Distributionally robust chance-constrained Markov decision processes

Hoang Nam Nguyen, Abdel Lisser, Vikas Vikram Singh

Noname manuscript No.
(will be inserted by the editor)

Distributionally robust chance-constrained Markov decision processes

Hoang Nam Nguyen · Abdel Lisser · Vikas Vikram Singh

Received: date / Accepted: date

Abstract Markov decision process (MDP) is a decision making framework where a decision maker is interested in maximizing the expected discounted value of a stream of rewards received at future stages at various states which are visited according to a controlled Markov chain. Many algorithms including linear programming methods are available in the literature to compute an optimal policy when the rewards and transition probabilities are deterministic. In this paper, we consider an MDP problem where the transition probabilities are known and the reward vector is a random vector whose distribution is partially known. We formulate the MDP problem using distributionally robust chance-constrained optimization framework under various types of moments based uncertainty sets, and statistical-distance based uncertainty sets defined using ϕ-divergence and Wasserstein distance metric. For each type of uncertainty set, we consider the case where a random reward vector has either a full support or a nonnegative support. For the case of full support, we show that the distributionally robust chance-constrained Markov decision process is equivalent to a second-order cone programming problem for the moments and ϕ-divergence distance based uncertainty sets, and it is equivalent to a mixed-integer second-order cone programming problem for an Wasserstein distance based uncertainty set. For the case of nonnegative support, it is equivalent to a copositive optimization problem and a biconvex optimization problem for the moments based uncertainty sets and Wasserstein distance based uncertainty set, respectively. As an application, we study a machine replacement problem and illustrate numerical experiments on randomly generated instances.

This research was supported by DST/CEFIPRA Project No. IFC/4117/DST-CNRS-5th call/2017-18/2 and CNRS Project No. AR/SB:2018-07-440.

University Paris Saclay, CNRS, CentraleSupelec, L2S, Bat Breguet, 3 rue Joliot Curie, 91190, Gif-sur-Yvette, France (E-mail: hoang-nam.nguyen3@centralesupelec.fr).

University Paris Saclay, CNRS, CentraleSupelec, L2S, Bat Breguet, 3 rue Joliot Curie, 91190, Gif-sur-Yvette, France (E-mail: abdel.lisser@centralesupelec.fr), Corresponding Author

Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, 110016, India (E-mail: vikassingh@maths.iitd.ac.in).

Address(es) of author(s) should be given

Keywords Markov decision processes · Distributionally robust chance-constrained optimization · Second-order cone programming · Copositive optimization · Mix-integer second-order cone programming · Biconvex optimization.

Mathematics Subject Classification (2000) 90C15 · 90C25 · 90C40 · 90C59

1 Introduction

An MDP is a decision making framework to model the performance of a stochastic system which evolves over time according to a controlled Markov chain. We consider the case where the system has a finite number of states. At time $t = 0$, the system is at some initial state $s_0 \in S$, where S is a finite state space, according to an initial distribution γ, and a decision maker chooses an action $a_0 \in A(s_0)$, where $A(s_0)$ denotes the set of finite number of actions available to the decision maker at state s_0. As a consequence a reward $R(s_0, a_0)$ is earned and at time $t = 1$, the system moves to a new state s_1 with probability $p(s_0, a_0, s_1)$. The same thing repeats at time $t = 1$ and it continues for the infinite horizon. The decision taken at time t, which could be deterministic or randomized, may depend on the history h_t at time t, where $h_t = (s_0, a_0, s_1, \ldots, s_{t-1}, a_{t-1}, s_t)$. Let H_t be the set of all possible histories at time t. A history dependent decision rule f_t at time t is defined as $f_t(h_t) \in \wp(A(s_t))$ for every history h_t with final state s_t, where $\wp(A(s_t))$ denotes the set of probability distributions on the action set $A(s_t)$. A sequence of history dependent decision rules $f^h = (f_t)_{t=0}^{\infty}$ is called a history dependent policy. The policy is called Markovian if each f_t in the sequence $(f_t)_{t=0}^{\infty}$ depends only on the state at time t. A Markovian policy $(f_t)_{t=0}^{\infty}$ is called a stationary policy if there exists a decision rule f such that $f_t = f$ for all t. Therefore, a stationary policy can be represented as a sequence of the same decision rules $(f, f, \ldots)$ and with some abuse of notations we can denote it as f, and define $f = (f(s))_{s\in S}$ such that $f(s) \in \wp(A(s))$ for every $s \in S$. As per a stationary policy f, whenever the Markov chain visits state s, the decision maker chooses an action a with probability $f(s, a)$. We denote the set of all history dependent and stationary policies by PO_{HD} and PO_S, respectively. A history dependent policy $f^h \in PO_{HD}$ and an initial distribution γ define a probability measure $P_\gamma^{f^h}$ over the state and action trajectories, and $E_\gamma^{f^h}$ denotes the expectation operator corresponding to the probability measure $P_\gamma^{f^h}$. For a given policy f^h and an initial distribution γ, the expected discounted reward at a discount factor $\alpha \in (0, 1)$ is defined as [1, 32]

$$\begin{aligned} V_\alpha(\gamma, f^h) &= (1-\alpha)\mathbb{E}_\gamma^{f^h}\left(\sum_{t=0}^{\infty}\alpha^t R(X_t, A_t)\right), \\ &= \sum_{s\in S}\sum_{a\in A(s)} g_\alpha(\gamma, f^h; s, a)R(s, a), \end{aligned} \tag{1}$$

where X_t and A_t represent the state and the action at time t, respectively. For a given policy f^h, the set $\{g_\alpha(\gamma, f^h; s, a)\}_{(s,a)}$ is the occupation measure defined by

$$g_\alpha(\gamma, f^h; s, a) = (1-\alpha)\sum_{t=0}^{\infty}\alpha^t P_\gamma^{f^h}(\mathbb{X}_t = s, \mathbb{A}_t = a),\ \forall\, s \in S,\ a \in A(s).$$

When the running rewards and the transition probabilities are stationary, i.e., $R(X_t = s, A_t = a) = R(s, a)$ and $P(X_{t+1} = s' | X_t = s, A_t = a) = p(s, a, s')$ for all t, we can restrict to stationary policies without loss of optimality [1, 32]. It follows from Theorem 3.2 on p. 28

in [1] that the set of occupation measures corresponding to history dependent policies is equal to the set of occupation measures corresponding to stationary policies and further it is equal to the set

$$Q_\alpha(\gamma) = \Bigg\{ \rho \in \mathbb{R}^{|\mathcal{K}|} \,\Big|\, \sum_{(s,a)\in\mathcal{K}} \rho(s,a)\Big(\delta(s',s) - \alpha p(s,a,s')\Big) = (1-\alpha)\gamma(s'),\ \forall\, s' \in S,$$
$$\rho(s,a) \geq 0,\ \forall\, s \in S,\ a \in A(s) \Bigg\},$$

such that the value of the expected discounted reward defined by (1) remains the same; $\delta(s', s)$ is the Kronecker delta and $\mathcal{K} = \{(s,a) \mid s \in S, a \in A(s)\}$. Therefore, the optimal policy of the MDP problem can be obtained by solving the following linear programming problem [32]

$$\max_{\rho\in Q_\alpha(\gamma)} \rho^{\mathrm{T}} R, \tag{2}$$

where $R = (R(s,a))_{s\in S, a\in A(s)}$ is a running reward vector and T denotes the transposition. If ρ^* is an optimal solution of (2), the stationary optimal policy f^* can be defined as

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

whenever the denominator is nonzero $\big($if it is zero, we choose $f^*(s)$ arbitrarily from $\wp(A(s))\big)$ [1]. In practice, the MDP model parameters $R(\cdot)$ and $p(\cdot)$ are not known in advance and are estimated from historical data. This leads to errors in the optimal policies [26]. Most efforts to take into account this uncertainty focused on the study of robust MDPs where the rewards or the transition probabilities are known to belong to a prespecified uncertainty set [22,28,41,46,47]. However, Delage and Mannor [9] showed that the robust MDP approach usually leads to conservative policies. For this reason, a chance-constrained Markov decision process (CCMDP) was introduced in [9], where the controller obtains the expected discounted reward with certain confidence. In [9], the case of random rewards and random transition probabilities are considered separately and it is shown that a CCMDP is equivalent to a second-order cone programming (SOCP) problem when the running reward vector follows a multivariate normal distribution and the transition probabilities are exactly known. When the transition probabilities follow Dirichlet distribution and the running rewards are exactly known, the CCMDP problem becomes intractable and the optimal policies can be computed using approximation methods. Varagapriya et al. [42] considered a CMDP problem with joint chance constraint where the running cost vectors are random vectors and the transition probabilities are known. They proposed two SOCP based approximations which give upper and lower bounds to the CMDP problem if the cost vectors follow multivariate elliptical distributions and the dependence among the constraints is driven by a Gumbel-Hougaard copula.

In many practical situations, it is often the case that only a partial information about the underlying distribution is available based on historical data. In that case, a distributionally robust approach, is used to model the uncertainties, which assumes that the true distribution belongs to an uncertainty set based on its partially available information. Such an approach has been used in modelling the uncertainties of many optimization and game problems [23, 24, 39]. There are at least two popular ways to construct an uncertainty set for the distribution of the uncertain parameters. The first one is based on the partial information on moments of the true distribution and the second one is based on the statistical distance

between the true distribution and a nominal distribution. The moments-based uncertainty sets assume certain conditions on the first two moments [8, 10, 31]. The statistical distance-based uncertainty sets contain all the distributions which lie inside a ball of small radius and center at a nominal distribution which is usually considered to be an empirical distribution or a normal distribution [13, 23]. To define a distance between the distributions, either a $\phi-$divergence [3, 23] or Wasserstein distance metric is used [13, 15, 51].

In this paper, we consider an infinite horizon MDP with discounted payoff criterion defined in Section 1 where the reward vector is a random vector and the transition probabilities are known. The distribution of the reward vector is not completely known and it is assumed to belong to a given uncertainty set. We formulate the random discounted reward with a distributionally robust chance constraint which guarantees the maximum reward for a given policy with at least a given level of confidence. We call this class of MDP as a distributionally robust chance-constrained Markov decision process (DRCCMDP). The random reward vector has either a full support or a nonnegative support. We consider both moments and statistical distance based uncertainty sets. The main contributions of the paper are as follows.

