho). \end{aligned} \tag{4}$$

If then vector $\hat{R}$ follows a multivariate normal distribution, the optimization problem (4) is equivalent to an SOCP problem [5]. The above result can be generalized for elliptically symmetric distributions because the linear chance constraint (i) present in (4) is equivalent to a second order cone constraint [10].

Nevertheless, in most of practical scenarios, we are confronted with limited information about the true probability distributions at hand. To address this, we use the distributionally robust optimization approach, where the distribution of $\hat{R}$ falls within an uncertainty set. This formulation leads us to the subsequent problem, namely the Distributionally Robust Chance-Constrained Markov Decision Process (DRCCMDP) problem

$$\begin{aligned} &\sup \quad y \\ &\text{s.t.} \quad \text{(i)}\ \inf_{F \in \mathcal{D}} \mathbb{P}_F\left(\chi^{\mathrm{T}} \hat{R} \geq y\right) \geq 1 - \epsilon,\ \text{(ii)}\, \chi \in \mathcal{Q}_\alpha(\varrho), \end{aligned} \tag{5}$$

where F is the distribution of $\hat{R}$ and $\mathcal{D}$ is a given uncertainty set. Note that $\mathbb{P}_F(\chi^{\mathrm{T}} \hat{R} \leq y - \theta) \leq \mathbb{P}_F(\chi^{\mathrm{T}} \hat{R} < y) \leq \mathbb{P}_F(\chi^{\mathrm{T}} \hat{R} \leq y)$ for every $\theta > 0$. Therefore, we can replace

$\sup_{F\in\mathcal{D}} \mathbb{P}_F\left(\chi^{\mathrm{T}}\hat{R} < y\right)$ by $\sup_{F\in\mathcal{D}} \mathbb{P}_F\left(\chi^{\mathrm{T}}\hat{R} \leq y\right)$. Then, the problem (5) is equivalent to the following problem

$$\begin{aligned} & \sup \quad y \\ \text{s.t.} \quad & \text{(i)} \ \sup_{F\in\mathcal{D}} \mathbb{P}_F\left(\chi^{\mathrm{T}}\hat{R} \leq y\right) \leq \epsilon, \ \text{(ii)} \ \chi \in \mathcal{Q}_\alpha(\varrho). \end{aligned} \tag{6}$$

In the following sections, we study different types of uncertainty sets of $\hat{R}$ which are defined using i) partial information of moments of $\hat{R}$, ii) ϕ-divergence distance, and iii) Wasserstein distance. For each uncertainty set, we consider the cases of full and nonnegative supports of $\hat{R}$. We derive equivalent reformulations of DRCCMDP problem (5) (or (6) equivalently) for each uncertainty set.

3 Uncertainty sets based on moments

Let $\mu \in \mathbb{R}^{|\mathcal{K}|}$ be the mean vector and $\Sigma \succ 0$ be a $|\mathcal{K}| \times |\mathcal{K}|$ positive definite matrix. We consider 3 types of moments based uncertainty sets of the distribution of $\hat{R}$ defined as follows:

1. **Known mean and known covariance uncertainty set:** The uncertainty set of the distribution of $\hat{R}$ in this case is defined by

$$\mathcal{D}_1(\varphi, \mu, \Sigma) = \left\{ F \in \mathcal{M}^+ \left| \begin{array}{l} \mathbb{E}(\mathbf{1}_{\{\hat{R}\in\varphi\}}) = 1, \\ \mathbb{E}(\hat{R}) = \mu, \\ \mathbb{E}[(\hat{R}-\mu)(\hat{R}-\mu)^{\mathrm{T}}] = \Sigma. \end{array} \right. \right\}, \tag{7}$$

2. **Known mean and unknown covariance uncertainty set:** The uncertainty set of the distribution of $\hat{R}$ in this case is defined by

$$\mathcal{D}_2(\varphi, \mu, \Sigma, \delta_0) = \left\{ F \in \mathcal{M}^+ \left| \begin{array}{l} \mathbb{E}(\mathbf{1}_{\{\hat{R}\in\varphi\}}) = 1, \\ \mathbb{E}(\hat{R}) = \mu, \\ \mathbb{E}[(\hat{R}-\mu)(\hat{R}-\mu)^{\mathrm{T}}] \preceq \delta_0\Sigma. \end{array} \right. \right\}, \tag{8}$$

3. **Unknown mean and unknown covariance uncertainty set:** The uncertainty set of the distribution of $\hat{R}$ in this case is defined by

$$\mathcal{D}_3(\varphi, \mu, \Sigma, \delta_1, \delta_2) = \left\{ F \in \mathcal{M}^+ \left| \begin{array}{l} \mathbb{E}(\mathbf{1}_{\{\hat{R}\in\varphi\}}) = 1, \\ [\mathbb{E}(\hat{R})-\mu]^{\mathrm{T}}\Sigma^{-1}[\mathbb{E}(\hat{R})-\mu] \leq \delta_1, \\ \mathbb{E}[(\hat{R}-\mu)(\hat{R}-\mu)^{\mathrm{T}}] \preceq \delta_2\Sigma. \end{array} \right. \right\}, \tag{9}$$

where $\varphi \subset \mathbb{R}^{|\mathcal{W}|}$ is the support of $\hat{R}$ which we assume to be a convex set, $\mathcal{M}^+$ is the set of all probability measures on $\mathbb{R}^{|\mathcal{W}|}$ with Borel $\sigma-$algebra, $\delta_1 \geq 0, \delta_2, \delta_0 \geq 1$, $\mu \in \mathrm{RI}(\varphi)$; $\mathrm{RI}(\varphi)$ denotes the relative interior of φ. The notation $C \preceq D$ implies that $D - C$ is a positive semidefinite matrix and $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function.

3.1 DRCCMDP with moments based uncertainty sets under full support

We consider the case when the random vector $\hat{R}$ has full support, i.e., $\varphi = \mathbb{R}^{|\mathcal{W}|}$. We show that the DRCCMDP problem is equivalent to an SOCP problem.

Theorem 1. *Consider the DRCCMDP problem* (5) *where the distribution of* $\hat{R}$ *belongs to the uncertainty sets defined by* (7), (8), (9), *and the support* $\varphi = \mathbb{R}^{|\mathcal{W}|}$. *Then, the* DRCCMDP (5) *can be reformulated equivalently as the following* SOCP *problem*

$$\begin{aligned} & \max \quad y \\ s.t. \quad & \text{(i) } \mu^{\mathrm{T}}\chi - \kappa\|\Sigma^{\frac{1}{2}}\chi\|_2 \geq y, \text{ (ii) } \chi \in \mathcal{Q}_\alpha(\varrho), \end{aligned} \tag{10}$$

where $||\cdot||_2$ *denotes the Euclidean norm and* κ *is a real number whose value for each uncertainty set is given in Table 1.*

Table 1 Value of κ for moments based uncertainty set

Uncertainty set	$\mathcal{D} = \mathcal{D}_1(\varphi, \mu, \Sigma)$	$\mathcal{D} = \mathcal{D}_2(\varphi, \mu, \Sigma, \delta_0)$	$\mathcal{D} = \mathcal{D}_3(\varphi, \mu, \Sigma, \delta_1, \delta_2)$
κ	$\sqrt{\frac{1-\epsilon}{\epsilon}}$	$\sqrt{\frac{(1-\epsilon)\delta_0}{\epsilon}}$	$\sqrt{\frac{(1-\epsilon)\delta_2}{\epsilon}} + \sqrt{\delta_1}$

Proof. The proof follows from the fact that for each uncertainty set the distributionally robust chance constraint (i) of (5) is equivalent to a second-order cone constraint. The uncertainty set (7) has been widely studied in the literature [3, 9]. For the uncertainty sets (8) and (9), it can be proved using similar arguments used in Lemma 3.1 and Lemma 3.2 of [16] which are based on the one-sided Chebyshev inequality [14]. □

3.2 DRCCMDP with moments based uncertainty sets under nonnegative support

We consider the case where the support of the random vector $\hat{R}$ is a nonnegative orthant of $|\mathcal{W}|$-dimensional Euclidean space, i.e., $\varphi = \mathbb{R}_+^{|\mathcal{W}|}$. We show that the DRCCMDP problem (6) is equivalent to a copositive optimization problem.

Theorem 2. *Consider a DRCCMDP problem* (6) *with* $\varphi = \mathbb{R}_+^{|\mathcal{W}|}$. *Then, the following results hold.*

1. *If the uncertainty set is defined by* (7), *the DRCCMDP problem* (6) *is equivalent to the following copositive optimization problem*

$$\begin{aligned} & \max \quad y \\ s.t. \quad & \text{(i) } -\bar{t} - \bar{Q} \circ \Sigma - \bar{q}^{\mathrm{T}}\mu \leq s\epsilon, \\ & \text{(ii) } \left(\begin{array}{c|c} -\bar{Q} & -\frac{1}{2}\bar{q} + \bar{Q}\mu \\ \hline -\frac{1}{2}\bar{q}^{\mathrm{T}} + \mu^{\mathrm{T}}\bar{Q} & -\bar{t} - \mu^{\mathrm{T}}\bar{Q}\mu \end{array}\right) \in \mathrm{COP}^{|\mathcal{W}|+1}, \end{aligned}$$

$$\text{(iii)} \left(\begin{array}{c|c} -\bar{Q} & -\frac{1}{2}\bar{q} + \bar{Q}\mu + \chi \\ \hline -\frac{1}{2}\bar{q}^{\mathrm{T}} + \mu^{\mathrm{T}}\bar{Q} + \chi^{\mathrm{T}} & -\bar{t} - \mu^{\mathrm{T}}\bar{Q}\mu - s - y \end{array}\right) \in \mathrm{COP}^{|\mathcal{W}|+1},$$
$$\text{(iv)}\ \bar{Q} \in \mathcal{S}^{|\mathcal{W}|}, s \geq 0,\ \chi \in \mathcal{Q}_\alpha(\varrho). \tag{11}$$

