Risk Estimation in a Markov Cost Process: Lower and Upper Bounds

Gugan Thoppe [1] **Prashanth L. A.** [2] **Sanjay P. Bhat** [3]

Abstract

We tackle the problem of estimating risk measures of the infinite-horizon discounted cost of a Markov cost process. The risk measures we study include variance, Value-at-Risk (VaR), and Conditional Value-at-Risk (CVaR). First, we show that estimating any of these risk measures with ϵ-accuracy, either in expected or high-probability sense, requires at least $\Omega(1/\epsilon^2)$ samples. Then, using a truncation scheme, we derive an upper bound for the CVaR and variance estimation. This bound matches our lower bound up to logarithmic factors. Finally, we discuss an extension of our estimation scheme that covers more general risk measures satisfying a certain continuity criterion, such as spectral risk measures and utility-based shortfall risk. To the best of our knowledge, our work is the first to provide lower and upper bounds for estimating any risk measure beyond the mean within a Markovian setting. Our lower bounds also extend to the infinite-horizon discounted costs' mean. Even in that case, our lower bound of $\Omega(1/\epsilon^2)$ improves upon the existing $\Omega(1/\epsilon)$ bound (Metelli et al., 2023).

1. Introduction

In a traditional discounted reinforcement learning (RL) problem (Bertsekas & Tsitsiklis, 1996; Sutton & Barto, 2018), the objective is to maximize the value function, which is the expected value of the infinite-horizon cumulative discounted cost. However, optimizing only the expected value is not appealing in several practical applications. For instance, in the financial domain, strategists like to consider the risk of investments. Similarly, in transportation, road users are sensitive to the variations in the delay incurred and would especially like to avoid a large delay even if it occurs infrequently. Risk-sensitive RL addresses such applications by incorporating a risk measure in the optimization process, either in the objective or as a constraint.

Going beyond the expected value, three well-known risk measures are variance, Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) (Rockafellar & Uryasev, 2000). For a Cumulative Distribution Function (CDF) $\mathcal{F}$, the VaR $v_\alpha(\mathcal{F})$ and CVaR $c_\alpha(\mathcal{F})$ at a given level $\alpha \in (0,1)$ is defined by

$$v_\alpha(\mathcal{F}) = \inf\{\xi : \mathbb{P}[X \leq \xi] \geq \alpha\}, \text{ and} \tag{1}$$

$$c_\alpha(\mathcal{F}) = v_\alpha(X) + \frac{\mathbb{E}[X - v_\alpha(X)]^+}{1-\alpha}, \tag{2}$$

where $X \sim \mathcal{F}$. From the above, it is apparent that $v_\alpha(\mathcal{F})$ is a certain quantile of the CDF $\mathcal{F}$. For a (strictly increasing) continuous distribution $\mathcal{F}$, $v_\alpha(\mathcal{F}) = \mathcal{F}^{-1}(\alpha)$, while CVaR can be equivalently written as $c_\alpha(\mathcal{F}) = \mathbb{E}[X \mid X \geq v_\alpha(X)]$ for $X \sim \mathcal{F}$. In words, CVaR is the expected value of X, conditioned on the event that X exceeds the VaR. Choosing a α close to 1 and taking X as modeling the losses of a financial position, CVaR can be understood as the expected loss given that losses have exceeded a certain threshold (specified by a quantile). In the financial domain, CVaR is preferred over VaR because CVaR is a coherent risk measure (Artzner et al., 1999), while VaR is not. In particular, with VaR as the risk measure, combining two investment portfolios (diversification) can cause an increase in the risk—a property that is undesirable in finance.

Estimation of VaR and CVaR using independent and identically distributed (i.i.d.) samples has received a lot of attention recently in the literature, cf. (Bhat & Prashanth, 2019; Brown, 2007; Wang & Gao, 2010; Thomas & Learned-Miller, 2019; Kolla et al., 2019a; Prashanth et al., 2020). The concentration bounds for CVaR have been useful in deriving sample complexity results in the context of empirical risk minimization and bandit applications.

In this paper, we are concerned with the problem of estimating a risk measure from a sample path of a discounted Markov Cost Process (MCP). In the context of RL, this is equivalent to the policy evaluation problem, albeit for a risk measure. For this problem, we derive minimax sample

[1]Dept. of Computer Science and Automation, Indian Institute of Science (IISc), Bengaluru, India; Robert Bosch Centre for Data Science and Artificial Intelligence, IIT Madras, Chennai, India. [2]Dept. of Computer Science and Engineering, Indian Institute of Technology Madras, Chennai, India. [3]TCS Research, Hyderabad, India.. Correspondence to: Gugan Thoppe <gthoppe@iisc.ac.in>, Prashanth L. A. <prashla@cse.iitm.ac.in>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

complexity lower bounds as well as upper bounds. As a parallel to the i.i.d. sampling framework mentioned above, the upper bound we derive is useful in the design and analysis of risk-sensitive policy gradient algorithms (see (Prashanth & Fu, 2022) for a recent survey).