1. We consider three different 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.
4. We illustrate our theoretical results on a machine replacement problem [9].

The paper is organized as follows. In Section 2, we define a DRCCMDP under a discounted payoff 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 metric and Wasserstein distance metric and their equivalent reformulations are presented in Section 4. The numerical results on a machine replacement problem is given in Section 5. We conclude the paper in Section 6.

2 Distributionally robust chance constrained Markov decision process

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 γ, the expected discounted reward defined by (1) becomes a random variable. We consider the case where the controller is interested in a maximum discounted reward which can be obtained with at least a given 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, it follows from the discussion given in Section 1 that we can represent the CCMDP problem (3) equivalently in terms of decision vector (y, ρ) as follows

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

If then vector $\hat{R}$ follows a multivariate normal distribution, the optimization problem (4) is equivalent to an SOCP problem [9]. 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 [20].

However, in most of the practical situations, we only have partial information about the underlying probability distributions. Such situations can be handled with the distributionally robust optimization approach, i.e., we assume that the distribution of $\hat{R}$ belongs to an uncertainty set. This leads to the following DRCCMDP problem

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

where F is the distribution of $\hat{R}$ and $\mathcal{D}$ is a given uncertainty set. The first constraint of (5) can be written as

$$\sup_{F\in\mathcal{D}} \mathbb{P}_F\left(\rho^{\mathrm{T}}\hat{R} < y\right) \leq \epsilon.$$

Note that $\mathbb{P}_F(\rho^{\mathrm{T}}\hat{R} \leq y-\theta) \leq \mathbb{P}_F(\rho^{\mathrm{T}}\hat{R} < y) \leq \mathbb{P}_F(\rho^{\mathrm{T}}\hat{R} \leq y)$ for every $\theta > 0$. Therefore, we can replace $\sup_{F\in\mathcal{D}} \mathbb{P}_F\left(\rho^{\mathrm{T}}\hat{R} < y\right)$ by $\sup_{F\in\mathcal{D}} \mathbb{P}_F\left(\rho^{\mathrm{T}}\hat{R} \leq y\right)$. Then, the problem (5) is equivalent to the following problem

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

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 Moments based uncertainty sets

Let $\mu \in \mathbb{R}^{|\mathcal{K}|}$ the mean vector and $\Sigma \succ 0$ 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. **Uncertainty set with known mean and known covariance matrix:** The uncertainty set of the distribution of $\hat{R}$ in this case is defined by

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

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

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

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

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

where $\varphi \subset \mathbb{R}^{|\mathcal{K}|}$ is the support of $\hat{R}$ which we assume to be a convex set, $\mathcal{M}^+$ is the set of all probability measures on $\mathbb{R}^{|\mathcal{K}|}$ 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 $A \preceq B$ implies that $B - A$ 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., $\varphi = \mathbb{R}^{|\mathcal{K}|}$. 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* $\varphi = \mathbb{R}^{|\mathcal{K}|}$*. Then, the* DRCCMDP (5) *can be reformulated equivalently as the following* SOCP

$$\begin{array}{lll} & \max & y \\ s.t. & \text{(i)} & \mu^{\mathrm{T}} \rho - \kappa \|\Sigma^{\frac{1}{2}} \rho\|_2 \geq y, \\ & \text{(ii)} & \rho \in Q_\alpha(\gamma), \end{array} \tag{10}$$

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

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 [7, 16]. For the uncertainty sets (8) and (9), it can be proved using similar arguments used in Lemma 3.1 and Lemma 3.2 of [27] which are based on the one-sided Chebyshev inequality [24].

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	$\sqrt{\frac{1-\epsilon}{\epsilon}}$	$\sqrt{\frac{(1-\epsilon)\delta_0}{\epsilon}}$	$\sqrt{\frac{(1-\epsilon)\delta_2}{\epsilon}} + \sqrt{\delta_1}$

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 $|\mathcal{K}|$-dimensional Euclidean space, i.e., $\varphi = \mathbb{R}_+^{|\mathcal{K}|}$. We show that the DRCCMDP problem (6) is equivalent to a copositive optimization problem.

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

1. *If the distribution of* $\hat{R}$ *belongs to the uncertainty set defined by* (7)*, the DRCCMDP problem* (6) *is equivalent to the following copositive optimization problem*

$$
\begin{aligned}
& \max \quad y \\
s.t. \quad & \text{(i)} \quad -t - Q \circ \Sigma - q^{\mathrm{T}}\mu \le \epsilon, \\
& \text{(ii)} \quad \begin{pmatrix} -Q & -\frac{1}{2}q + Q\mu \\ -\frac{1}{2}q^{\mathrm{T}} + \mu^{\mathrm{T}}Q & -t - \mu^{\mathrm{T}}Q\mu \end{pmatrix} \in \mathrm{COP}^{|\mathcal{K}|+1}, \\
& \text{(iii)} \begin{pmatrix} -Q & -\frac{1}{2}q + Q\mu + \lambda\rho \\ -\frac{1}{2}q^{\mathrm{T}} + \mu^{\mathrm{T}}Q + \lambda\rho^{\mathrm{T}} & -t - \mu^{\mathrm{T}}Q\mu - 1 - \lambda y \end{pmatrix} \in \mathrm{COP}^{|\mathcal{K}|+1}, \\
& \text{(iv)} \quad Q \in \mathcal{S}^{|\mathcal{K}|}, \lambda \ge 0, \\
& \text{(v)} \quad \rho \in \mathcal{Q}_\alpha(\gamma).
\end{aligned} \tag{11}
$$

2. *If the distribution of* $\hat{R}$ *belongs to the uncertainty set defined by* (8)*, the DRCCMDP problem* (6) *is equivalent to the following copositive optimization problem*

$$
\begin{aligned}
& \max \quad y \\
s.t. \quad & \text{(i)} \quad -t - \mu^{\mathrm{T}}q - \mu^{\mathrm{T}}Q\mu + \delta_0\Sigma \circ Q \le \epsilon, \\
& \text{(ii)} \quad \begin{pmatrix} Q & -\frac{1}{2}q - Q\mu \\ -\frac{1}{2}q^{\mathrm{T}} - \mu^{\mathrm{T}}Q & -t \end{pmatrix} \in \mathrm{COP}^{|\mathcal{K}|+1}, \\
& \text{(iii)} \quad \begin{pmatrix} Q & \frac{1}{2}(-q + \lambda\rho) - Q\mu \\ \frac{1}{2}(-q + \lambda\rho)^{\mathrm{T}} - \mu^{\mathrm{T}}Q & -t - 1 - \lambda y \end{pmatrix} \in \mathrm{COP}^{|\mathcal{K}|+1}, \\
& \text{(iv)} \quad Q \in \mathcal{S}_+^{|\mathcal{K}|}, \lambda \ge 0, \\
& \text{(v)} \quad \rho \in \mathcal{Q}_\alpha(\gamma).
\end{aligned} \tag{12}
$$

3. *If the distribution of $\hat{R}$ belongs to the uncertainty set defined by* (9)*, the DRCCMDP problem* (6) *is equivalent to the following copositive optimization problem*

$$\begin{aligned}
&\max \quad y \\
s.t.\ &\text{(i)} \quad r + t \leq \epsilon, \\
&\text{(ii)} \quad \begin{pmatrix} Q & \frac{1}{2}q \\ \frac{1}{2}q^{\mathrm{T}} & r \end{pmatrix} \in \mathrm{COP}^{|\mathcal{K}|+1}, \\
&\text{(iii)} \quad t \geq (\delta_2 \Sigma + \mu\rho^{\mathrm{T}}) \circ Q + \rho^{\mathrm{T}} q + \sqrt{\delta_1}||\Sigma^{\frac{1}{2}}(q + 2Q\mu)||_2, \\
&\text{(iv)} \quad \begin{pmatrix} Q & \frac{1}{2}(q+\lambda\rho) \\ \frac{1}{2}(q+\lambda\rho)^{\mathrm{T}} & r - 1 - \lambda y \end{pmatrix} \in \mathrm{COP}^{|\mathcal{K}|+1}, \\
&\text{(v)} \quad Q \in \mathcal{S}_+^{|\mathcal{K}|}, \lambda \geq 0, \\
&\text{(vi)} \quad \rho \in \mathcal{Q}_\alpha(\gamma),
\end{aligned} \tag{13}$$

where $\mathrm{COP}^{|\mathcal{K}|+1} = \left\{ M \in \mathcal{S}^{|\mathcal{K}|+1} \mid x^{\mathrm{T}} M x \geq 0,\ \forall\, x \in \mathbb{R}_+^{|\mathcal{K}|+1} \right\}$, *$\mathcal{S}^n$ 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 $\left(\!\!\begin{array}{c}-\\-\end{array}\!\!\right)$ 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*

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

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

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

such that strong duality holds.

The proof is given in Appendix A.

Proof (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

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

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

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

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

$$U = \begin{pmatrix} -Q & -\frac{1}{2}q + Q\mu \\ \hline -\frac{1}{2}q^{\mathrm{T}} + \mu^{\mathrm{T}}Q & -t - \mu^{\mathrm{T}}Q\mu \end{pmatrix}.$$

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

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

Using Proposition 5.1 in [21], we deduce that the constraint (ii) of (15) is equivalent to $U \in \mathrm{COP}^{|\mathcal{K}|+1}$. The constraint (iii) of (15) is equivalent to:

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

Define,

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

where $\mathcal{L}(\lambda, \xi, U, \rho, y) = -1 + (\xi^{\mathrm{T}}, 1)U(\xi^{\mathrm{T}}, 1)^{\mathrm{T}} + \lambda(\rho^{\mathrm{T}}\xi - y)$. In [8], the authors use the Sion's minimax theorem [40] 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 ρ, we have

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

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

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

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

$$\min_{\xi \in \mathbb{R}_{+}^{|\mathcal{K}|}} \max_{\lambda \geq 0} \quad \mathcal{L}(\lambda, \xi, U_i, \rho_i, y) \xrightarrow{i \to \infty} \min_{\xi \in \mathbb{R}_{+}^{|\mathcal{K}|}} \max_{\lambda \geq 0} \quad \mathcal{L}(\lambda, \xi, U, \rho, y).$$

$$\max_{\lambda \geq 0} \min_{\xi \in \mathbb{R}_{+}^{|\mathcal{K}|}} \quad \mathcal{L}(\lambda, \xi, U_i, \rho_i, y) \xrightarrow{i \to \infty} \max_{\lambda \geq 0} \min_{\xi \in \mathbb{R}_{+}^{|\mathcal{K}|}} \quad \mathcal{L}(\lambda, \xi, U, \rho, y).$$