2. *If the uncertainty set is defined by (8), the DRCCMDP problem (6) is equivalent to the following copositive optimization problem*

$$\begin{array}{ll} & \max \quad y \\ s.t. & \text{(i)}\ -\bar{t} - \mu^{\mathrm{T}}\bar{q} - \mu^{\mathrm{T}}\bar{Q}\mu + \delta_0\Sigma \circ \bar{Q} \leq s\epsilon, \\ & \text{(ii)}\ \left(\begin{array}{c|c} \bar{Q} & -\frac{1}{2}\bar{q} - \bar{Q}\mu \\ \hline -\frac{1}{2}\bar{q}^{\mathrm{T}} - \mu^{\mathrm{T}}\bar{Q} & -\bar{t} \end{array}\right) \in \mathrm{COP}^{|\mathcal{W}|+1}, \\ & \text{(iii)}\ \left(\begin{array}{c|c} \bar{Q} & \frac{1}{2}(-\bar{q}+\chi) - \bar{Q}\mu \\ \hline \frac{1}{2}(-\bar{q}+\chi)^{\mathrm{T}} - \mu^{\mathrm{T}}\bar{Q} & -\bar{t} - s - y \end{array}\right) \in \mathrm{COP}^{|\mathcal{W}|+1}, \\ & \text{(iv)}\ \bar{Q} \in \mathcal{S}_+^{|\mathcal{W}|}, s \geq 0,\ \chi \in \mathcal{Q}_\alpha(\varrho). \end{array} \tag{12}$$

3. *If the uncertainty set is defined by (9), the DRCCMDP problem (6) is equivalent to the following copositive optimization problem*

$$\begin{array}{ll} & \max \quad y \\ s.t. & \text{(i)}\ \bar{r} + \bar{t} \leq s\epsilon, \\ & \text{(ii)}\ \left(\begin{array}{c|c} \bar{Q} & \frac{1}{2}\bar{q} \\ \hline \frac{1}{2}\bar{q}^{\mathrm{T}} & \bar{r} \end{array}\right) \in \mathrm{COP}^{|\mathcal{W}|+1}, \\ & \text{(iii)}\ \bar{t} \geq (\delta_2\Sigma + \mu\chi^{\mathrm{T}}) \circ \bar{Q} + \chi^{\mathrm{T}}\bar{q} + \sqrt{\delta_1}||\Sigma^{\frac{1}{2}}(\bar{q} + 2\bar{Q}\mu)||_2, \\ & \text{(iv)}\ \left(\begin{array}{c|c} \bar{Q} & \frac{1}{2}(\bar{q}+\chi) \\ \hline \frac{1}{2}(\bar{q}+\chi)^{\mathrm{T}} & \bar{r} - s - y \end{array}\right) \in \mathrm{COP}^{|\mathcal{W}|+1}, \\ & \text{(v)}\ \bar{Q} \in \mathcal{S}_+^{|\mathcal{W}|}, s \geq 0,\ \chi \in \mathcal{Q}_\alpha(\varrho), \end{array} \tag{13}$$

where $\mathrm{COP}^{|\mathcal{W}|+1} = \left\{M \in \mathcal{S}^{|\mathcal{W}|+1} \mid x^{\mathrm{T}}Mx \geq 0,\ \forall\ x \in \mathbb{R}_+^{|\mathcal{W}|+1}\right\}$, $\mathcal{S}^n$ *is the set of all real symmetric matrix of size* $n \times n$, $\mathcal{S}_+^n$ *is the set of positive semidefinite matrices of size* $n \times n$, $\circ$ *denotes the Frobenius inner product and* $\left(\begin{array}{c|c} & \\ \hline & \end{array}\right)$ *denotes a block matrix (or a partitioned matrix).*

In order to prove the first result of Theorem 2, we need the following lemma.

Lemma 1. *Consider an optimization problem*

$$\sup_{F \in \mathcal{D}_1(\varphi,\mu,\Sigma)} \mathbb{P}_F(\chi^{\mathrm{T}}\hat{R} \leq y), \tag{14}$$

where $\varphi = \mathbb{R}_+^{|\mathcal{W}|}$. *If the feasible set of (14) is non-empty, the dual of (14) is given by*

$$\inf \quad -t - Q \circ \Sigma - q^T\mu$$

$$
\begin{aligned}
s.t. \quad & \text{(i) } \mathbf{1}_{\{\chi^T\xi\leq y\}} + q^T\xi + \xi^T Q\xi - 2\xi^T Q\mu + \mu^T Q\mu + t \leq 0,\ \forall\ \xi \in \mathbb{R}_+^{|\mathcal{W}|}, \\
& \text{(ii) } Q \in \mathcal{S}^{|\mathcal{W}|},
\end{aligned}
$$

such that strong duality holds.

Proof. The proof is given in Appendix A. □

Proof of Theorem 2. 1. Let the distribution of $\hat{R}$ belongs to the uncertainty set $\mathcal{D}_1(\phi, \mu, \Sigma)$. Using Lemma 1, the optimization problem (6) is equivalent to the following problem

$$
\begin{aligned}
& \sup \quad y \\
\text{s.t.} \quad & \text{(i) } -t - Q \circ \Sigma - q^{\mathrm{T}}\mu \leq \epsilon, \\
& \text{(ii) } q^{\mathrm{T}}\xi + \xi^{\mathrm{T}} Q\xi - 2\xi^{\mathrm{T}} Q\mu + \mu^{\mathrm{T}} Q\mu + t \leq 0,\ \forall\ \xi \in \mathbb{R}_+^{|\mathcal{W}|}, \\
& \text{(iii) } 1 + q^{\mathrm{T}}\xi + \xi^{\mathrm{T}} Q\xi - 2\xi^{\mathrm{T}} Q\mu + \mu^{\mathrm{T}} Q\mu + t \leq 0,\ \forall\ \xi \in \mathbb{R}_+^{|\mathcal{W}|},\ \chi^{\mathrm{T}}\xi \leq y, \\
& \text{(iv) } Q \in \mathcal{S}^{|\mathcal{W}|}, \chi \in \mathcal{Q}_\alpha(\varrho).
\end{aligned}
\tag{15}
$$

The constraint (ii) of (15) is equivalent to:

$$
(\xi^{\mathrm{T}}, 1) U (\xi^{\mathrm{T}}, 1)^{\mathrm{T}} \geq 0,\ \forall\ \xi \in \mathbb{R}_+^{|\mathcal{W}|},
$$

where $U \in \mathcal{S}^{|\mathcal{W}|+1}$ such that

$$
U = \left(\begin{array}{c|c} -Q & -\frac{1}{2}q + Q\mu \\ \hline -\frac{1}{2}q^{\mathrm{T}} + \mu^{\mathrm{T}} Q & -t - \mu^{\mathrm{T}} Q\mu \end{array}\right).
$$

Here, $(\xi^{\mathrm{T}}, 1)$ denotes the row vector of size $1 \times (|\mathcal{W}| + 1)$ with the last component equals 1 and the first $|\mathcal{W}|$ components are the components of ξ. The above inequality can be rewritten as

$$
x^{\mathrm{T}} U x \geq 0,\ \forall\ x \in \mathbb{R}_+^{|\mathcal{W}|+1}, ||x||_2 = 1.
$$

Using Proposition 5.1 in [11], we deduce that the constraint (ii) of (15) is equivalent to $U \in \mathrm{COP}^{|\mathcal{W}|+1}$. The constraint (iii) of (15) is equivalent to:

$$
-1 + (\xi^{\mathrm{T}}, 1) U (\xi^{\mathrm{T}}, 1)^{\mathrm{T}} \geq 0,\ \forall\ \xi \in \mathbb{R}_+^{|\mathcal{W}|},\ \chi^{\mathrm{T}}\xi \leq y. \tag{16}
$$

Define,

$$
\begin{cases}
s_{\mathrm{P}} = \min\limits_{\xi \in \mathbb{R}_+^{|\mathcal{W}|}} \max\limits_{\lambda \geq 0} & \mathcal{L}(\lambda, \xi, U, \chi, y). \\
s_{\mathrm{D}} = \max\limits_{\lambda \geq 0} \min\limits_{\xi \in \mathbb{R}_+^{|\mathcal{W}|}} & \mathcal{L}(\lambda, \xi, U, \chi, y).
\end{cases}
\tag{17}
$$

where $\mathcal{L}(\lambda, \xi, U, \chi, y) = -1 + (\xi^{\mathrm{T}}, 1)U(\xi^{\mathrm{T}}, 1)^{\mathrm{T}} + \lambda(\chi^{\mathrm{T}}\xi - y)$. In [4], the authors use the Sion's minimax theorem [24] to interchange the minimum and the maximum. However, since φ is not compact, we cannot apply the Sion's minimax theorem directly in this case. We show that φ can be restricted to a compact set without loss of optimality. For a given U and χ, we have

$$\begin{aligned} s_{\mathrm{P}} &\leq \max_{\lambda \geq 0} \mathcal{L}(\lambda, 0, U, \chi, y) \\ &= \max_{\lambda \geq 0}(-t - \mu^{\mathrm{T}} Q\mu - \lambda y - 1) = -t - \mu^{\mathrm{T}} Q\mu - 1 < \infty \end{aligned} \tag{18}$$

Therefore, using the min-max inequality $s_{\mathrm{D}} \leq s_{\mathrm{P}} < \infty$. Let $U_i = U + \frac{1}{2^i}\mathbf{I}_{|\mathcal{W}|+1}$ and $\chi_i = \chi + \frac{1}{2^i}\mathbf{1}$, for every $i \in \mathbb{N}$, where $\mathbf{I}_{|\mathcal{W}|+1}$ denotes the identity matrix of size $|\mathcal{W}| + 1$, $\mathbf{1}$ denotes the vector with all components equal to 1. It is clear from the construction that $\chi_i > 0$ componentwise. Since, $\mathcal{L}$ is a continuous function w.r.t U and χ, we have

$$\mathcal{L}(\lambda, \xi, U_i, \chi_i, y) \xrightarrow{i \to \infty} \mathcal{L}(\lambda, \xi, U, \chi, y), \ \forall\ \xi \in \mathbb{R}_+^{|\mathcal{W}|}, \lambda \geq 0.$$