Our key contributions are summarized below. Tables 1 and 2 provide a summary of the bounds we have derived.

1. **Lower bounds**: We derive a minimax sample complexity lower bound of $\Omega(1/\epsilon^2)$ for risk estimation in two types of MCP problem instances: one with deterministic costs and the other with stochastic costs. In either case, our bounds are order optimal and the first of its kind for risk estimation. Our first bound applies to VaR and CVaR of the infinite-horizon discounted cost, while the second applies *even* to its mean and variance.
2. **Lower bound proof**: We obtain our two lower bounds via novel proof techniques. In the deterministic costs case, the bound is obtained by identifying a 'hard' problem instance, involving a two-state Markov chain with a cost function that suitably diverges as $\epsilon \to 0$, and then solving a constrained optimization problem. Our proof builds upon (Metelli et al., 2023) which looks at a lower bound for mean estimation, the analogous optimization problem there, though, is unconstrained. With stochastic costs, we show that our lower bound holds even when the cost mean is bounded, as a function of ϵ. The 'hard' problem instance in this case involves a MCP with a single state and a Gaussian single-stage cost function.
3. **Upper bound**: Using a truncated horizon estimation scheme, we derive upper bounds of $\widetilde{\mathcal{O}}\left(\frac{1}{\epsilon^2}\right)$ for CVaR and variance estimation. These bounds match our corresponding lower bounds up to logarithmic factors.
4. **Other risk measures**: Finally, we also propose an extension of the estimation scheme to cover risk measures that satisfy a certain Lipschitz continuity criterion. Prominent risk measures that are covered in this extension are spectral risk measures (Acerbi, 2002) and utility-based shortfall risk (Föllmer & Schied, 2002).

We now compare our contributions to (Metelli et al., 2023), which is the closest related work. In the aforementioned reference, the authors derive a minimax sample-complexity lower bound of $\Omega\left(\frac{1}{\epsilon}\right)$ in a probabilistic sense for estimating the mean of the infinite-horizon discounted cost of an MCP; see row 1 of Table 2. Their proof involves a two-state Markov chain with $\{0,1\}$ rewards. In this setting, the mean of the cumulative discounted cost can be explicitly written as a function of the transition probabilities. In contrast, the proofs of our lower bounds are more challenging owing to the lack of a closed form expression for the risk measures we consider. Moreover, our lower bounds, when specialized to mean estimation, leads to an improvement in comparison to (Metelli et al., 2023); see Remarks 3.3 and A.1 for details.

The rest of the paper is organized as follows: Section 2 provides the problem formulation. Section 3 presents our two lower bounds for estimating VaR, CVaR, and variance. Section 4 presents our upper bounds for estimation of the above risk measures and also a general class of Lipschitz risk measures. Our lower and upper bounds are of two types: 1) 'in expectation' and 2) 'with high probability'. Section 5 provides proof outlines for the above results, leaving the details to the appendix. Finally, we conclude in Section 6.

2. Problem Formulation and Preliminaries

We formally provide all our notations, describe our setup, and state our research goals here. We also give a detailed description of a general risk estimation algorithm.

Notations: $\mathcal{U}$ denotes an arbitrary (possibly infinite) set, and $\mathfrak{F}$ a σ-algebra on $\mathcal{U}$ containing all its singleton subsets. For $\mathcal{S} \subseteq \mathcal{U}$, $\mathfrak{F}_{\mathcal{S}} := \{A \cap \mathcal{S} : A \in \mathfrak{F}\}$ is the induced σ-algebra on $\mathcal{S}$. Also, $\mathscr{P}(\mathcal{S})$, $\mathscr{P}(\mathbb{R})$, and $\mathscr{P}(\{0,1\})$ denote the sets of probability measures on $(\mathcal{S}, \mathfrak{F}_{\mathcal{S}})$, $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and $(\{0,1\}, 2^{\{0,1\}})$, respectively, where $\mathcal{B}(\mathbb{R})$ is the Borel-σ-algebra on $\mathbb{R}$, while $2^{\{0,1\}}$ is the power set of $\{0,1\}$. Furthermore, $\mathscr{B}(\mathcal{S})$ denotes the set of $\mathcal{S} \mapsto \mathscr{P}(\mathbb{R})$ functions.