This implies that, if $s_{\mathrm{P}} = s_{\mathrm{D}}$ holds for any U_i, ρ_i, $i \in \mathbb{N}$, it also holds for U, ρ. For an arbitrary U_i and ρ_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, ρ_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 $\rho_i^{\mathrm{T}}\xi_{\mathrm{P}} - y \leq 0$. 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, ρ_i and y. As $\xi \in \mathbb{R}_+^{|\mathcal{K}|}$ and $\rho_i > 0$, componentwise, we have

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

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

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

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, ρ_i and y. Let $\Upsilon = \max(\Upsilon_{\mathrm{P}}, \Upsilon_{\mathrm{D}})$. Then, (17) is equivalent to

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

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

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

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

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

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

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

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

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

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

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

where $Z \in \mathcal{S}^{|\mathcal{K}|+1}$ and

$$Z = \begin{pmatrix} -Q & -\frac{1}{2}q + Q\mu + \lambda\rho \\ -\frac{1}{2}q^{\mathrm{T}} + \mu^{\mathrm{T}}Q + \lambda\rho^{\mathrm{T}} & -t - \mu^{\mathrm{T}}Q\mu - 1 - \lambda y \end{pmatrix}.$$

Using similar arguments as above, the constraint (20) is equivalent to

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

This implies that the constraint (iii) of (15) is equivalent to (21). Hence, DRCCMDP problem (6) is equivalent to (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 [8], the dual of the optimization problem $\sup_{F\in\mathcal{D}} \mathbb{P}_F \left(\rho^{\mathrm{T}}\hat{R} \leq y\right)$ can be written as

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

and the strong duality holds. The rest of the proof follows from the similar 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 [10] the dual of the problem $\sup_{F\in\mathcal{D}} \mathbb{P}_F \left(\rho^{\mathrm{T}}\hat{R} \leq y\right)$ is given by

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

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 Copositive optimization has been studied in the literature. In practical applications, the copositive constraints can be approximated conservatively by SDP (semidefinite programming) constraints. We refer to [5, 6, 48] for some recent researches about SDP approximations.

4 Statistical distance based uncertainty sets

In this section, we consider uncertainty sets defined using statistical distance metric known as ϕ-divergence and Wasserstein distance. For each uncertainty set, we propose equivalent reformulation of DRCCMDP problem (5) (or (6)).

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 based on the available estimated data. The decision maker believes 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 $\mathbb{R}^{|\mathcal{K}|}$ is given by

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

For different choices of ϕ, we refer to [3] and [29]. 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

$$\mathcal{D}_4(\nu, \theta_\phi) = \left\{F \in \mathcal{M}^+ \mid I_\phi(F, \nu) \leq \theta_\phi\right\}, \tag{22}$$

where $\theta_\phi > 0$.

Definition 2 The conjugate of ϕ is a function $\phi^* : \mathbb{R} \to \mathbb{R} \cup \infty$ such that

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

Lemma 2 *Consider an optimization problem*

$$\inf_{F \in \mathcal{D}_4(\nu, \theta_\phi)} \mathbb{P}_F(\rho^{\mathrm{T}} \hat{R} \geq y). \tag{23}$$

Then, the dual problem of (23) *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 $O = \left\{\xi \in \mathbb{R}^{|\mathcal{K}|} \mid \rho^{\mathrm{T}}\xi \geq y\right\}$*, such that the strong duality holds.*

Proof We rewrite the primal problem (23) as a following semi-infinite programming problem

$$\begin{aligned} v_{\mathrm{P}} = \inf_{F \geq 0} & \int_{\mathbb{R}^{|\mathcal{K}|}} \mathbf{1}_O(\xi) F(\xi) \mathrm{d}\xi \\ \text{s.t.} \quad \text{(i)} \quad & \int_{\mathbb{R}^{|\mathcal{K}|}} f_\nu(\xi) \phi\left(\frac{F(\xi)}{f_\nu(\xi)}\right) \mathrm{d}\xi \leq \theta_\phi, \\ \text{(ii)} \quad & \int_{\mathbb{R}^{|\mathcal{K}|}} F(\xi) \mathrm{d}\xi = 1. \end{aligned} \tag{24}$$

The dual problem of (24) is given by

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

where λ is the dual variable of the constraint (i) of (24) and β is the dual variable of the constraint (ii) of (24). 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 [23].

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

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

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

2, the following result holds.

Theorem 3 *Consider the DRCCMDP problem* (5) *under the uncertainty set defined by* (22) *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*

$$\begin{aligned} & \max \quad y \\ s.t. \quad & \text{(i)} \quad \rho^{\mathrm{T}}\mu_\nu - \Phi^{(-1)}[f(\theta_\phi,\epsilon)]\|\Sigma_\nu^{\frac{1}{2}}\rho\|_2 \geq y, \\ & \text{(ii)} \quad \rho \in \mathcal{Q}_\alpha(\gamma), \end{aligned} \tag{25}$$

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 ϕ-divergences are given in Table 3.

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

$$\mathbb{P}_\nu(\rho^{\mathrm{T}}\hat{R} \geq y) \geq f(\theta_\phi,\epsilon). \tag{26}$$

Since ν is a normal distribution with mean vector μ_ν and covariance matrix Σ_ν, it is well known that (26) is equivalent to the constraint (i) of (25). The details of the proof for the Hellinger distance case is given in Appendix B. The proofs for Kullback-Leibler, Variation distance and Modified χ^2 - distance follow from Propositions 2, 3 and 4 of [23].

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

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

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 [43,44] for more details of the Wasserstein distance metric.

Let φ be a closed, convex subset of $\mathbb{R}^{|\mathcal{K}|}$ 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

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

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

$$\mathcal{D}_5(\varphi,\nu,p,\theta_W) = \left\{F \in \mathcal{P}_p(\varphi) \mid W_p(F,\nu) \le \theta_W\right\}, \tag{27}$$

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

Lemma 3 *Consider an optimization problem*

$$\sup_{F\in\mathcal{D}_5(\varphi,\nu,p,\theta_W)} \mathbb{P}_F(\rho^{\mathrm{T}}\hat{R} \le y). \tag{28}$$

Then, the dual problem of (28) *is given by*

$$\inf_{\lambda\ge 0}\left\{\lambda\theta_W^p - \int_\varphi \inf_{z\in\varphi}\left[\lambda||x-z||_2^p - \mathbf{1}_{\{\rho^T z\le y\}}\right]\nu(\mathrm{d}x)\right\}, \tag{29}$$

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 [15] that the following strong duality holds

$$\begin{aligned}&\sup_{\mu\in\mathcal{P}(\Xi)}\left\{\int_{\Xi}\Psi(\xi)\mu(d\xi)\ \mid\ W_p(\mu,\nu)\le\theta_W\right\}\\&=\inf_{\lambda\in\mathbb{R},\lambda\ge0}\left\{\lambda\theta_W^p-\int_{\Xi}\inf_{\xi\in\Xi}\left[\lambda d^p(\xi,\zeta)-\Psi(\xi)\right]\nu(\mathrm{d}\zeta)\right\}<\infty,\end{aligned}\tag{30}$$

provided the growth factor given by Definition 4 of [15] is finite. We apply this result in our case by choosing $\Xi=\varphi$, d as an Euclidean metric and $\Psi(\xi)=\mathbf{1}_{\{\rho^{\mathrm{T}}\xi\le y\}}$ for all $\xi\in\varphi$. For this choice of $\Psi(\xi)$, it is easy to see from Definition 4 of [15] that the growth factor is zero. Since $\left\{\xi\in\varphi\mid\rho^{\mathrm{T}}\xi\le y\right\}$ 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, (30) reduces to

$$\sup_{F\in\mathcal{D}_5(\varphi,\nu,p,\theta_W)}\mathbb{P}_F\left(\rho^{\mathrm{T}}\hat{R}\le y\right)=\inf_{\lambda\ge0}\left\{\lambda\theta_W^p-\int_{\varphi}\inf_{\xi\in\varphi}\left[\lambda||\zeta-\xi||_2^p-\mathbf{1}_{\{\rho^{\mathrm{T}}\xi\le y\}}\right]\nu(\mathrm{d}\zeta)\right\}.$$

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* (27)*, the* DRCCMDP (6) *can be reformulated equivalently as the following deterministic problem*

$$\begin{aligned}&\sup\quad y\\ s.t.\quad&\text{(i)}\quad\theta_W-\frac{1}{H}\sum_{i=1}^{H}g_i\le l\epsilon,\\&\text{(ii)}\quad\inf_{z\in\varphi,\rho^{\mathrm{T}}z\le y}||\tilde{\xi}_i-z||_2\ge l+g_i,\ \forall\, i=1,\ldots,H,\\&\text{(iii)}\quad l>0,\ \rho\in\mathcal{Q}_\alpha(\gamma),\ g_i\le0,\ \forall\, i=1,\ldots,H.\end{aligned}\tag{31}$$

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

$$\lambda\theta_W-\frac{1}{H}\sum_{i=1}^{H}\inf_{z\in\varphi}\left[\lambda||\tilde{\xi}_i-z||_2-\mathbf{1}_{\{\rho^{\mathrm{T}}z\le y\}}\right]\le\epsilon,\ \lambda\ge0.$$

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

$$\begin{cases}\text{(i)}\quad\lambda\theta_W-\frac{1}{H}\sum_{i=1}^{H}t_i\le\epsilon,\ \lambda\ge0\\ \text{(ii)}\quad\inf_{z\in\varphi}\left[\lambda||\tilde{\xi}_i-z||_2-\mathbf{1}_{\{\rho^{\mathrm{T}}z\le y\}}\right]\ge t_i,\ \forall\, i=1,\ldots,H.\end{cases}\tag{32}$$

The constraint (ii) of (32) is equivalent to the following two constraints

$$\begin{cases}\text{(i)}\quad\inf_{z\in\varphi}\lambda||\tilde{\xi}_i-z||_2\ge t_i,\ \forall\, i=1,\ldots,H,\\ \text{(ii)}\quad\inf_{z\in\varphi,\rho^{\mathrm{T}}z\le y}\lambda||\tilde{\xi}_i-z||_2-1\ge t_i,\ \forall\, i=1,\ldots,H.\end{cases}\tag{33}$$

Since $\lambda\ge0$, $\inf_{z\in\varphi}\lambda||\tilde{\xi}_i-z||_2=0$. Then, the constraint (i) of (33) is equivalent to $t_i\le0$, for every $i=1,\ldots,H$. Moreover, if $\lambda=0$, from the constraint (ii) of (33), $t_i\le-1$, 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 (32). 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

$$\begin{cases} \text{(i)} \quad \theta_W - \frac{1}{H}\sum_{i=1}^{H} g_i \leq l\epsilon, \\ \text{(ii)} \quad \inf_{z\in\varphi, \rho^{\mathrm{T}} z \leq y} ||\tilde{\xi}_i - z||_2 \geq l + g_i, \ \forall\, i = 1, \ldots, H, \\ \text{(iii)} \quad l > 0, \ g_i \leq 0, \ \forall\, i = 1, \ldots, H. \end{cases} \tag{34}$$

This implies that the DRCCMDP (6) is equivalent to (31).