Since, the min and max operators preserve the continuity, we have

$$\begin{aligned} \min_{\xi \in \mathbb{R}_+^{|\mathcal{W}|}} \max_{\lambda \geq 0} \quad \mathcal{L}(\lambda, \xi, U_i, \chi_i, y) &\xrightarrow{i \to \infty} \min_{\xi \in \mathbb{R}_+^{|\mathcal{W}|}} \max_{\lambda \geq 0} \quad \mathcal{L}(\lambda, \xi, U, \chi, y). \\ \max_{\lambda \geq 0} \min_{\xi \in \mathbb{R}_+^{|\mathcal{W}|}} \quad \mathcal{L}(\lambda, \xi, U_i, \chi_i, y) &\xrightarrow{i \to \infty} \max_{\lambda \geq 0} \min_{\xi \in \mathbb{R}_+^{|\mathcal{W}|}} \quad \mathcal{L}(\lambda, \xi, U, \chi, y). \end{aligned}$$

This implies that, if $s_{\mathrm{P}} = s_{\mathrm{D}}$ holds for any U_i, χ_i, $i \in \mathbb{N}$, it also holds for U, χ. For an arbitrary U_i and χ_i, let the the optimal solutions of minimax and maximin problems defined by (17) are $(\xi_{\mathrm{P}}, \lambda_{\mathrm{P}})$ and $(\xi_{\mathrm{D}}, \lambda_{\mathrm{D}})$, respectively. We prove that ξ_{P} and ξ_{D} are bounded, i.e., there exists $\Upsilon_{\mathrm{P}} > 0$ and $\Upsilon_{\mathrm{D}} > 0$ depending on U_i, χ_i and y such that $||\xi_{\mathrm{P}}||_2 \leq \Upsilon_{\mathrm{P}}$ and $||\xi_{\mathrm{D}}||_2 \leq \Upsilon_{\mathrm{D}}$. It is clear that $\lambda_{\mathrm{P}} = 0$ and $\chi_i^{\mathrm{T}}\xi_{\mathrm{P}} - y \leq 0$. Hence, we have

$$\begin{aligned} s_{\mathrm{P}} &= -1 + (\xi_{\mathrm{P}}^{\mathrm{T}}, 1)U_i(\xi_{\mathrm{P}}^{\mathrm{T}}, 1)^{\mathrm{T}}, \\ &= -1 + (\xi_{\mathrm{P}}^{\mathrm{T}}, 1)U(\xi_{\mathrm{P}}^{\mathrm{T}}, 1)^{\mathrm{T}} + \frac{1}{2^i}||\xi_{\mathrm{P}}||_2^2 + \frac{1}{2^i}. \end{aligned}$$

From constraint (ii) of (15), it follows that $(\xi_{\mathrm{P}}^{\mathrm{T}}, 1)U(\xi_{\mathrm{P}}^{\mathrm{T}}, 1)^{\mathrm{T}} \geq 0$. Therefore, if $||\xi_{\mathrm{P}}||_2 \to \infty$, $s_{\mathrm{P}} \to \infty$. Therefore, $||\xi_{\mathrm{P}}||_2$ is bounded by some real number $\Upsilon_P > 0$ which depends on U_i, χ_i and y. As $\xi \in \mathbb{R}_+^{|\mathcal{W}|}$ and $\chi_i > 0$, componentwise, we have

$$\liminf_{||\xi||_2 \to \infty} \lambda(\xi)(\chi_i^{\mathrm{T}}\xi - y) \geq 0,$$

for any $\lambda(\xi) \geq 0$. Then,

$$s_{\mathrm{D}} = -1 + (\xi_{\mathrm{D}}^{\mathrm{T}}, 1)U_i(\xi_{\mathrm{D}}^{\mathrm{T}}, 1)^{\mathrm{T}} + \lambda_{\mathrm{D}}(\chi_i^{\mathrm{T}}\xi_{\mathrm{D}} - y),$$

$$= -1 + (\xi_{\mathrm{D}}^{\mathrm{T}}, 1)U(\xi_{\mathrm{D}}^{\mathrm{T}}, 1)^{\mathrm{T}} + \frac{1}{2^i}||\xi_{\mathrm{D}}||_2^2 + \frac{1}{2^i} + \lambda_{\mathrm{D}}(\chi_i^{\mathrm{T}}\xi_{\mathrm{D}} - y).$$

It is clear that $\frac{1}{2^i}||\xi_{\mathrm{D}}||_2^2 \to \infty$ and the other terms are lower bounded by some nonnegative number. Therefore, $s_{\mathrm{D}} \to \infty$ when $||\xi_{\mathrm{D}}||_2 \to \infty$. Hence, $||\xi_{\mathrm{D}}||_2$ is bounded by some real number $\Upsilon_D > 0$ which depends on U_i, χ_i and y. Let $\Upsilon = \max(\Upsilon_{\mathrm{P}}, \Upsilon_{\mathrm{D}})$. Then, (17) is equivalent to

$$\begin{aligned} s_{\mathrm{P}} &= \min_{\xi \in \mathbb{R}_+^{|\mathcal{W}|}, ||\xi||_2 \leq \Upsilon} \max_{\lambda \geq 0} \quad \mathcal{L}(\lambda, \xi, U^i, \chi^i, y). \\ s_{\mathrm{D}} &= \max_{\lambda \geq 0} \min_{\xi \in \mathbb{R}_+^{|\mathcal{W}|}, ||\xi||_2 \leq \Upsilon} \quad \mathcal{L}(\lambda, \xi, U^i, \chi^i, y). \end{aligned}$$

Note that the set $\left\{\xi \mid \xi \in \mathbb{R}_+^{|\mathcal{W}|}, ||\xi||_2 \leq \Upsilon\right\}$ is compact. Therefore, from Sion's minimax theorem $s_{\mathrm{P}} = s_{\mathrm{D}}$ for every U_i, χ_i, $i \in \mathbb{N}$. For any ξ such that $\chi^{\mathrm{T}}\xi > y$, it is easy to see that

$$\max_{\lambda \geq 0} \quad \mathcal{L}(\lambda, \xi, U, \chi, y) = \infty$$

The condition $s_{\mathrm{P}} < \infty$ gives $\chi^{\mathrm{T}}\xi \leq y$ and $\lambda = 0$ which in turn implies that

$$s_{\mathrm{P}} = \min_{\chi^{\mathrm{T}}\xi \leq y} \quad \mathcal{L}(0, \xi, U, \chi, y) \geq 0.$$

Therefore, (16) is equivalent to $s_{\mathrm{D}} \geq 0$. Then, there exists a sequence of nonnegative numbers $\lambda_j \geq 0$ and a decreasing sequence of positive numbers $\theta_j > 0$, such that $\theta_j \to 0$ as $j \to \infty$, for which the following condition holds

$$\begin{cases} -1 + (\xi^{\mathrm{T}}, 1)U(\xi^{\mathrm{T}}, 1)^{\mathrm{T}} + \lambda_j(\chi^{\mathrm{T}}\xi - y) \geq -\theta_j, \ \forall\ \xi \in \mathbb{R}_+^{|\mathcal{W}|}, \ j \in \mathbb{N}, \\ \lambda_j \geq 0, \ \forall\ j \in \mathbb{N}. \end{cases} \tag{19}$$

For each $j \in \mathbb{N}$, define

$$Fea(\theta_j) = \{(U, \chi, y, \lambda) \mid -1 + (\xi^{\mathrm{T}}, 1)U(\xi^{\mathrm{T}}, 1)^{\mathrm{T}} + \lambda(\chi^{\mathrm{T}}\xi - y) \geq -\theta_j, \ \lambda \geq 0\}.$$

The feasible region defined by (19) is equivalent to $\bigcap_{j \in \mathbb{N}} Fea(\theta_j)$. For any $i < j$, $Fea(\theta_j) \subset Fea(\theta_i)$. Therefore, $Fea(\theta_j) \downarrow \bigcap_{i \in \mathbb{N}} Fea(\theta_i)$ as $j \to \infty$. The feasible set $Fea(\theta_j)$ as $j \to \infty$ is given by

$$\begin{cases} (\xi^{\mathrm{T}}, 1)Z(\xi^{\mathrm{T}}, 1)^{\mathrm{T}} \geq 0, \ \forall\ \xi \in \mathbb{R}_+^{|\mathcal{W}|}, \\ \lambda \geq 0, \end{cases} \tag{20}$$

where $Z \in \mathcal{S}^{|\mathcal{W}|+1}$ and $Z = \left(\begin{array}{c|c} -Q & -\frac{1}{2}q + Q\mu + \lambda\chi \\ \hline -\frac{1}{2}q^{\mathrm{T}} + \mu^{\mathrm{T}}Q + \lambda\chi^{\mathrm{T}} & -t - \mu^{\mathrm{T}}Q\mu - 1 - \lambda y \end{array}\right)$. Using similar arguments as above, the constraint (20) is equivalent to

$$Z \in \mathrm{COP}^{|\mathcal{W}|+1},\ \lambda \geq 0. \tag{21}$$

This implies that the constraint (iii) of (15) is equivalent to (21). Hence, DRCCMDP problem (6) is equivalent to the following optimization problem

$$\begin{array}{ll} & \max \quad y \\ \text{s.t.} & (i)\ -t - Q \circ \Sigma - q^T\mu \leq \epsilon, \\ & (ii)\ \left(\begin{array}{c|c} -Q & -\frac{1}{2}q + Q\mu \\ \hline -\frac{1}{2}q^T + \mu^T Q & -t - \mu^T Q\mu \end{array}\right) \in COP^{|\mathcal{W}|+1}, \\ & (iii)\ \left(\begin{array}{c|c} -Q & -\frac{1}{2}q + Q\mu + \lambda\chi \\ \hline -\frac{1}{2}q^T + \mu^T Q + \lambda\chi^T & -t - \mu^T Q\mu - 1 - \lambda y \end{array}\right) \in COP^{|\mathcal{W}|+1}, \\ & (iv)\ Q \in \mathcal{S}^{|\mathcal{W}|}, \lambda \geq 0,\ \chi \in \mathcal{Q}_\alpha(\varrho). \end{array} \tag{22}$$

If $\lambda = 0$ is an optimal solution of (22), it is clear that either (22) is infeasible or its optimal value is ∞, which leads to contradiction. Then, we can restrict to $\lambda > 0$. Let $\bar{t} = \frac{t}{\lambda},\ \bar{Q} = \frac{Q}{\lambda},\ \bar{q} = \frac{q}{\lambda},\ s = \frac{1}{\lambda}$, we deduce that (6) is equivalent to (11).