The constraint (ii) of (31) includes inf term which makes it difficult to solve the problem directly. We show that the optimization problem (31) 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

Lemma 5 *If* $\varphi = \mathbb{R}^{|\mathcal{K}|}$,

$$\inf_{\rho^{\mathrm{T}} z \leq y} ||\tilde{\xi}_i - z||_2 = \max\left(0, \frac{\rho^{\mathrm{T}}\tilde{\xi}_i - y}{||\rho||_2}\right), \ \forall\, i = 1, \ldots, H.$$

The proof is given in Appendix C. Using Lemma 5, we have the following result.

Lemma 6 *The optimization problem* (31) *is equivalent to the following optimization problem*

$$\begin{aligned} & \sup \quad y \\ s.t. \quad & \text{(i)} \quad \beta\theta_W - \frac{1}{H}\sum_{i=1}^{H} b_i \leq t\epsilon, \\ & \text{(ii)} \quad \max\left(0, \rho^{\mathrm{T}}\tilde{\xi}_i - y\right) \geq b_i + t, \ \forall\, i = 1, \ldots, H, \\ & \text{(iii)} \quad ||\rho||_2 \leq \beta, \ t \geq 0, \ \beta > 0, \rho \in Q_\alpha(\gamma), \ b_i \leq 0, \ \forall\, i = 1, \ldots, H. \end{aligned} \tag{35}$$

Proof Using Lemma 5, the constraint (ii) of problem (31) can be written as

$$\max\left(0, \frac{\rho^{\mathrm{T}}\tilde{\xi}_i - y}{||\rho||_2}\right) \geq l + g_i, \ \forall\, i = 1, ..., H.$$

Let $\beta > 0$ be an auxiliary variable. Then, under the transformations $t = \beta l$, $b_i = \beta g_i$, for every $i = 1, ..., H$, it is easy to see that (31) is equivalent to (35).

It is clear that a vector $(y, \rho, \beta, (b_i)_{i=1}^{H}, t)$ such that $\rho \in Q_\alpha(\gamma)$, $\beta = ||\rho||_2$, $b_i = 0$, for every $i = 1, \ldots, H$, $t = \frac{\theta_W}{\epsilon}||\rho||_2$ and $y = \min_{i=1,\ldots,H}(\rho^{\mathrm{T}}\tilde{\xi}_i) - \frac{\theta_W}{\epsilon}||\rho||_2$ is a feasible solution of (35). Therefore, the optimal solutions of (35) and the following optimization problem are

the same

$$\begin{aligned}
&\sup \quad y \\
\text{s.t.} \quad &\text{(i)} \quad \beta\theta_W - \frac{1}{H}\sum_{i=1}^{H} b_i \le t\epsilon, \\
&\text{(ii)} \quad \max\left(0, \rho^{\mathrm{T}}\tilde{\xi}_i - y\right) \ge b_i + t,\ \forall\, i = 1, \ldots, H, \\
&\text{(iii)} \quad y \ge \min_{i=1,\ldots,H} (\rho^{\mathrm{T}}\tilde{\xi}_i) - \frac{\theta_W}{\epsilon}||\rho||_2, \\
&\text{(iv)} \quad ||\rho||_2 \le \beta,\ t \ge 0,\ \beta > 0, \rho \in Q_\alpha(\gamma),\ b_i \le 0,\ \forall\, i = 1, \ldots, H.
\end{aligned} \tag{36}$$

We reformulate the problem (36) as an MISOCP problem. In order to do that, we define a constant $M = \left(\frac{\theta_W}{\epsilon} + 2\max_{i=1,\ldots,H} ||\tilde{\xi}_i||_2\right)$ for which the following result holds.

Lemma 7 *For every feasible solution of* (36)*, $M \ge |y - \rho^{\mathrm{T}}\tilde{\xi}_i|$ for all $i = 1, \ldots, H$.*

The proof is given in Appendix D.

Theorem 4 *Consider the* DRCCMDP *problem* (6)*. We assume that the distribution of $\hat{R}$ belongs to the uncertainty set defined by* (27) *and $\varphi = \mathbb{R}^{|\mathcal{K}|}$. Then, the* DRCCMDP (6) *can be reformulated equivalently as the following MISOCP*

$$\begin{aligned}
&\max \quad y \\
\textit{s.t.} \quad &\text{(i)} \quad \beta\theta_W - \frac{1}{H}\sum_{i=1}^{H} b_i \le t\epsilon, \\
&\text{(ii)} \quad M\eta_i \ge b_i + t,\ \forall\, i = 1, \ldots, H, \\
&\text{(iii)} \quad M(1 - \eta_i) + \rho^{\mathrm{T}}\tilde{\xi}_i - y \ge b_i + t,\ \forall\, i = 1, \ldots, H, \\
&\text{(iv)} \quad \eta_i \in \{0, 1\},\ \forall\, i = 1, \ldots, H, \\
&\text{(v)} \quad ||\rho||_2 \le \beta, t \ge 0, \beta > 0, \rho \in Q_\alpha(\gamma), b_i \le 0,\ \forall\, i = 1, \ldots, H.
\end{aligned} \tag{37}$$

Notice that the parameter M is the well known big-M constant.

Proof Since, the distribution of $\hat{R}$ belongs to the uncertainty set defined by (27), the DR-CCMDP problem is equivalent to (36). We show that (36) and (37) are equivalent. It is clear that a vector $(y, \rho, \beta, (b_i)_{i=1}^H, (\eta_i)_{i=1}^H, t)$ such that $\rho \in Q_\alpha(\gamma)$, $\beta = ||\rho||_2, b_i = 0$, $t = \frac{\theta_W}{\epsilon}||\rho||_2$, $\eta_i = 1$, for every $i = 1, \ldots, H$, and $y = \min_{i=1,\ldots,H}(\rho^{\mathrm{T}}\tilde{\xi}_i) - \frac{\theta_W}{\epsilon}||\rho||_2$ is a feasible solution of (37). Therefore, the optimal solution of (37) does not change if we add constraint (38) given below

$$y \ge \min_{i=1,\ldots,H} (\rho^{\mathrm{T}}\tilde{\xi}_i) - \frac{\theta_W}{\epsilon}||\rho||_2, \tag{38}$$

to the feasible region of (37). Now, it is enough to show that the constraint (ii) of (36) is equivalent to (ii) – (iv) of (37). Let the constraint (ii) of (36) be satisfied, i.e.,

$$\max\left(0, \rho^{\mathrm{T}}\tilde{\xi}_i - y\right) \ge b_i + t,\ \forall\, i = 1, \ldots, H. \tag{39}$$

For each $i = 1, \ldots, H$, we consider two cases as follows:

Case 1: If $\max\left(0, \rho^{\mathrm{T}}\tilde{\xi}_i - y\right) = 0$, by choosing $\eta_i = 0$, (39) is equivalent to the constraint (ii) of (37). Moreover, using Lemma 7, we have

$$M \geq |y - \rho^{\mathrm{T}}\tilde{\xi}_i|.$$

Therefore,

$$M(1-\eta_i) + \rho^{\mathrm{T}}\tilde{\xi}_i - y \geq M - |y - \rho^{\mathrm{T}}\tilde{\xi}_i| \geq 0 \geq b_i + t.$$

Case 2: If $\max\left(0, \rho^{\mathrm{T}}\tilde{\xi}_i - y\right) = \rho^{\mathrm{T}}\tilde{\xi}_i - y$, by choosing $\eta_i = 1$, (39) is equivalent to the constraint (iii) of (37). Moreover, using Lemma 7, we have

$$M\eta_i = M \geq \rho^{\mathrm{T}}\tilde{\xi}_i - y \geq b_i + t.$$

This implies that there exists $\eta_i \in \{0, 1\}$ such that (ii) – (iv) of (37) are satisifed. Conversely, suppose (ii) – (iv) of (37) has a feasible solution. If $\eta_i = 1$, the constraint (iii) of (37) implies the constraint (ii) of (36). If $\eta_i = 0$, the constraint (ii) of (37) implies the constraint (ii) of (36).

Remark 2 An MISOCP problem can be solved efficiently with BONMIN, PAJARITO or BARON solvers.

4.2.2 DRCCMDP under Wasserstein distance based uncertainty set with nonnegative support

Lemma 8 *Let $\varphi = \mathbb{R}_+^{|\mathcal{K}|}$ and consider an optimization problem*

$$\inf_{z \in \varphi, \rho^{\mathrm{T}} z \leq y} ||\tilde{\xi}_i - z||_2. \tag{40}$$

The dual problem of (40) *is given by*

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

such that the strong duality holds.

The proof is given in Appendix E.

Theorem 5 *Consider the* DRCCMDP *problem* (6)*. We assume that the distribution of $\hat{R}$ belongs to the uncertainty set defined by* (27) *and $\varphi = \mathbb{R}_+^{|\mathcal{K}|}$. Then, the* DRCCMDP (6) *can be reformulated equivalently as the following biconvex optimization problem*

$$\begin{aligned} & \max \quad y \\ s.t. \quad & \text{(i)} \quad \theta_W - \frac{1}{H}\sum_{i=1}^{H} g_i \leq l\epsilon, \\ & \text{(ii)} \quad \lambda_i(\rho^{\mathrm{T}}\tilde{\xi}_i - y) - \zeta_i^{\mathrm{T}}\tilde{\xi}_i \geq l + g_i,\ \forall\, i = 1, \ldots, H, \\ & \text{(iii)} \quad ||\zeta_i - \lambda_i\rho||_2 \leq 1,\ \forall\, i = 1, \ldots, H, \\ & \text{(iv)} \quad \lambda_i \geq 0, \zeta_i \in \mathbb{R}_+^{|\mathcal{K}|},\ l > 0,\ g_i \leq 0,\ \rho \in \mathcal{Q}_\alpha(\gamma),\ \forall\, i = 1, \ldots, H. \end{aligned} \tag{41}$$

The proof follows directly from Lemma 4 and Lemma 8.