2. Let the distribution of $\hat{R}$ belongs to the uncertainty set $\mathcal{D}_2(\varphi, \mu, \Sigma, \delta_0)$. From Theorem 3.4 [4], the dual of the optimization problem $\sup_{F\in\mathcal{D}} \mathbb{P}_F\left(\chi^{\mathrm{T}}\hat{R} \leq y\right)$ can be written as

$$\begin{array}{ll} & \inf \quad (-t - \mu^{\mathrm{T}}q - \mu^{\mathrm{T}}Q\mu + \delta_0\Sigma \circ Q) \\ \text{s.t.} & \text{(i)}\ \mathbf{1}_{\{\chi^{\mathrm{T}}\xi \leq y\}} + t + q^{\mathrm{T}}\xi - \xi^{\mathrm{T}}Q\xi + 2\mu^{\mathrm{T}}Q\xi \leq 0,\ \forall\ \xi \in \mathbb{R}_+^{|\mathcal{W}|}, \\ & \text{(ii)}\ Q \in \mathcal{S}_+^{|\mathcal{W}|}, \end{array}$$

and the strong duality holds. We follow the same arguments used for the case of the uncertainty set $\mathcal{D}_1(\varphi, \mu, \Sigma)$.

3. If the distribution of $\hat{R}$ belongs to the uncertainty set $\mathcal{D}_3(\varphi, \mu, \Sigma, \delta_1, \delta_2)$, using Lemma 1 of [6] the dual of the problem $\sup_{F\in\mathcal{D}} \mathbb{P}_F\left(\chi^{\mathrm{T}}\hat{R} \leq y\right)$ is given by

$$\begin{array}{ll} & \inf \quad (r + t) \\ \text{s.t.} & \text{(i)}\ r \geq \mathbf{1}_{\{\chi^{\mathrm{T}}\xi \leq y\}} - \xi^{\mathrm{T}}Q\xi - \xi^{\mathrm{T}}q,\ \forall\ \xi \in \mathbb{R}_+^{|\mathcal{W}|}, \\ & \text{(ii)}\ t \geq (\delta_2\Sigma + \mu\chi^{\mathrm{T}}) \circ Q + \chi^{\mathrm{T}}q + \sqrt{\delta_1}||\Sigma^{\frac{1}{2}}(q + 2Q\mu)||_2, \\ & \text{(iii)}\ Q \in \mathcal{S}_+^{|\mathcal{W}|}, \end{array}$$

and strong duality holds. Again, the rest of the proof follows using similar arguments used in the case of $\mathcal{D}_1(\varphi, \mu, \Sigma)$.

□

Remark 1. *The optimization problems (11)-(13) are similar as they have the same objective function and same copositive constraints of the form $A \in COP^n$, for some n. To the best of our knowledge, there is no commercial solver, which can solve optimization problems with copositive constraints. We propose relaxing the copositive constraints by an SDP constraint and a linear constraint, i.e., $A \succ 0$ and $A \geq 0$, componentwise. It is clear that these two cases are special cases of the copositive constraint, that leads to an outer approximation of the copositive problem (lower bound).*

4 Uncertainty sets based on statistical distance

4.1 Uncertainty set with ϕ -divergence distance

We consider an uncertainty set defined using statistical distance metric called ϕ-divergence. In such uncertainty set, a nominal distribution is known to the decision maker relied on the available estimated data. The decision maker thinks that the true distribution of $\hat{R}$ belongs to a ball of radius θ_ϕ and centered at a nominal distribution ν and the distance between the true distribution and ν is given by a ϕ-divergence. We show that the DRCCMDP problem (5) is equivalent to an SOCP problem for various ϕ-divergences.

Definition 1. *The $\phi-$divergence distance between two probability measures ν_1 and ν_2 with densities f_{ν_1} and f_{ν_2}, respectively, and full support $\mathbb{R}^{|\mathcal{W}|}$ is given by*

$$I_\phi(\nu_1, \nu_2) = \int_{\mathbb{R}^{|\mathcal{W}|}} \phi\left(\frac{f_{\nu_1}(\xi)}{f_{\nu_2}(\xi)}\right) f_{\nu_2}(\xi) d\xi.$$

For different choices of ϕ, we refer to [2] and [19]. Let $\nu \in \mathcal{M}^+$ be a nominal distribution with a density function f_ν. The uncertainty set of the distribution of $\hat{R}$ based on ϕ-divergence is defined by

$$\mathcal{D}_4(\nu, \theta_\phi) = \left\{F \in \mathcal{M}^+ \mid I_\phi(F, \nu) \leq \theta_\phi\right\}, \tag{23}$$

where $\theta_\phi > 0$.

Definition 2. *A function $\phi^* : \mathbb{R} \to \mathbb{R} \cup \infty$ is the conjugate of ϕ, where*

$$\phi^*(r) = \sup_{t \geq 0} \left\{rt - \phi(t)\right\}, \ \forall\ r \in \mathbb{R}.$$

Lemma 2. *Consider an optimization problem*

$$\inf_{F \in \mathcal{D}_4(\nu, \theta_\phi)} \mathbb{P}_F(\chi^{\mathrm{T}} \hat{R} \geq y). \tag{24}$$

Then, the dual problem of (24) is given by

$$\sup_{\lambda > 0, \beta \in \mathbb{R}} \left\{\beta - \lambda\theta_\phi - \lambda\phi^*\left(\frac{-1+\beta}{\lambda}\right)\mathbb{P}_\nu(O) - \lambda\phi^*\left(\frac{\beta}{\lambda}\right)(1 - \mathbb{P}_\nu(O))\right\},$$

where $O = \{\xi \in \mathbb{R}^{|\mathcal{W}|} \mid \chi^{\mathrm{T}}\xi \geq y\}$, *such that the strong duality holds.*

Proof. We rewrite the primal problem (24) as a following semi-infinite programming problem

$$\begin{aligned} v_{\mathrm{P}} = \inf_{F\geq 0} & \int_{\mathbb{R}^{|\mathcal{W}|}} \mathbf{1}_O(\xi)F(\xi)\mathrm{d}\xi \\ \text{s.t.} \quad \text{(i)} & \int_{\mathbb{R}^{|\mathcal{W}|}} f_\nu(\xi)\phi\left(\frac{F(\xi)}{f_\nu(\xi)}\right)\mathrm{d}\xi \leq \theta_\phi, \text{ (ii)} \int_{\mathbb{R}^{|\mathcal{W}|}} F(\xi)\mathrm{d}\xi = 1. \end{aligned} \tag{25}$$

The dual problem of (25) is given by

$$v_{\mathrm{D}} = \sup_{\lambda\geq 0,\beta\in\mathbb{R}} \left\{\beta - \lambda\theta_\phi + \inf_{F(\xi)\geq 0}\left\{\int_{\mathbb{R}^{|\mathcal{W}|}} \left(\mathbf{1}_O(\xi)F(\xi) - \beta F(\xi) + \lambda f_\nu(\xi)\phi\left(\frac{F(\xi)}{f_\nu(\xi)}\right)\right)\mathrm{d}\xi\right\}\right\},$$

where λ is the dual variable of the constraint (i) of (25) and β is the dual variable of the constraint (ii) of (25). Since $\theta_\phi > 0$, the Slater's condition holds which implies that the strong duality holds, i.e., $v_{\mathrm{P}} = v_{\mathrm{D}}$. The rest of the proof follows from Theorem 1 of [13]. □

We study 4 cases of $\phi-$divergences whose conjugates are given in Table 2. Using

Table 2 List of selected $\phi-$divergences with their conjugate

Divergence	$\phi(u), u \geq 0$	$\phi^*(v)$
Kullback-Leibler	$u\log(u) - u + 1.$	$\mathrm{e}^v - 1$
Variation distance	$\lvert u-1\rvert.$	$\begin{cases} -1, & v \leq -1, \\ v, & -1 \leq v \leq 1, \\ \infty, & v > 1. \end{cases}$
Modified χ^2 - distance	$(u-1)^2.$	$\begin{cases} -1, & v \leq -2, \\ v + \frac{v^2}{4}, & v > -2. \end{cases}$
Hellinger distance	$(\sqrt{u}-1)^2.$	$\begin{cases} \frac{v}{1-v}, & v < 1, \\ \infty, & v \geq 1. \end{cases}$

Lemma 2, the following result holds.

Theorem 3. *Consider the DRCCMDP problem* (5)*, where the uncertainty set is defined by* (23) *for the* ϕ*-divergences listed in Table 3. If the reference distribution* ν *is a normal distribution with mean vector* μ_ν *and positive definite covariance matrix* Σ_ν*, the DRCCMDP problem* (5) *is equivalent to the following* SOCP *problem*

$$\begin{aligned} & \max \quad y \\ s.t. \quad & \text{(i) } \chi^{\mathrm{T}}\mu_\nu - \Phi^{(-1)}[f(\theta_\phi,\epsilon)]\|\Sigma_\nu^{\frac{1}{2}}\chi\|_2 \geq y, \text{ (ii) } \chi \in \mathcal{Q}_\alpha(\varrho), \end{aligned} \tag{26}$$

where $\Phi^{(-1)}$ is the quantile of the standard normal distribution and the values of θ_ϕ, ϵ and $f(\theta_\phi, \epsilon)$ for different choices of divergences are given in Table 3.