Remark 3 The optimization problem (41) is a non-convex reformulation with biconvex terms. It can be solved by DMCP solver in CVXPY or nonlinear nonconvex optimization solvers, e.g., IPOPT without any guarantee of running time.

5 Machine replacement problem

In this section, we consider a machine replacement problem where a machine in a factory has a life-time of N years. At every stage a maintenance of the machine is scheduled but a factory owner can decide whether to repair or do not repair the machine. There is a high probability that the machine behaves like a new one if it is being repaired and its life gets reduced by a year if it is not being repaired. The factory owner incurs maintenance cost if he decides to repair the machine. It can be modelled as an MDP problem where the life of a machine represents the state of underlying Markov chain, i.e., there are $N + 1$ states. The first state represents a brand new machine. At each state there are two actions: i) "repair", ii) "do not repair". The transition probabilities of the Markov chain with respect to each action is given by Figure 5, where the solid lines and dotted lines correspond to action "repair" and "do not repair", respectively. The maintenance cost corresponding to every state-action pair is not exactly known and is realised after the decision is made. Therefore, it is modelled with a random variable. We assume that for every state action pair (s, a), the maintenance cost is defined as $\hat{c}(s, a) = K + \hat{Z}(s, a)$, where K represents the fixed cost and $\hat{Z}(s, a)$ represents a variable cost which is a random variable. The machine generates a revenue $L(s, a)$ at state-action pair (s, a) and the profit for each $(s, a) \in \mathcal{K}$ is given by

$$\hat{R}(s, a) = L(s, a) - K - \hat{Z}(s, a). \tag{42}$$

The factory owner is interested in maximizing the expected discounted profit. We assume that the factory owner has a finite number of the same machines which are modelled using the same Markov chain. Therefore, we compute the optimal repair policy with respect to a single machine and the same repair policy can be applied for all other machines.

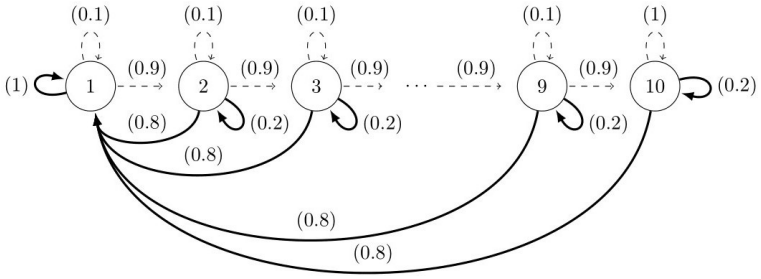

Fig. 1 Machine replacement *MDP* with two actions: "repair" (with solid lines) and "do not repair" (with dashed lines)

All the numerical results below are performed using Python 3.8.8 on an Intel Core i5-1135G7, Processor 2.4 GHz (8M Cache, up to 4.2 GHz), RAM 16G, 512G SSD. 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 set the number of states to 10, the threshold value $\epsilon = 0.1$, the discount parameter $\alpha = 0.85$ and the initial distribution of states γ to be uniformly distributed. For the above instance, $|\mathcal{K}| = 20$ and $\hat{R}$ is a 20×1 random vector with mean vector μ given by

$$\mu(s, a) = L(s, a) - K - \mu_{\hat{Z}}(s, a), \tag{43}$$

where $\mu_{\hat{Z}}$ is the mean vector of the random cost vector $\hat{Z}$. We take $K = 10$, the functio L and the mean cost $\mu_{\hat{Z}}$ corresponding to each state-action pair are summarized in Table 4. For example, at state 1, if the "repair" action is taken, the factory owner has to pay a random cost with mean $\mu_{\hat{Z}}(1,1) = 10$. If the action "do not repair" is taken, the mean value of the random cost is $\mu_{\hat{Z}}(1,2) = 0$. The last state is considered to be risky and not repairing may lead to the machine breakdown. This is the reason we take the mean cost equal to 5 if "do not repair" action is taken at state 10. The covariance matrix Σ of $\hat{R}$ is randomly generated using the following formula

$$\Sigma = \frac{AA^{\mathrm{T}}}{20} + D_{20}, \tag{44}$$

where A is a 20×20 random matrix whose all the entries are real numbers belonging to $[0, 1]$ generated by the command "*A=numpy.random.random(size=(20, 20))*", 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. For the above instance, Σ is diagonally dominant with high values at entries $(10, 10)$ and $(20, 20)$ which is due to the fact that action at risky state can have large variance corresponding to both actions. We use the above μ and Σ for all the moments based uncertainty sets. For ϕ-divergence based uncertainty set, we take the

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	0	30	30
2	10.1	0	30	29.9
3	10.2	0	30	29.8
4	10.3	0	30	29.7
5	10.4	0	30	29.6
6	10.5	0	30	29.5
7	10.6	0	30	29.4
8	10.7	0	30	29.3
9	10.8	0	30	29.2
10	10.9	5	30	29.1

Table 5 Other parameters

Known mean unknown covariance	$\delta_0 = 0.9$
Unknown mean unknown covariance	$\delta_1 = \delta_2 = 1$
$\phi-$divergence	$\theta_\phi = 0.01$
Wasserstein distance	$\theta_W = 0.01$ $H = 1000$

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 (43) and (44), 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 the reference distribution ν. We generate a standard Gaussian vector by the command "*x=numpy.random.normal(0,1,20)*". Using vector x, we generate a Gaussian vector with μ_ν and Σ_ν by using $\tilde{\xi}_i = Bx + \mu_\nu$, where μ_ν and Σ_ν are the mean vector and the covariance matrix defined by (43) and (44), respectively, and B is the Cholesky factorization of Σ_ν. To get the Cholesky factorization of a matrix, we use the command "*numpy.linalg.cholesky*". We summarize the other parameters related to all the uncertainty sets 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 (43) and (44), by solving an equivalent SOCP problem [9]. 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 p is the probability of "repair" action and $1 - p$ is the probability of "do not repair" action. It is clear from Tables 6 and 7 that the optimal repair

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

State(s) \ Optimal policies	CCMDP Gaussian (p,1-p)	Full support known mean known covariance (p,1-p)	Full support known mean unknown covariance (p,1-p)	Full support unknown mean unknown covariance (p,1-p)	$\phi-$divergence (Modified χ^2) (p,1-p)	$\phi-$divergence (variation) (p,1-p)
1	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
2	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
3	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
4	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
5	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
6	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
7	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
8	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)	(0, 1)
9	(0, 1)	(0.64, 0.36)	(0.64, 0.36)	(0.6, 0.4)	(0.27, 0.73)	(0.05, 0.95)
10	(0.9, 0.1)	(0.91, 0.09)	(0.91, 0.09)	(0.91, 0.09)	(0.9, 0.1)	(0.9, 0.1)

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

$\phi-$divergence (Kullbach-Leibler) (p,1-p)	$\phi-$divergence (Hellinger) (p,1-p)	Full support Wasserstein (p,1-p)	Nonnegative known mean known covariance (p,1-p)	Nonnegative known mean unknown covariance (p,1-p)	Nonnegative unknown mean unknown covariance (p,1-p)	Nonnegative Wasserstein (p,1-p)
(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, 1)
(0.25, 0.75)	(0.28, 0.72)	(0.02, 0.98)	(0.62, 0.38)	(0.62, 0.38)	(0.59, 0.41)	(0.01, 0.99)
(0.9, 0.1)	(0.9, 0.1)	(0.9, 0.1)	(0.91, 0.09)	(0.91, 0.09)	(0.91, 0.09)	(0.9, 0.1)

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. This shows that the statistical distance based uncertainty sets give better optimal policy as compared to moments based 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 present the time analysis by considering the number of states for all uncertainty sets between 1000 and 10000. 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 (37) while additional CPU time is required to solve the SDP relaxations of the copositive optimization problem (11) and the biconvex optimization problem (41).

6 Conclusions

We study a DRCCMDP problem under various moments and statistical distance based uncertainty sets defined using ϕ-divergence and Wasserstein distance metric. 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.

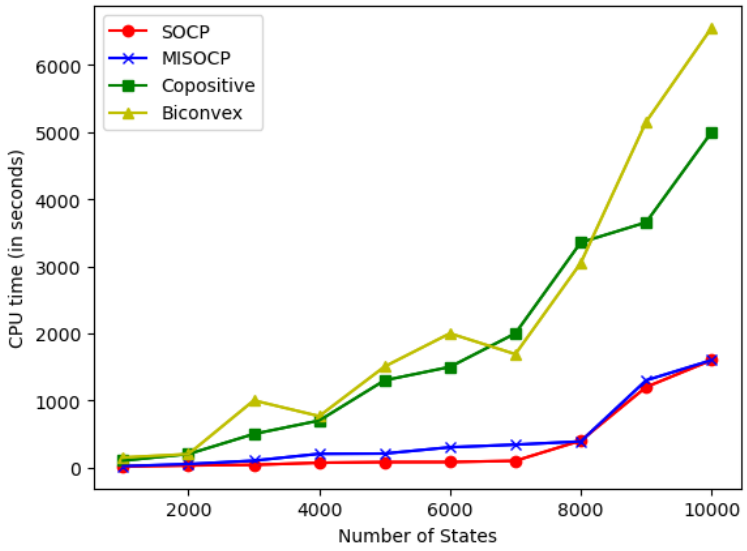

Fig. 2 CPU time (in seconds) to solve SOCP (10) with $\kappa = \sqrt{\frac{1-\epsilon}{\epsilon}}$, MISOCP (37), copositive optimization problem (11) and biconvex optimization problem (41) with different number of states.

Acknowledgement

This research was supported by DST/CEFIPRA Project No. IFC/4117/DST-CNRS-5th call/2017-18/2 and CNRS Project No. AR/SB:2018-07-440.

A Proof of Lemma 1

Consider the optimization problem

$$\begin{aligned} & v_{\mathrm{P}}(\mu,\Sigma) = \sup_{F\in\mathcal{C}^+} \int_{\varphi} \mathbf{1}_{\{\rho^{\mathrm{T}}\hat{R}\leq y\}} \mathrm{d}F(\hat{R}) \\ \text{s.t..}\ & \text{(i)} \quad \int_{\varphi} \mathrm{d}F(\hat{R}) = 1, \\ & \text{(ii)} \quad \int_{\varphi} (\hat{R}-\mu)(\hat{R}-\mu)^{\mathrm{T}} \mathrm{d}F(\hat{R}) = \Sigma, \\ & \text{(iii)} \quad \int_{\varphi} \hat{R}\,\mathrm{d}F(\hat{R}) = \mu, \end{aligned} \tag{45}$$

where $\mathcal{C}^+$ is the set of all positive measures on $\mathbb{R}_+^{|\mathcal{K}|}$. The dual problem of (45) is given by

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

where t, q, and Q are the dual variables associated with the constraints (i), (ii) and (iii) of (45), respectively. In Theorem 3.4 of [8], 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 [37] 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 (46) such that

$$-t_j^* - Q_j^* \circ \Sigma - q_j^{*\mathrm{T}}\mu \to v_{\mathrm{D}}(\mu, \Sigma), \text{ as } j \to \infty. \tag{47}$$

For each $j \in \mathbb{N}$, let $r_j^* = \max(0, q_j^*) - q_j^*$, where $\max(0, q_j^*)$ denotes a $|\mathcal{K}|$-dimensional vector with i^{th} component equal to the maximum value between 0 and the i^{th} component of q_j^*, for every $i = 1, \ldots, |\mathcal{K}|$. 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

$$\begin{aligned} v_{\mathrm{P}}^j(\mu, \Sigma) = \sup_{F\in\mathcal{C}^+} & \int_{\varphi} \mathbf{1}_{\{\rho^{\mathrm{T}} R \le y\}} \mathrm{d}F(R) \\ \text{s.t. (i)} \quad & \int_{\varphi} \mathrm{d}F(R) = 1, \\ \text{(ii)} \quad & \int_{\varphi} (R-\mu)(R-\mu)^{\mathrm{T}} \mathrm{d}F(R) = \Sigma, \\ \text{(iii)} \quad & \mu \le \int_{\varphi} R \mathrm{d}F(R) \le \mu + x_j. \end{aligned} \tag{48}$$

The dual problem of (48) is given by

$$\begin{aligned} v_{\mathrm{D}}^j(\mu, \Sigma) = \inf \quad & -t - Q \circ \Sigma + (r-h)^{\mathrm{T}}\mu + r^T x_j \\ \text{s.t. (i)} \quad & \mathbf{1}_{\{\rho^{\mathrm{T}}\xi \le y\}} + (h-r)^{\mathrm{T}}\xi + \xi^{\mathrm{T}} Q \xi - 2\xi^{\mathrm{T}} Q\mu + \mu^{\mathrm{T}} Q \mu + t \le 0, \ \forall\ \xi \in \mathbb{R}_+^{|\mathcal{K}|}, \\ \text{(ii)} \quad & h, r \in \mathbb{R}_+^{|\mathcal{K}|}, Q \in \mathcal{S}^{|\mathcal{K}|}, \end{aligned} \tag{49}$$

where t, Q, r and h are the dual variables of the constraint (i), (ii) and (iii) of (48), 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 (49). Hence,

$$v_{\mathrm{D}}^j(\mu, \Sigma) \le -t_j^* - Q_j^* \circ \Sigma - q_j^{*\mathrm{T}}\mu + r_j^{*\mathrm{T}} x_j, \ \forall\ j \in \mathbb{N}. \tag{50}$$

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 $\varphi_j := \left\{\xi \parallel \xi \in \mathbb{R}_+^{\mathcal{K}}, \xi \ge \frac{x_j}{2}, \text{ componentwise}\right\}$, defined by

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

It is clear that F_j is a feasible solution of (48) 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 (48) 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 (49) and (46) are equivalent and objective function of (49) has additional positive term. Therefore,

$$v_{\mathrm{D}}^j(\mu, \Sigma) \ge v_{\mathrm{D}}(\mu, \Sigma), \ \forall\ j \in \mathbb{N}. \tag{51}$$

Using (47), (50) and (51) 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 $j \to \infty$.

B Proof of Theorem 3 - Case Hellinger distance

From Table 2, the conjugate of ϕ has the following form

$$\phi^*(r) = \begin{cases} \frac{r}{1-r}, & \text{if } r < 1, \\ \infty, & \text{if } r \ge 1. \end{cases} \tag{52}$$

Let

$$L = \sup_{\lambda>0,\beta\in\mathbb{R}} \left\{ \beta - \lambda\theta_\phi - \lambda\phi^*\left(\frac{-1+\beta}{\lambda}\right)\mathbb{P}_\nu(O) - \lambda\phi^*\left(\frac{\beta}{\lambda}\right)(1-\mathbb{P}_\nu(O)) \right\}. \tag{53}$$

The constraint (i) of (5) is equivalent to

$$L \geq 1-\epsilon. \tag{54}$$

We consider two cases as follows:
Case 1: Let $\frac{\beta}{\lambda} < 1$. Since $\lambda > 0$, the following inequality holds

$$\frac{\beta-1}{\lambda} < \frac{\beta}{\lambda} < 1.$$

From (52), we have

$$\phi^*\left(\frac{\beta}{\lambda}\right) = \frac{\beta}{\lambda-\beta},\ \phi^*\left(\frac{\beta-1}{\lambda}\right) = \frac{\beta-1}{\lambda+1-\beta}.$$

Consequently, it follows from (53) that

$$L = \sup_{\lambda>0,\beta<\lambda} \left\{ \mathbb{P}_\nu(O)\frac{\lambda^2}{(\lambda-\beta)(\lambda-\beta+1)} - \frac{\beta^2}{\lambda-\beta} - \lambda\theta_\phi \right\}.$$

Let $\eta = \lambda - \beta$. Then, we can write

$$L = \sup_{\lambda>0,\eta>0} \left\{ \lambda^2\left(\frac{\mathbb{P}_\nu(O)}{\eta(\eta+1)} - \frac{1}{\eta}\right) + \lambda(2-\theta_\phi) - \eta \right\}.$$

Let $g(\lambda,\eta) = \lambda^2\left(\frac{\mathbb{P}_\nu(O)}{\eta(\eta+1)} - \frac{1}{\eta}\right) + \lambda(2-\theta_\phi) - \eta$. It is a second-order polynomial of λ and the coefficient of λ^2 is negative because $0 \leq \mathbb{P}_\nu(O) \leq 1$ and $\eta > 0$. It is well known that the maximum value of a second order polynomial $f(x) = ax^2 + bx + c$ with $a < 0$ is $c - \frac{b^2}{4a}$ and it holds at $x = \frac{-b}{2a}$. Hence, the maximum value of $g(\lambda,\eta)$ holds at $\lambda^* = \frac{\eta(\eta+1)(2-\theta_\phi)}{2(1+\eta-\mathbb{P}_\nu(O))}$. Since $\theta_\phi < 2$, $\lambda^* > 0$. Therefore, for a given $\eta > 0$, the optimal value L holds at λ^* and $L = c - \frac{b^2}{4a}$, where $c = -\eta$, $b = 2-\theta_\phi$, $a = \frac{\mathbb{P}_\nu(O)}{\eta(\eta+1)} - \frac{1}{\eta}$, which implies that

$$L = \sup_{\eta>0} \left\{ -\eta + \frac{(2-\theta_\phi)^2\eta(\eta+1)}{4(\eta+1-\mathbb{P}_\nu(O))} \right\}. \tag{55}$$

Let $u = \eta + 1 - \mathbb{P}_\nu(O)$, then $\eta > 0$ is equivalent to $u > 1 - \mathbb{P}_\nu(O)$ and we can write

$$\begin{aligned} L &= \sup_{u>1-\mathbb{P}_\nu(O)} \left\{ \left(\frac{(2-\theta_\phi)^2}{4} - 1\right)u + \frac{(2-\theta_\phi)^2\mathbb{P}_\nu(O)(\mathbb{P}_\nu(O)-1)}{4}\frac{1}{u} \right. \\ &\quad \left. + 1 - \mathbb{P}_\nu(O) + \frac{(2-\theta_\phi)^2(2\mathbb{P}_\nu(O)-1)}{4} \right\}, \\ &= \sup_{u>1-\mathbb{P}_\nu(O)} G(u), \end{aligned}$$

where $G(u) = a_1u + \frac{b_1}{u} + c_1$ such that

$$\begin{aligned} a_1 &= \frac{(2-\theta_\phi)^2}{4} - 1,\ b_1 = \frac{(2-\theta_\phi)^2\mathbb{P}_\nu(O)(\mathbb{P}_\nu(O)-1)}{4}, \\ c_1 &= 1 - \mathbb{P}_\nu(O) + \frac{(2-\theta_\phi)^2(2\mathbb{P}_\nu(O)-1)}{4}. \end{aligned}$$

Since $0 < \theta_\phi < 2$ and $0 \leq \mathbb{P}_\nu(O) \leq 1$, $a_1 < 0$ and $b_1 \leq 0$. It is clear that G is decreasing on (u^*, ∞), increasing on $(-u^*, u^*)$ and decreasing on $(-\infty, -u^*)$, where

$$u^* = \sqrt{\frac{b_1}{a_1}} = \sqrt{\frac{(2-\theta_\phi)^2}{4-(2-\theta_\phi)^2}\mathbb{P}_\nu(O)(1-\mathbb{P}_\nu(O))}, \tag{56}$$
$$G(u^*) = a_1 u^* + \frac{b_1}{u^*} + c_1 = -2\sqrt{a_1 b_1} + c_1.$$