Table 3 The function f for selected $\phi-$divergences

Divergence	$f(\theta_\phi, \epsilon)$	θ_ϕ, ϵ
K-L	$\inf_{x\in(0,1)} \frac{e^{-\theta_\phi x^{1-\epsilon}-1}}{x-1}$	$\theta_\phi > 0$, $0 < \epsilon < 1$
Variation	$1-\epsilon+\frac{\theta_\phi}{2}$	$\theta_\phi > 0$, $0 < \epsilon < 1$
Modified χ^2	$1-\epsilon+\frac{\sqrt{\theta_\phi^2+4\theta_\phi(\epsilon-\epsilon^2)}-(1-2\epsilon)\theta_\phi}{2\theta_\phi+2}$	$\theta_\phi > 0$, $0 < \epsilon < \frac{1}{2}$
Hellinger	$\frac{-B+\sqrt{\Delta}}{2}$, where $B = -(2-(2-\theta_\phi)^2)\epsilon - \frac{(2-\theta_\phi)^2}{2}$, $C = \left(\frac{(2-\theta_\phi)^2}{4}-\epsilon\right)^2$, $\Delta = B^2-4C = (2-\theta_\phi)^2\left[4-(2-\theta_\phi)^2\right]\epsilon(1-\epsilon)$.	$0 < \theta_\phi < 2-\sqrt{2}$, $0 < \epsilon < 1$

Proof. Using Lemma 2, we prove that the constraint (i) of (5) is equivalent to the following constraint

$$\mathbb{P}_\nu(\chi^{\mathrm{T}}\hat{R} \geq y) \geq f(\theta_\phi, \epsilon). \tag{27}$$

Since ν is a normal distribution with mean vector μ_ν and covariance matrix Σ_ν, it is well known that (27) is equivalent to the constraint (i) of (26). The Hellinger distance case proof is given in [17], where the Kullback-Leibler, Variation distance and Modified χ^2 - distance ones follow from Propositions 2, 3 and 4 of [13]. □

4.2 Uncertainty set with Wasserstein distance

We consider an uncertainty set defined using statistical distance metric called Wasserstein distance. We show that the DRCCMDP problem (6) is tractable if the reference distribution ν follows a discrete distribution whose scenarios are taken from historical data. We refer to Villani [27, 28] for more details of the Wasserstein distance metric.

Let φ be a closed, convex subset of $\mathbb{R}^{|\mathcal{W}|}$ and $p \in [1, \infty)$. Let $\mathcal{B}(\varphi)$ denotes the Borel $\sigma-$ algebra on φ. Let $\mathcal{P}(\varphi)$ be the set of all probability measures defined on $\mathcal{B}(\varphi)$ and $\mathcal{P}_p(\varphi)$ denote the subset of $\mathcal{P}(\varphi)$ with finite $p-$ moment and it is defined as

$$\mathcal{P}_p(\varphi) = \left\{\mu \in \mathcal{P}(\varphi) \mid \int_{\xi\in\varphi} ||\xi - \xi_0||_2^p \mu(\mathrm{d}\xi) < \infty \text{ for some } \xi_0 \in \varphi\right\}.$$

It follows from the triangle inequality that the above definition of $\mathcal{P}_p(\varphi)$ does not depend on ξ_0.

Definition 3 (Wasserstein distance)**.** *The Wasserstein distance* $W_p(\mu,\nu)$ *between* $\nu_1,\nu_2 \in \mathcal{P}_p(\varphi)$ *is defined by*

$$W_p(\nu_1,\nu_2) = \left(\inf_{\varrho\in\mathcal{P}_{\nu_1,\nu_2}(\varphi\times\varphi)} \int_{\varphi\times\varphi} ||x-z||_2^p \varrho(dx,dz)\right)^{\frac{1}{p}},$$

where $\mathcal{P}_{\nu_1,\nu_2}(\varphi\times\varphi)$ *denotes the set of all probability measures defined on* $\mathcal{B}(\varphi\times\varphi)$ *such that the marginal laws are* ν_1 *and* ν_2*.*

The uncertainty set using Wasserstein distance is defined by

$$\mathcal{D}_5(\varphi,\nu,p,\theta_W) = \{F \in \mathcal{P}_p(\varphi) \mid W_p(F,\nu) \le \theta_W\}, \tag{28}$$

where $\nu \in \mathcal{P}_p(\varphi)$ and $\theta_W > 0$.

Lemma 3. *Consider an optimization problem*

$$\sup_{F\in\mathcal{D}_5(\varphi,\nu,p,\theta_W)} \mathbb{P}_F(\chi^{\mathrm{T}}\hat{R} \le y). \tag{29}$$

Then, the dual problem of (29) *is given by*

$$\inf_{\lambda\ge 0}\left\{\lambda\theta_W^p - \int_\varphi \inf_{z\in\varphi}\left[\lambda||x-z||_2^p - \mathbf{1}_{\{\chi^{\mathrm{T}}z\le y\}}\right]\nu(\mathrm{d}x)\right\}, \tag{30}$$

such that the strong duality holds and the optimal values of the primal and the dual problems are finite.

Proof. Let Ξ be a Polish space with metric d, $\mathcal{P}(\Xi)$ be the set of Borel probability measures on Ξ, $\nu \in \mathcal{P}(\Xi)$ and $\Psi \in L^1(\nu)$, where $L^1(\nu)$ represents the L^1 space of ν - measurable functions. It follows from Theorem 1 of [8] that the following strong duality holds

$$\begin{aligned}&\sup_{\mu\in\mathcal{P}(\Xi)}\left\{\int_\Xi \Psi(\xi)\mu(d\xi) \mid W_p(\mu,\nu)\le\theta_W\right\}\\ &= \inf_{\lambda\in\mathbb{R},\lambda\ge 0}\left\{\lambda\theta_W^p - \int_\Xi \inf_{\xi\in\Xi}\left[\lambda d^p(\xi,\zeta) - \Psi(\xi)\right]\nu(\mathrm{d}\zeta)\right\} < \infty,\end{aligned} \tag{31}$$

provided the growth factor given by Definition 4 of [8] is finite. We apply this result in our case by choosing $\Xi = \varphi$, d as an Euclidean metric and $\Psi(\xi) = \mathbf{1}_{\{\chi^{\mathrm{T}}\xi\le y\}}$ for all $\xi\in\varphi$. For this choice of $\Psi(\xi)$, it is easy to see from Definition 4 of [8] that the growth factor is zero. Since $\left\{\xi\in\varphi \mid \chi^{\mathrm{T}}\xi \le y\right\}$ is a closed set, it is a Borel measurable set. Hence, it is clear that $\Psi\in L^1(\nu)$ for all $\nu\in\mathcal{P}(\varphi)$. Then, (31) reduces to

$$\sup_{F\in\mathcal{D}_5(\varphi,\nu,p,\theta_W)} \mathbb{P}_F\left(\chi^{\mathrm{T}}\hat{R}\le y\right) = \inf_{\lambda\ge 0}\left\{\lambda\theta_W^p - \int_\varphi \inf_{\xi\in\varphi}\left[\lambda||\zeta-\xi||_2^p - \mathbf{1}_{\{\chi^{\mathrm{T}}\xi\le y\}}\right]\nu(\mathrm{d}\zeta)\right\}.$$

□

We consider the case when $p = 1$ and ν is a data-driven reference distribution, i.e., it is a discrete distribution with H scenarios $\tilde{\xi}_1, \ldots, \tilde{\xi}_H$, where $\tilde{\xi}_i \in \varphi$, for every $i = 1, \ldots, H$. Using Lemma 3, we propose a deterministic reformulation of the DRCCMDP problem (6).

Lemma 4. *If the distribution of $\hat{R}$ belongs to the uncertainty set defined by (28), the* DRCCMDP (6) *can be reformulated equivalently as the following deterministic problem*

$$\begin{aligned} & \sup \quad y \\ s.t. \quad & \text{(i) } \theta_W - \frac{1}{H}\sum_{i=1}^{H} g_i \leq l\epsilon, \\ & \text{(ii) } \inf_{z\in\varphi, \chi^{\mathrm{T}} z \leq y} ||\tilde{\xi}_i - z||_2 \geq l + g_i, \ \forall\ i = 1, \ldots, H, \\ & \text{(iii) } l > 0, \ \chi \in \mathcal{Q}_\alpha(\varrho), \ g_i \leq 0, \ \forall\ i = 1, \ldots, H. \end{aligned} \tag{32}$$

Proof. Using Lemma 3, since ν is a discrete distribution with H scenarios $\tilde{\xi}_1, ..., \tilde{\xi}_H$, the constraint (i) of (6) can be equivalently written as

$$\lambda\theta_W - \frac{1}{H}\sum_{i=1}^{H} \inf_{z\in\varphi} \left[\lambda||\tilde{\xi}_i - z||_2 - \mathbf{1}_{\{\chi^{\mathrm{T}} z \leq y\}}\right] \leq \epsilon, \ \lambda \geq 0.$$

By introducing auxiliary variables t_i, $i = 1, ..., H$, the above constraint can be rewritten as

$$\begin{cases} \text{(i) } \lambda\theta_W - \frac{1}{H}\sum_{i=1}^{H} t_i \leq \epsilon, \ \lambda \geq 0 \\ \text{(ii) } \inf_{z\in\varphi} \left[\lambda||\tilde{\xi}_i - z||_2 - \mathbf{1}_{\{\chi^{\mathrm{T}} z \leq y\}}\right] \geq t_i, \ \forall\ i = 1, \ldots, H. \end{cases} \tag{33}$$

The constraint (ii) of (33) is equivalent to the following two constraints

$$\begin{cases} \text{(i) } \inf_{z\in\varphi} \lambda||\tilde{\xi}_i - z||_2 \geq t_i, \ \forall\ i = 1, \ldots, H, \\ \text{(ii) } \inf_{z\in\varphi, \chi^{\mathrm{T}} z \leq y} \lambda||\tilde{\xi}_i - z||_2 - 1 \geq t_i, \ \forall\ i = 1, \ldots, H. \end{cases} \tag{34}$$

Since $\lambda \geq 0$, $\inf_{z\in\varphi} \lambda||\tilde{\xi}_i - z||_2 = 0$. Then, the constraint (i) of (34) is equivalent to $t_i \leq 0$, for every $i = 1, \ldots, H$. Moreover, if $\lambda = 0$, from the constraint (ii) of (34), $t_i \leq -1$, for every $i = 1, \ldots, H$, which in turn implies $-\frac{1}{H}\sum_{i=1}^{H} t_i \geq 1$. This violates the constraint (i) of (33). Hence, $\lambda > 0$. Let $l = \frac{1}{\lambda}$ and $g_i = \frac{t_i}{\lambda}$, for every $i = 1, \ldots, H$. Therefore, the constraint (i) of (6) is equivalent to the following constraints

$$\begin{cases} \text{(i) } \theta_W - \frac{1}{H}\sum_{i=1}^{H} g_i \leq l\epsilon, \\ \text{(ii) } \inf_{z\in\varphi, \chi^{\mathrm{T}} z \leq y} ||\tilde{\xi}_i - z||_2 \geq l + g_i, \ \forall\ i = 1, \ldots, H, \\ \text{(iii) } l > 0, \ g_i \leq 0, \ \forall\ i = 1, \ldots, H. \end{cases} \tag{35}$$

This implies that the DRCCMDP (6) is equivalent to (32). □

The constraint (ii) of (32) includes inf term which makes it difficult to solve the problem directly. We show that the optimization problem (32) is equivalent to a MISOCP problem and a biconvex optimization problem for the case of full support and nonnegative support, respectively.