If $u^* \leq 1 - \mathbb{P}_\nu(O)$, we deduce that $(1-\mathbb{P}_\nu(O), \infty) \subset (u^*, \infty)$. Since G is decreasing on (u^*, ∞), it implies that G is decreasing on $(1-\mathbb{P}_\nu(O), \infty)$. Hence, the optimal value of G is attained when $u = 1 - \mathbb{P}_\nu(O)$, i.e, $\eta = 0$. From (55), $L = 0$ which violates the constraint (54). Therefore, $u^* > 1 - \mathbb{P}_\nu(O) > 0$. Since, G is decreasing on (u^*, ∞) and increasing on $(1-\mathbb{P}_\nu(O), u^*)$, then $u = u^*$ is the optimal solution of $G(u)$ and $L = -2\sqrt{a_1 b_1} + c_1$. Therefore,

$$\begin{aligned} L =& -2\sqrt{\frac{(2-\theta_\phi)^2}{4}\left(1 - \frac{(2-\theta_\phi)^2}{4}\right)\mathbb{P}_\nu(O)(1-\mathbb{P}_\nu(O))} \\ &+ 1 - \mathbb{P}_\nu(O) + \frac{(2-\theta_\phi)^2(2\mathbb{P}_\nu(O)-1)}{4}. \end{aligned}$$

Then, (54) is rewritten equivalently as follows

$$\begin{aligned} &-2\sqrt{\frac{(2-\theta_\phi)^2}{4}\left(1 - \frac{(2-\theta_\phi)^2}{4}\right)\mathbb{P}_\nu(O)(1-\mathbb{P}_\nu(O))} \\ &\geq \left(1 - \frac{(2-\theta_\phi)^2}{2}\right)\mathbb{P}_\nu(O) + \frac{(2-\theta_\phi)^2}{4} - \epsilon. \end{aligned} \tag{57}$$

By taking the square on both side of (57), we get

$$\begin{aligned} &(2-\theta_\phi)^2\left(1 - \frac{(2-\theta_\phi)^2}{4}\right)\mathbb{P}_\nu(O)(1-\mathbb{P}_\nu(O)) \\ &\leq \left[\left(1 - \frac{(2-\theta_\phi)^2}{2}\right)\mathbb{P}_\nu(O) + \frac{(2-\theta_\phi)^2}{4} - \epsilon\right]^2. \end{aligned} \tag{58}$$

By rewriting (58), we get the following second-order inequality in $\mathbb{P}_\nu(O)$

$$\left(\mathbb{P}_\nu(O)\right)^2 + B\,\mathbb{P}_\nu(O) + C \geq 0,$$

which is equivalent to

$$\left(\mathbb{P}_\nu(O) - x_{\max}\right)\left(\mathbb{P}_\nu(O) - x_{\min}\right) \geq 0, \tag{59}$$

where $x_{\max} = \frac{-B+\sqrt{\Delta}}{2}$, $x_{\min} = \frac{-B-\sqrt{\Delta}}{2}$ and B, C, Δ are given in Table 3. It is clear that (57) is equivalent to either $\mathbb{P}_\nu(O) \geq x_{\max}$ or $\mathbb{P}_\nu(O) \leq x_{\min}$. Moreover, $x_{\max}$ and $x_{\min}$ are solutions of the following two equalities

$$-2\sqrt{\frac{(2-\theta_\phi)^2}{4}\left(1 - \frac{(2-\theta_\phi)^2}{4}\right)x(1-x)} = \left(1 - \frac{(2-\theta_\phi)^2}{2}\right)x + \frac{(2-\theta_\phi)^2}{4} - \epsilon, \tag{60}$$

and

$$2\sqrt{\frac{(2-\theta_\phi)^2}{4}\left(1 - \frac{(2-\theta_\phi)^2}{4}\right)x(1-x)} = \left(1 - \frac{(2-\theta_\phi)^2}{2}\right)x + \frac{(2-\theta_\phi)^2}{4} - \epsilon. \tag{61}$$

Since $\theta_\phi < 2 - \sqrt{2}$, we deduce that $1 - \frac{(2-\theta_\phi)^2}{2} < 0$. Therefore, we have

$$\left(1 - \frac{(2-\theta_\phi)^2}{2}\right)x_{\min} + \frac{(2-\theta_\phi)^2}{4} - \epsilon > \left(1 - \frac{(2-\theta_\phi)^2}{2}\right)x_{\max} + \frac{(2-\theta_\phi)^2}{4} - \epsilon,$$

which implies that $x_{\max}$ is a solution of (60) and $x_{\min}$ is a solution of (61). Hence, the condition $\mathbb{P}_\nu(O) \leq x_{\min}$ implies that

$$\left(1 - \frac{(2-\theta_\phi)^2}{2}\right)\mathbb{P}_\nu(O) + \frac{(2-\theta_\phi)^2}{4} - \epsilon \geq \left(1 - \frac{(2-\theta_\phi)^2}{2}\right)x_{\min} + \frac{(2-\theta_\phi)^2}{4} - \epsilon > 0,$$

which violates the constraint (57). Then, (57) is equivalent to $\mathbb{P}_\nu(O) \geq x_{\max}$, i.e., the constraint (i) of (5) is equivalent to

$$\mathbb{P}_\nu(\rho^T \hat{R} \geq y) \geq \frac{-B + \sqrt{\Delta}}{2}.$$

Case 2: Let $1 \leq \frac{\beta}{\lambda}$. From (52), $\phi^*\left(\frac{\beta}{\lambda}\right) = \infty$, which in turn implies that $L = -\infty$ and it violates the constraint (54).

C Proof of Lemma 5

For each $i = 1, \ldots, H$, we consider two cases as follows:
Case 1: Let $\rho^{\mathrm{T}}\tilde{\xi}_i \leq y$. In this case, it is clear that $\inf_{\rho^{\mathrm{T}}z \leq y} ||\tilde{\xi}_i - z||_2 = 0$ and the optimal value holds at $z = \tilde{\xi}_i$.
Case 2: Let $\rho^{\mathrm{T}}\tilde{\xi}_i > y$. Geometrically, the term $\inf_{\rho^{\mathrm{T}}z \leq y} ||\tilde{\xi}_i - z||_2$ can be interpreted as the distance between $\tilde{\xi}_i$ and the hyper plane $\{z \mid \rho^{\mathrm{T}}z = y\}$. Assume that the optimal value of $\inf_{\rho^{\mathrm{T}}z \leq y} ||\tilde{\xi}_i - z||_2$ holds at $z = z^*$. If $\rho^{\mathrm{T}}z^* < y$, since $\rho^{\mathrm{T}}\tilde{\xi}_i > y$, we deduce that there exists z_0 on $\mathrm{Seg}(z^*, \tilde{\xi}_i)$ such that $\rho^{\mathrm{T}}z_0 = y$, where $\mathrm{Seg}(z^*, \tilde{\xi}_i) := \{z \mid z = z^* + t(\tilde{\xi}_i - z^*), \quad 0 < t < 1\}$. It is clear that $||\tilde{\xi}_i - z^*||_2 > ||\tilde{\xi}_i - z_0||_2$. However, $||\tilde{\xi}_i - z^*||_2 = \inf_{\rho^{\mathrm{T}}z \leq y} ||\tilde{\xi}_i - z||_2$, which gives a contradiction. Therefore, $\rho^{\mathrm{T}}z^* = y$. We can write $\inf_{\rho^{\mathrm{T}}z \leq y} ||\tilde{\xi}_i - z||_2$ equivalently as

$$\begin{aligned} & \inf ||\tilde{\xi}_i - z||_2 \\ \text{s.t.} \quad & \rho^{\mathrm{T}}z = y. \end{aligned} \tag{62}$$

Using the KKT conditions, the optimal solution of (62) satisfies

$$2(\tilde{\xi}_i - z^*) - \lambda\rho = 0, \tag{63}$$

where λ is the Lagrange multiplier associated with the equality constraint. By taking the inner product of (63) with ρ, we have

$$2(\tilde{\xi}_i - z^*)^T\rho - \lambda||\rho||_2^2 = 0,$$

which implies that

$$\lambda = \frac{2(\tilde{\xi}_i - z^*)^T\rho}{||\rho||_2^2}. \tag{64}$$

On the other hand, by taking inner product of (63) with $\tilde{\xi}_i - z^*$, we get

$$2||\tilde{\xi}_i - z^*)||_2^2 - \lambda\rho^T(\tilde{\xi}_i - z^*) = 0. \tag{65}$$

Using (64), (65) and $\rho^T z^* = y$, we have

$$||\tilde{\xi}_i - z^*)||_2 = \frac{\rho^T\tilde{\xi}_i - y}{||\rho||_2}.$$

D Proof of Lemma 7

Let (y, ρ) be a feasible solution of (36) which implies that the constraint (i) of (6) holds. Since, reference distribution ν always belongs to uncertainty set (27), we have

$$\frac{1}{H}\sum_{i=1}^{H}\mathbf{1}_{\{\rho^{\mathrm{T}}\tilde{\xi}_i \le y\}} = \mathbb{P}_{\nu}\left(\rho^{\mathrm{T}}\hat{R} \le y\right) \le \epsilon. \tag{66}$$

It follows from (66) that there exists $\tilde{\xi}_i$ such that $\rho^{\mathrm{T}}\tilde{\xi}_i > y$ which in turn implies that

$$y < \max_{i=1,\ldots,H}(\rho^{\mathrm{T}}\tilde{\xi}_i) < \max_{i=1,\ldots,H}|\rho^{\mathrm{T}}\tilde{\xi}_i| + \frac{\theta_W}{\epsilon}||\rho||_2. \tag{67}$$

Moreover, from the constraint (iii) of (36), we have

$$y \ge \min_{i=1,\ldots,H}(\rho^{\mathrm{T}}\tilde{\xi}_i) - \frac{\theta_W}{\epsilon}||\rho||_2 \ge -\max_{i=1,\ldots,H}|\rho^{\mathrm{T}}\tilde{\xi}_i| - \frac{\theta_W}{\epsilon}||\rho||_2. \tag{68}$$

Using (68) and (67), we get the following inequality

$$|y| + |\rho^{\mathrm{T}}\tilde{\xi}_i| \le 2\max_{i=1,\ldots,H}|\rho^{\mathrm{T}}\tilde{\xi}_i| + \frac{\theta_W}{\epsilon}||\rho||_2, \ \forall\, i = 1,\ldots,H. \tag{69}$$

Using (69), Cauchy-Schwartz inequality, and the fact that ρ is a probability measure, we have

$$|y - \rho^{\mathrm{T}}\tilde{\xi}_i| \le M.$$

E Proof of Lemma 8

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

$$\begin{aligned} & \min \quad t \\ \text{s.t.} \quad & \text{(i)} \quad \rho^{\mathrm{T}}z \le y, \\ & \text{(ii)} \quad t \ge ||\tilde{\xi}_i - z||_2, \\ & \text{(iii)} \quad z \in \mathbb{R}_+^{|\mathcal{K}|}. \end{aligned} \tag{70}$$