4.2.1 DRCCMDP under Wasserstein distance based uncertainty set with full support

Theorem 4. *Consider the* DRCCMDP *problem* (6)*. We assume that the distribution of* $\hat{R}$ *belongs to the uncertainty set defined by* (28) *and* $\varphi = \mathbb{R}^{|\mathcal{W}|}$*. Then, the* DRCCMDP *problem* (6) *is equivalent to the following MISOCP*

$$
\begin{aligned}
& \max \quad y \\
s.t. \quad & \text{(i) } \beta\theta_W - \frac{1}{H}\sum_{i=1}^{H} b_i \leq t\epsilon, \\
& \text{(ii) } M\eta_i \geq b_i + t, \ \forall\ i = 1, \ldots, H, \\
& \text{(iii) } M(1-\eta_i) + \chi^{\mathrm{T}}\tilde{\xi}_i - y \geq b_i + t, \ \forall\ i = 1, \ldots, H, \\
& \text{(iv) } \eta_i \in \{0,1\}, \ \forall\ i = 1, \ldots, H, \\
& \text{(v) } ||\chi||_2 \leq \beta, t \geq 0, \beta > 0, \chi \in \mathcal{Q}_\alpha(\varrho), b_i \leq 0, \ \forall\ i = 1, \ldots, H.
\end{aligned} \tag{36}
$$

Proof. The proof follows from Theorem 2 [31]. □

4.2.2 DRCCMDP under Wasserstein distance based uncertainty set with nonnegative support

Lemma 5. *Let* $\varphi = \mathbb{R}_+^{|\mathcal{W}|}$ *and consider an optimization problem*

$$
\inf_{z \in \varphi, \chi^{\mathrm{T}} z \leq y} ||\tilde{\xi}_i - z||_2. \tag{37}
$$

The dual problem of (37) *is given by*

$$
\begin{aligned}
\max \quad & \lambda_i(\chi^{\mathrm{T}}\tilde{\xi}_i - y) - \zeta_i^{\mathrm{T}}\tilde{\xi}_i \\
s.t. \quad & ||\zeta_i - \lambda_i\chi||_2 \leq 1, \ \zeta_i \in \mathbb{R}_+^{|\mathcal{W}|}, \lambda_i \geq 0,
\end{aligned}
$$

such that the strong duality holds.

Proof. The proof is given in Appendix B. □

Theorem 5. *Consider the* DRCCMDP *problem* (6)*. We assume that the distribution of* $\hat{R}$ *belongs to the uncertainty set defined by* (28) *and* $\varphi = \mathbb{R}_+^{|\mathcal{W}|}$*. Then, the* DRCCMDP (6) *is equivalent to the following biconvex optimization problem*

$$
\max \quad y
$$

$$
\begin{aligned}
s.t. \quad & \text{(i) } \theta_W - \frac{1}{H}\sum_{i=1}^{H} g_i \leq l\epsilon, \\
& \text{(ii) } \lambda_i(\chi^{\mathrm{T}}\tilde{\xi}_i - y) - \zeta_i^{\mathrm{T}}\tilde{\xi}_i \geq l + g_i, \ \forall\ i = 1, \ldots, H, \\
& \text{(iii) } ||\zeta_i - \lambda_i\chi||_2 \leq 1, \ \forall\ i = 1, \ldots, H, \\
& \text{(iv) } \lambda_i \geq 0, \zeta_i \in \mathbb{R}_+^{|\mathcal{W}|}, \ l > 0, \ g_i \leq 0, \ \chi \in \mathcal{Q}_\alpha(\varrho), \ \forall\ i = 1, \ldots, H.
\end{aligned} \tag{38}
$$

Proof. The proof follows directly from Lemma 4 and Lemma 5. □

5 Machine replacement problem

In this section, we address a machine replacement problem which involves a factory machine with a lifespan of N years. At each stage, maintenance is scheduled for the machine , and the factory owner must decide whether to repair it or leave it unrepaired. If repaired, the machine might function as a new one, while not repairing it reduces its remaining lifespan by one year. Repairing the machine incurs a maintenance cost, which can be modeled as an MDP problem with the machine's lifespan representing the states of the underlying Markov chain, resulting in a total of $N + 1$ states. The initial state corresponds to a new machine, and at each state, there are two actions: i) "repair" and ii) "do not repair". The transition probabilities of the Markov chain corresponding to each action, are detailed in Figure 5. The maintenance cost for each state-action pair is uncertain and realized after the decision is decided. For every state-action pair (s, a), the maintenance cost is defined as $\hat{c}(s, a) = K + \hat{Z}(s, a)$, with K representing a fixed cost and $\hat{Z}(s, a)$ representing a random variable for the variable cost. Additionally, the machine generates revenue $L(s, a)$ for each state-action pair (s, a) and the profit for each $(s, a) \in \mathcal{W}$ is given by

$$
\hat{R}(s, a) = L(s, a) - K - \hat{Z}(s, a). \tag{39}
$$

The factory owner aims to maximize the expected discounted profit. We assume that the factory owner possesses a finite number of identical machines, all of which are modeled using the same Markov chain. Consequently, we calculate the optimal repair policy for a single machine, and this identical repair policy can be applied to all other machines in the fleet.

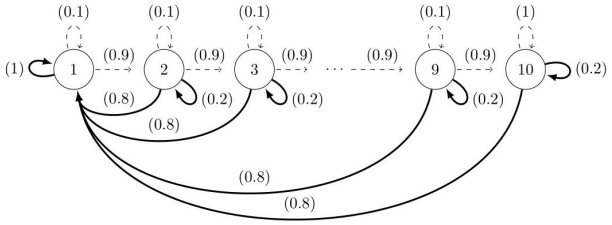

Fig. 1 Machine replacement with 10 states and 2 actions: "repair" (with continuous lines) and "no repair" (with dotted lines)

We compare the performance of DRCCMDP for each uncertainty set with the CCMDP model (4) where the distribution of $\hat{R}$ is assumed to be a normal distribution. In our numerical experiments, we consider a set of 10 states, $\epsilon = 0.15$, the discount parameter $\alpha = 0.8$ and the initial distribution of states ϱ to be uniformly distributed. For the above instance, $|\mathcal{W}| = 20$ and $\hat{R}$ is a 20×1 random vector with mean vector μ given by

$$\mu(s,a) = L(s,a) - K - \mu_{\hat{Z}}(s,a), \tag{40}$$

where $\mu_{\hat{Z}}$ is the mean vector of the random cost vector $\hat{Z}$. We take $K = 10$, the function L and the mean cost $\mu_{\hat{Z}}$ corresponding to each state-action pair are summarized in Table 4. The covariance matrix Σ of $\hat{R}$ is randomly generated using the following formula

$$\Sigma = \frac{AA^{\mathrm{T}}}{20} + D_{20}, \tag{41}$$

where A is a 20×20 random matrix whose all the entries are real numbers belonging to $[0,1]$, and D_{20} is a 20×20 diagonal matrix with $D_{20}(10,10) = 4$, $D_{20}(20,20) = 9$, $D_{20}(i,i) = 1$, for every $i \neq 10, 20$ and all other entries equal to zero. We use the above μ and Σ for all the moments based uncertainty sets. For ϕ-divergence based uncertainty

Table 4 Random cost $\hat{Z}$ and Revenue L

State(s) \ Action(a)	"Repair" $\mu_{\hat{Z}}(s,1)$	"Do not repair" $\mu_{\hat{Z}}(s,2)$	"Repair" $L(s,1)$	"Do not repair" $L(s,2)$
1	10.1	0	30	29.9
2	10.2	0	30	29.8
3	10.3	0	30	29.7
4	10.4	0	30	29.6
5	10.5	0	30	29.5
6	10.6	0	30	29.4
7	10.7	0	30	29.3
8	10.8	0	30	29.2
9	10.9	0	30	29.1
10	11	5	30	29.0

Table 5 Other parameters

K-U	$\delta_0 = 0.9$
U-U	$\delta_1 = \delta_2 = 1$
$\phi-$divergence	$\theta_\phi = 0.01$
Wasserstein	$\theta_W = 0.01$ $H = 1000$

set, we take the nominal distribution ν as a normal distribution with mean $\mu_\nu = \mu$ and covariance matrix $\Sigma_\nu = \Sigma$ where μ and Σ are defined by (40) and (41), respectively. For Wasserstein distance based uncertainty set, we take the number of observations $H = 1000$. The scenarios $(\tilde{\xi}_i)_{i=1}^H$ are randomly generated by taking $\tilde{\xi}_i = Bx + \mu_\nu$, where x is a standard normal vector, μ_ν is defined by (40) and B is the Cholesky factorization of Σ_ν defined by (41). We summarize the other parameters in Table 5.

We compute an optimal policy of the CCMDP problem (4), where $\hat{R}$ follows a normal distribution with mean vector and covariance matrix defined by (40) and (41),

Table 6 Optimal policies of *CCMDP* and *DRCCMDP* with full and nonnegative supports

Policies / State(s)	CCMDP Normal (1-q,q)	Full support K-K (1-q,q)	Full support K-U (1-q,q)	Full support U-U (1-q,q)	Modified χ^2 (1-q,q)	variation (1-q,q)
1	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
2	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
3	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
4	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
5	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
6	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
7	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
8	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
9	(1, 0)	(0.36, 0.64)	(0.36, 0.64)	(0.3, 0.7)	(0.73, 0.27)	(0.95, 0.05)
10	(0.1, 0.9)	(0.09, 0.91)	(0.09, 0.91)	(0.09, 0.91)	(0.1, 0.9)	(0.1, 0.9)
Optimal discounted reward	89.3	74	74	71.9	84.2	88.1

Table 7 Optimal policies of *CCMDP* and *DRCCMDP* with full and nonnegative supports (continued)

Kullbach-Leibler (1-q,q)	Hellinger (1-q,q)	Full support Wasserstein (1-q,q)	Nonnegative K-K (1-q,q)	Nonnegative K-U (1-q,q)	Nonnegative U-U (1-q,q)	Nonnegative Wasserstein (1-q,q)
(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)	(1, 0)
(0.75, 0.25)	(0.72, 0.28)	(0.98, 0.02)	(0.39, 0.61)	(0.39, 0.61)	(0.34, 0.66)	(0.99, 0.01)
(0.1, 0.9)	(0.1, 0.9)	(0.1, 0.9)	(0.09, 0.91)	(0.09, 0.91)	(0.09, 0.91)	(0.1, 0.9)
85	84	88.9	75.3	75.3	73.1	89.1

by solving an equivalent SOCP problem [5]. The optimal policies of the DRCCMDP problem for all the uncertainty sets are computed by solving the proposed equivalent optimization problems. We present the optimal policies of CCMDP and DRCCMDP with full support and nonnegative support in Tables 6 and 7, where q is the probability of "repair" and $1-q$ is the probability of "no repair". It is clear from Tables 6 and 7 that the optimal repair policy corresponding to all the uncertainty sets for first eight states is same. At state 9 the probability of repair is greater than the probability of do not repair for moments based uncertainty sets whereas for statistical distance based uncertainty sets the probability of repair is less than the probability of do not repair. By comparing the optimal discounted reward of all uncertainty models with the CCMDP model, it is easy to see that the statistical distance based uncertainty sets give better optimal policy as compared to moments based uncertainty sets and the nonnegative support uncertainty sets give better optimal policy as compared to

full support uncertainty sets. At the last state, the optimal policy is to choose repair action with a very high probability for all the uncertainty sets. We conduct a time analysis by examining the number of states within the range of 1000 to 10000 for all uncertainty sets. All the parameters are taken similar to the case of 10 states. The results are presented in Figure 5 which shows that the CPU time is almost always the same to solve SOCP (10) with $\kappa = \sqrt{\frac{1-\epsilon}{\epsilon}}$ and the MISOCP (36) while additional CPU time is required to solve the SDP + LP relaxations of the copositive optimization problem (11) and the biconvex optimization problem (38).

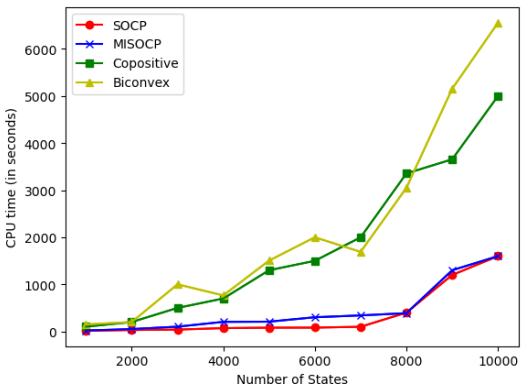

Fig. 2 CPU time (in seconds) to solve SOCP (10) with $\kappa = \sqrt{\frac{1-\epsilon}{\epsilon}}$, MISOCP (36), copositive optimization problem (11) and biconvex optimization problem (38) with different number of states.

6 Conclusions

We consider a DRCCMDP problem under various uncertainty sets. We propose equivalent SOCP, MISOCP, copositive optimization problem and biconvex optimization problem, depending on the choice of the uncertainty set, for the DRCCMDP problem. All these optimization problems except biconvex optimization problems and copositive optimization problems can be solved efficiently using known optimization solvers. We perform numerical experiments, using the optimization solvers in python, on a machine replacement problem using randomly generated data. The numerical experiments are performed on the DRCCMDP problem up to 10000 states and it is very clear from our time analysis that these problems can be solved very efficiently.

Acknowledgement

This research was supported by the French government under the France 2030 program, reference ANR-11-IDEX-0003 within the OI H-Code.

Compliance with Ethical Standards

The authors have no relevant financial or non-financial interests to disclose.

A Proof of Lemma 1

Consider the optimization problem

$$\begin{aligned} & v_{\mathrm{P}}(\mu,\Sigma) = \sup_{F\in\mathcal{C}^+} \int_{\varphi} \mathbf{1}_{\{\chi^{\mathrm{T}}\hat{R}\leq y\}} \mathrm{d}F(\hat{R}) \\ \text{s.t.} \quad & \text{(i)} \int_{\varphi} \mathrm{d}F(\hat{R}) = 1, \ \text{(ii)} \int_{\varphi} (\hat{R}-\mu)(\hat{R}-\mu)^{\mathrm{T}} \mathrm{d}F(\hat{R}) = \Sigma, \\ & \text{(iii)} \int_{\varphi} \hat{R}\mathrm{d}F(\hat{R}) = \mu, \end{aligned} \tag{42}$$

where $\mathcal{C}^+$ is the set of all positive measures on $\mathbb{R}_+^{|\mathcal{W}|}$. The dual problem of (42) is given by

$$\begin{aligned} & v_{\mathrm{D}}(\mu,\Sigma) = \inf \quad -t - Q\circ\Sigma - q^{\mathrm{T}}\mu \\ \text{s.t..} \ & \text{(i)} \ \mathbf{1}_{\{\chi^{\mathrm{T}}\xi\leq y\}} + q^{\mathrm{T}}\xi + \xi^{\mathrm{T}}Q\xi - 2\xi^{\mathrm{T}}Q\mu + \mu^{\mathrm{T}}Q\mu + t \leq 0, \ \forall\ \xi\in\mathbb{R}_+^{|\mathcal{W}|}, \\ & \text{(ii)} \ Q\in\mathcal{S}^{|\mathcal{W}|}, \end{aligned} \tag{43}$$

where t, q, and Q are the dual variables associated with the constraints (i), (ii) and (iii) of (42), respectively. In Theorem 3.4 of [4], under the assumption $\mu\in\mathrm{RI}(\varphi)$, the authors show that the Dirac distribution δ_μ lies in the relative interior of the distributional uncertainty set which implies that the weaker condition of Proposition 3.4 of [22] holds. However, it is not trivial to find a strictly feasible point inside our distributional uncertainty set. Let $(t_j^*, Q_j^*, q_j^*)_{j\in\mathbb{N}}$ be a sequence of feasible solutions of (43) such that

$$-t_j^* - Q_j^*\circ\Sigma - q_j^{*\mathrm{T}}\mu \to v_{\mathrm{D}}(\mu,\Sigma), \ \text{as } j\to\infty. \tag{44}$$

For each $j\in\mathbb{N}$, let $r_j^* = \max(0,q_j^*) - q_j^*$, where $\max(0,q_j^*)$ denotes a $|\mathcal{W}|$-dimensional vector with i^{th} component equal to the maximum value between 0 and the i^{th} component of q_j^*, for every $i=1,\ldots,|\mathcal{W}|$. Let ϵ_j be a strictly positive decreasing sequence such that $\epsilon_j r_j^*\to 0$ componentwise and $\epsilon_j\to 0$, when $j\to\infty$. Let $x_j = \epsilon_j\mathbf{1}$, where $\mathbf{1}$ denotes the vector with all components equal to 1. Then, $r_j^{*\mathrm{T}}x_j\to 0$ as $j\to\infty$. For each $j\in\mathbb{N}$, consider the following conic optimization problem

$$\begin{aligned} & v_{\mathrm{P}}^j(\mu,\Sigma) = \sup_{F\in\mathcal{C}^+} \int_{\varphi} \mathbf{1}_{\{\chi^{\mathrm{T}}R\leq y\}} \mathrm{d}F(R) \\ \text{s.t.} \quad & \text{(i)} \int_{\varphi} \mathrm{d}F(R) = 1, \ \text{(ii)} \int_{\varphi} (R-\mu)(R-\mu)^{\mathrm{T}} \mathrm{d}F(R) = \Sigma, \end{aligned}$$

$$\text{(iii) } \mu \leq \int_{\varphi} R \mathrm{d}F(R) \leq \mu + x_j. \tag{45}$$

The dual problem of (45) is given by

$$\begin{aligned} & v_{\mathrm{D}}^{j}(\mu, \Sigma) = \inf \quad -t - Q \circ \Sigma + (r-h)^{\mathrm{T}}\mu + r^{T}x_j \\ \text{s.t.} \quad & \text{(i) } \mathbf{1}_{\{\chi^{\mathrm{T}}\xi \leq y\}} + (h-r)^{\mathrm{T}}\xi + \xi^{\mathrm{T}}Q\xi - 2\xi^{\mathrm{T}}Q\mu + \mu^{\mathrm{T}}Q\mu + t \leq 0, \ \forall\ \xi \in \mathbb{R}_{+}^{|\mathcal{W}|}, \\ & \text{(ii) } h, r \in \mathbb{R}_{+}^{|\mathcal{W}|}, Q \in \mathcal{S}^{|\mathcal{W}|}, \end{aligned} \tag{46}$$

where t, Q, r and h are the dual variables of the constraint (i), (ii) and (iii) of (45), respectively. The vector (t, Q, h, r) such that $t = t_j^*, Q = Q_j^*$, $h = \max(0, q_j^*)$, $r = r_j^*$ is a feasible solution of (46). Hence,

$$v_{\mathrm{D}}^{j}(\mu, \Sigma) \leq -t_j^* - Q_j^* \circ \Sigma - q_j^{*\mathrm{T}}\mu + r_j^{*\mathrm{T}}x_j, \ \forall\ j \in \mathbb{N}. \tag{47}$$