The Lagrangian dual problem of (70) is given by

$$\max_{\lambda_i\ge 0,\zeta_i\in\mathbb{R}_+^{|\mathcal{K}|},\beta\ge 0} \ \min_{t\in\mathbb{R},z\in\mathbb{R}^{|\mathcal{K}|}} \mathcal{L}(t,\rho,z,\lambda_i,\beta,\zeta_i),$$

where $\mathcal{L}(t,z,\lambda_i,\beta,\zeta_i) = t + \lambda_i(\rho^{\mathrm{T}}z - y) - \zeta_i^{\mathrm{T}}z + \beta(||\tilde{\xi}_i - z||_2 - t)$ such that λ_i, β and ζ_i are the Lagrange multipliers associated with constraints (i), (ii) and (iii) of (70), respectively. The inner minimization problem can be written as

$$J(\lambda_i,\zeta_i,\beta) = \min_{t\in\mathbb{R},z\in\mathbb{R}^{|\mathcal{K}|}}\left\{t(1-\beta) + \beta||\tilde{\xi}_i - z||_2 + \lambda_i\rho^{\mathrm{T}}z - \zeta_i^{\mathrm{T}}z - \lambda_i y\right\}. \tag{71}$$

It is easy to see that $J(\lambda_i,\zeta_i,\beta) = -\infty$ if $\beta \ne 1$ 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., $\inf_{z\in\mathbb{R}_+^{|\mathcal{K}|},\rho^{\mathrm{T}}z\le y}||\tilde{\xi}_i - z||_2 = -\infty$ which is a contradiction. Therefore, $\beta = 1$ and the dual problem of (70) is given by

$$\max_{\lambda_i\ge 0,\zeta_i\in\mathbb{R}_+^{|\mathcal{K}|}} J(\lambda_i,\zeta_i,1).$$

Using a change of variable $z_1 = \tilde{\xi}_i - z$, we have

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

The above minimization problem is unbounded unless $||\zeta_i - \lambda_i \rho||_2 \leq 1$ and it leads to the following dual problem of (70).

$$\begin{aligned} & \max \quad \lambda_i (\rho^{\mathrm{T}} \tilde{\xi}_i - y) - \zeta_i^{\mathrm{T}} \tilde{\xi}_i \\ \text{s.t.} \quad & \text{(i)} \quad ||\zeta_i - \lambda_i \rho||_2 \leq 1, \\ & \text{(ii)} \quad \lambda_i \geq 0, \zeta_i \in \mathbb{R}_+^{|\mathcal{K}|}. \end{aligned} \tag{72}$$

References

1. Altman, E. *Constrained Markov Decision Processes*. Routledge, New York, 1999.
2. Bagnell, J. A., Ng, A. Y., and Schneider, J. G. Solving uncertain markov decision problems. *Robotics Institute, Carnegie-Mellon University, Journal Contribution* (2001).
3. Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science 59*, 2 (2013), 341–357.
4. Berger, J. O. *The Robust Bayesian Viewpoint*. Department of Statistics, Purdue University, West Lafayette, 1982.
5. Bomze, I. M., Cheng, J., Dickinson, P. J., and Lisser, A. A fresh cp look at mixed-binary qps: new formulations and relaxations. *Mathematical Programming 166*, 1-2 (2017), 159–184.
6. Bomze, I. M., Cheng, J., Dickinson, P. J., Lisser, A., and Liu, J. Notoriously hard (mixed-)binary qps: empirical evidence on new completely positive approaches. *Computational Management Science 16*, 4 (2019), 593–619.
7. Calafiore, G. C., and El Ghaoui, L. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications 130*, 1 (2006), 1–22.
8. Cheng, J., Delage, E., and Lisser, A. Distributionally robust stochastic knapsack problem. *SIAM Journal on Optimization 24*, 3 (2014), 1485–1506.
9. Delage, E., and Mannor, S. Percentile optimization for Markov decision processes with parameter uncertainty. *Operations Research 58*, 1 (2010), 203–213.
10. Delage, E., and Ye, Y. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research 58*, 3 (2010), 595–612.
11. Dupačová, J. The minimax approach to stochastic programming and an illustrative application. *Stochastics 20*, 1 (1987), 73–88.
12. Epstein, L. G., and Schneider, M. Learning under ambiguity. *The Review of Economic Studies 74*, 4 (2007), 1275–1303.
13. Esfahani, P. M., and Kuhn, D. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming 171*, 1 (2018), 115–166.
14. Gallego, G., and Moon, I. The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society 44*, 8 (1993), 825–834.
15. Gao, R., and Kleywegt, A. J. Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research* (2022).
16. Ghaoui, L. E., Oks, M., and Oustry, F. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research 51*, 4 (2003), 543–556.
17. Givan, R., Leach, S., and Dean, T. Bounded-parameter Markov decision processes. *Artificial Intelligence 122*, 1-2 (2000), 71–109.
18. Goh, J., and Sim, M. Distributionally robust optimization and its tractable approximations. *Operations research 58*, 4-part-1 (2010), 902–917.
19. Hanasusanto, G. A., and Kuhn, D. Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Operations Research 66*, 3 (2018), 849–869.
20. Henrion, R. Structural properties of linear probabilistic constraints. *Optimization 56*, 4 (2007), 425–440.
21. Hiriart-Urruty, J.-B., and Seeger, A. A variational approach to copositive matrices. *SIAM Review 52*, 4 (2010), 593–629.
22. Iyengar, G. N. Robust dynamic programming. *Mathematics of Operations Research 30*, 2 (2005), 257–280.

23. Jiang, R., and Guan, Y. Data-driven chance constrained stochastic program. *Mathematical Programming 158*, 1 (2016), 291–327.
24. Liu, J., Lisser, A., and Chen, Z. Distributionally robust chance constrained geometric optimization. *Mathematics of Operations Research 47*, 4 (2022), 2950–2988.
25. Mannor, S., Simester, D., Sun, P., and Tsitsiklis, J. N. Bias and variance in value function estimation. In *Proceedings of the twenty-first international conference on Machine learning* (2004), p. 72.
26. Mannor, S., Simester, D., Sun, P., and Tsitsiklis, J. N. Bias and variance approximation in value function estimates. *Management Science 53*, 2 (2007), 308–322.
27. Nguyen, H. N., Singh, V. V., Lisser, A., and Arora, M. Zero-sum games with distributionally robust chance constraints. In *17th International Conference on Internet and Web Applications and Services (ICIW)* (2022), pp. 7–12.
28. Nilim, A., and El Ghaoui, L. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research 53*, 5 (2005), 780–798.
29. Pardo, L. *Statistical Inference Based on Divergence Measures*. Chapman and Hall/CRC Press, New York, 2018.
30. Pólik, I., and Terlaky, T. A survey of the s-lemma. *SIAM review 49*, 3 (2007), 371–418.
31. Popescu, I. Robust mean-covariance solutions for stochastic optimization. *Operations Research 55*, 1 (2007), 98–112.
32. Puterman, M., et al. *Markov Decision Processes*. John Wiley & Sons, Inc., New York, 1994.
33. Rockafellar, R. Conjugate duality and optimization, regional conference series in appl. *Math. SIAM, Philadelphia* (1974).
34. Rockafellar, R. T. Convex analysis princeton university press. *Princeton, NJ* (1970).
35. Rujeerapaiboon, N., Kuhn, D., and Wiesemann, W. Chebyshev inequalities for products of random variables. *Mathematics of Operations Research 43*, 3 (2018), 887–918.
36. Scarf, H. E. *A Min-Max Solution of an Inventory Problem*. Rand Corporation, Santa Monica Office, California, 1957.
37. Shapiro, A. On duality theory of conic linear problems. In *Semi-Infinite Programming. Nonconvex Optimization and Its Applications*, M. Á. Goberna and M. A. López, Eds., vol. 57. Springer US, Boston, MA, 2001, pp. 135–165.
38. Shapiro, A., Dentcheva, D., and Ruszczynski, A. *Lectures on Stochastic Programming: Modeling and Theory, Third Edition*. MOS-SIAM, Philadelphia, 2021.
39. Singh, V. V., Jouini, O., and Lisser, A. Distributionally robust chance-constrained games: Existence and characterization of Nash equilibrium. *Optimization Letters 11*, 7 (2017), 1385–1405.
40. Sion, M. On general minimax theorems. *Pacific Journal of Mathematics 8*, 1 (1958), 171–176.
41. Varagapriya, V., Singh, V. V., and Lisser, A. Constrained Markov decision processes with uncertain costs. *Operations Research Letters 50* (2022), 218–223.
42. Varagapriya, V., Singh, V. V., and Lisser, A. Joint chance-constrained Markov decision processes. *Annals of Operations Research* (2022).
43. Villani, C. *Optimal Transport*. Springer, Berlin, Heidelberg, 2009.
44. Villani, C. *Topics in Optimal Transportation*. American Mathematical Society, Providence, 2021.
45. Wang, Z., Glynn, P. W., and Ye, Y. Likelihood robust optimization for data-driven problems. *Computational Management Science 13*, 2 (2016), 241–261.
46. White III, C. C., and Eldeib, H. K. Markov decision processes with imprecise transition probabilities. *Operations Research 42*, 4 (1994), 739–749.
47. Wiesemann, W., Kuhn, D., and Rustem, B. Robust Markov decision processes. *Mathematics of Operations Research 38*, 1 (2012), 153–183.
48. Xu, G., and Burer, S. A copositive approach for two-stage adjustable robust optimization with uncertain right-hand sides. *Computational Optimization and Applications 70*, 1 (2018), 33–59.
49. Yu, Y.-L., Li, Y., Schuurmans, D., and Szepesvári, C. A general projection property for distribution families. *Advances in Neural Information Processing Systems 22* (2009).
50. Žáčková, J. On minimax solutions of stochastic linear programming problems. *Časopis Pro Pěstování Matematiky 91*, 4 (1966), 423–430.
51. Zhao, C., and Guan, Y. Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters 46*, 2 (2018), 262–267.
52. Zymler, S., Kuhn, D., and Rustem, B. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming 137*, 1 (2013), 167–198.