Since the feasibility set of (14) is non-empty, there exists a nonnegative distribution F^* such that $\mathbb{E}(F^*) = \mu$ and $\mathrm{Cov}(F^*) = \Sigma$. Let F_j be a distribution with support $\varphi_j := \left\{\xi \parallel \xi \in \mathbb{R}_{+}^{\mathcal{W}}, \xi \geq \frac{x_j}{2}, \text{ componentwise}\right\}$, defined by

$$F^*(\xi) = F_j(\xi + \frac{x_j}{2}), \ \forall\ \xi \in \mathbb{R}_{+}^{\mathcal{W}}.$$

It is clear that F_j is a feasible solution of (45) and $\varphi_j \subset \mathrm{RI}(\varphi)$. Hence, F_j belongs to the relative interior of the distributional uncertainty set which implies that strong duality holds, i.e., $v_{\mathrm{P}}^{j}(\mu, \Sigma) = v_{\mathrm{D}}^{j}(\mu, \Sigma)$ for all $j \in \mathbb{N}$. Since the objective function of (45) is a continuous function of F and $x_j \to 0$ as $j \to \infty$, then $v_{\mathrm{P}}^{j}(\mu, \Sigma) \to v_{\mathrm{P}}(\mu, \Sigma)$ as $j \to \infty$. Therefore, it is sufficient to prove that $v_{\mathrm{D}}^{j}(\mu, \Sigma) \to v_{\mathrm{D}}(\mu, \Sigma)$ as $j \to \infty$. It is clear that the feasible sets of (46) and (43) are equivalent and objective function of (46) has additional positive term. Therefore,

$$v_{\mathrm{D}}^{j}(\mu, \Sigma) \geq v_{\mathrm{D}}(\mu, \Sigma), \ \forall\ j \in \mathbb{N}. \tag{48}$$

Using (44), (47) and (48) and the fact that $r_j^{*T}x_j \to 0$ as $j \to \infty$, we have $v_{\mathrm{D}}^{j}(\mu, \Sigma) \to v_{\mathrm{D}}(\mu, \Sigma)$ as $j \to \infty$.

that: $L = \sup_{\beta<\lambda,\lambda>0} \left\{ \frac{\lambda^2}{(\lambda-\beta)(\lambda-\beta+1)} \mathbb{P}_{\nu}(O) - \frac{\beta^2}{\lambda-\beta} - \lambda\theta_{\phi} \right\}$. Let $\eta = \lambda - \beta$. Then, we can write

B Proof of Lemma 5

The optimization problem $\inf_{z \in \mathbb{R}_{+}^{|\mathcal{W}|}, \chi^{\mathrm{T}}z \leq y} ||\tilde{\xi}_i - z||_2$ can be reformulated as following SOCP problem

$$\begin{aligned} & \min \quad t \\ \text{s.t.} \quad & \chi^{\mathrm{T}}z \leq y, \ t \geq ||\tilde{\xi}_i - z||_2, \ z \in \mathbb{R}_{+}^{|\mathcal{W}|}. \end{aligned} \tag{49}$$

The Lagrangian dual problem of (49) is given by

$$\max_{\lambda_i \geq 0, \zeta_i \in \mathbb{R}_+^{|\mathcal{W}|}, \beta \geq 0} \quad \min_{t \in \mathbb{R}, z \in \mathbb{R}^{|\mathcal{W}|}} \mathcal{L}(t, \chi, z, \lambda_i, \beta, \zeta_i),$$

where $\mathcal{L}(t, z, \lambda_i, \beta, \zeta_i) = t + \lambda_i(\chi^{\mathrm{T}} z - y) - \zeta_i^{\mathrm{T}} z + \beta(||\tilde{\xi}_i - z||_2 - t)$. The inner minimization problem can be written as

$$J(\lambda_i, \zeta_i, \beta) = \min_{t \in \mathbb{R}, z \in \mathbb{R}^{|\mathcal{W}|}} \left\{ t(1-\beta) + \beta||\tilde{\xi}_i - z||_2 + \lambda_i \chi^{\mathrm{T}} z - \zeta_i^{\mathrm{T}} z - \lambda_i y \right\}. \tag{50}$$

It is easy to see that $J(\lambda_i, \zeta_i, \beta) = -\infty$ if $\beta \neq 1$ and it implies that the dual objective function value is $-\infty$. By using the strong duality of a primal-dual pair of SOCPs, the objective function value of primal problem is $-\infty$, i.e., $\inf_{z \in \mathbb{R}_+^{|\mathcal{W}|}, \chi^{\mathrm{T}} z \leq y} ||\tilde{\xi}_i - z||_2 = -\infty$ which is a contradiction. Therefore, $\beta = 1$ and using a change of variable $z_1 = \tilde{\xi}_i - z$, we have

$$J(\lambda_i, \zeta_i, 1) = \min_{z_1 \in \mathbb{R}^{|\mathcal{W}|}} \left\{ ||z_1||_2 + (\zeta_i - \lambda_i \chi)^{\mathrm{T}} z_1 \right\} + \lambda_i(\chi^{\mathrm{T}} \tilde{\xi}_i - y) - \zeta_i^{\mathrm{T}} \tilde{\xi}_i.$$

The above minimization problem is unbounded unless $||\zeta_i - \lambda_i \chi||_2 \leq 1$ and it leads to the following dual problem of (49).

$$\begin{aligned} \max \quad & \lambda_i(\chi^{\mathrm{T}} \tilde{\xi}_i - y) - \zeta_i^{\mathrm{T}} \tilde{\xi}_i \\ \text{s.t.} \quad & ||\zeta_i - \lambda_i \chi||_2 \leq 1, \ \lambda_i \geq 0, \zeta_i \in \mathbb{R}_+^{|\mathcal{W}|}. \end{aligned} \tag{51}$$

References

[1] E. Altman. *Constrained Markov Decision Processes*. Routledge, New York, 1999.

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

[3] G. C. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[4] J. Cheng, E. Delage, and A. Lisser. Distributionally robust stochastic knapsack problem. *SIAM Journal on Optimization*, 24(3):1485–1506, 2014.

[5] E. Delage and S. Mannor. Percentile optimization for Markov decision processes with parameter uncertainty. *Operations Research*, 58(1):203–213, 2010.

[6] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3): 595–612, 2010.

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

[8] R. Gao and A. Kleywegt. Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*, 48(2):603–655, 2023.

[9] L. E. Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4): 543–556, 2003.

[10] R. Henrion. Structural properties of linear probabilistic constraints. *Optimization*, 56(4):425–440, 2007.

[11] J.-B. Hiriart-Urruty and A. Seeger. A variational approach to copositive matrices. *SIAM Review*, 52(4):593–629, 2010.

[12] G. N. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280, 2005.

[13] R. Jiang and Y. Guan. Data-driven chance constrained stochastic program. *Mathematical Programming*, 158(1):291–327, 2016.

[14] J. Liu, A. Lisser, and Z. Chen. Distributionally robust chance constrained geometric optimization. *Mathematics of Operations Research*, 47(4):2950–2988, 2022.

[15] S. Mannor, D. Simester, P. Sun, and J. N. Tsitsiklis. Bias and variance approximation in value function estimates. *Management Science*, 53(2):308–322, 2007.

[16] H. N. Nguyen, V. V. Singh, A. Lisser, and M. Arora. Zero-sum games with distributionally robust chance constraints. In *17th International Conference on Internet and Web Applications and Services (ICIW)*, pages 7–12, 2022.

[17] H. N. Nguyen, V. V. Singh, A. Lisser, and M. Arora. Distributionally robust chance-constrained zero-sum games with moments based and statistical based uncertainty sets. *International Journal on Advances in Systems and Measurements*, 15(3-4):107–120, 2022.

[18] A. Nilim and L. El Ghaoui. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.

[19] L. Pardo. *Statistical Inference Based on Divergence Measures*. Chapman and Hall/CRC Press, New York, 2018.

[20] I. Popescu. Robust mean-covariance solutions for stochastic optimization. *Operations Research*, 55(1):98–112, 2007.

[21] M. Puterman et al. *Markov Decision Processes*. John Wiley & Sons, Inc., New York, 1994.

[22] A. Shapiro. On duality theory of conic linear problems. In M. Á. Goberna and M. A. López, editors, *Semi-Infinite Programming. Nonconvex Optimization and Its Applications*, volume 57, pages 135–165. Springer US, Boston, MA, 2001.

[23] V. V. Singh, O. Jouini, and A. Lisser. Distributionally robust chance-constrained games: Existence and characterization of Nash equilibrium. *Optimization Letters*, 11(7):1385–1405, 2017.

[24] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1): 171–176, 1958.

[25] V. Varagapriya, V. V. Singh, and A. Lisser. Constrained Markov decision processes with uncertain costs. *Operations Research Letters*, 50(2):218–223, 2022.

[26] V. Varagapriya, V. V. Singh, and A. Lisser. Joint chance-constrained Markov decision processes. *Annals of Operations Research*, 322(2):1013–1035, 2023.

[27] C. Villani. *Optimal Transport*. Springer, Berlin, Heidelberg, 2009.

[28] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, Providence, 2021.

[29] C. C. White III and H. K. Eldeib. Markov decision processes with imprecise transition probabilities. *Operations Research*, 42(4):739–749, 1994.

[30] W. Wiesemann, D. Kuhn, and B. Rustem. Robust Markov decision processes. *Mathematics of Operations Research*, 38(1):153–183, 2012.

[31] W. Xie. On distributionally robust chance constrained programs with wasserstein distance. *Mathematical Programming*, 186(1-2):115–155, 2021.

[32] C. Zhao and Y. Guan. Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters*, 46(2):262–267, 2018.