The tuple (M, f) denotes a MCP. Here, $M \equiv (\mathcal{S}, P, \nu)$ is a Markov chain on the state-space $\mathcal{S} \subseteq \mathcal{U}$ with transition kernel $P : \mathcal{S} \to \mathscr{P}(\mathcal{S})$ and initial state distribution $\nu \in \mathscr{P}(\mathcal{S})$, while $f \in \mathscr{B}(\mathcal{S})$ is a (possibly stochastic) cost function. Further, $P(\cdot|s)$ is the distribution that s maps to under P. Separately, $(s_n)_{n \in \mathbb{Z}_+}$ denotes a trajectory of M. Also, $\gamma \in [0,1)$ is the discount factor. The cumulative discounted cost over a t-length horizon is the random variable

$$\sum_{n=0}^{t-1} \gamma^n f_n, \text{ where } f_n \sim f(s_n). \tag{3}$$

Let $\mathcal{F}_t(M, f)$ be the CDF of this random variable. Finally,

$$\mathscr{M} := \{(M, f) : \mathcal{F}_t(M, f) \text{ converges weakly to a CDF}\},$$

and, for $(M, f) \in \mathscr{M}$,

$$\mathcal{F}(M, f) := \lim_{t \to \infty} \mathcal{F}_t(M, f), \tag{4}$$

where the limit is in the sense of weak convergence. Also, for $(M, f) \in \mathscr{M}$, the expressions $\mu(M, f)$, $\mathrm{V}(M, f)$, $v_\alpha(M, f)$, and $c_\alpha(M, f)$, with $\alpha \in (0,1)$, denote the mean, variance, VaR and CVaR (both at the confidence level $\alpha \in (0,1)$) of the CDF $\mathcal{F}(M, f)$, respectively.

Setup: We presume we have access to an MCP $(M, f) \in \mathscr{M}$, where M and f are unknown, but whose state and single-stage cost trajectory can be observed. We further

Table 1. Summary of the sample complexity lower and upper bounds, in an expected sense, for estimating various risk measures. For a given $\epsilon > 0$, sample complexity is the number of sample transitions N such that the estimation error $\mathbb{E}|\hat{\eta}_n - \eta(D)| < \epsilon$ for all $n \geq N$. Here η is the risk measure, D is the cumulative discounted cost, and $\hat{\eta}_N$ the risk estimate. Here $\tilde{O}(\cdot)$ is a variant of the big-O notation that ignores logarithmic factors.

Bound type	Risk measure	Sample complexity	Reference
Lower bound	Mean, VaR, CVaR, variance	$\Omega\left(\frac{1}{\epsilon^2}\right)$	Theorems 3.1 and 3.5
Upper bound	CVaR	$\widetilde{\mathcal{O}}\left(\frac{1}{\epsilon^2}\right)$	Theorem 4.1
Upper bound	Lipschitz risk measure	$\widetilde{\mathcal{O}}\left(\frac{1}{\epsilon^2}\right)$	Theorem 4.10
Upper bound	Variance	$\widetilde{\mathcal{O}}\left(\frac{1}{\epsilon^2}\right)$	Theorem 4.12

Table 2. Summary and comparison of sample complexity lower and upper bounds, in a high-probability sense, for various risk measures. For a given $\epsilon > 0$ and $\delta \in (0,1)$, sample complexity is the number of state transitions N such that $\mathbb{P}\{|\hat{\eta}_N - \eta(D)| > \epsilon\} \leq \delta$. Here ϵ denotes the estimation accuracy (and is presumed to be below some threshold to ensure N has a simple dependence on ϵ) , δ is the confidence, η the risk measure, D is the cumulative discounted cost, and $\hat{\eta}_N$ the risk estimate using N sample transitions of the MCP. In the bounds below, T is the truncation parameter used in our risk estimator, γ is the discount factor, and K is an upper bound on the costs of the MCP. In the bounds, the constants c, c' vary between rows. In the fourth row, L denotes the Lipschitz constant for a risk measure (see (17) below). Except the first row, other rows concern our work.

Bound type	Risk measure	Bound on $\mathbb{P}\{\lvert\hat{\eta}_N - \eta(D)\rvert > \epsilon\}$	Sample complexity	Reference
Lower bound	Mean	$\exp\left(-\frac{cN\epsilon}{1-\epsilon}\right)$	$\Omega\left(\epsilon^{-1}\ln\left(\frac{1}{\delta}\right)\right)$	(Metelli et al., 2023)
Lower bound	Mean, VaR, CVaR	$\exp\left(-c\sqrt{N\epsilon^2}\right)$	$\Omega\left(\epsilon^{-2}\ln(\frac{1}{\delta})\right)$	Theorems 3.1, 3.5
Upper bound	VaR, CVaR	$\exp\left(-\frac{cN}{T}\left(\epsilon - \frac{\gamma^T K}{1-\gamma}\right)^2\right)$	$\widetilde{\mathcal{O}}\left(\epsilon^{-2}\ln(\frac{1}{\delta})\right)$	Theorems 4.4, 4.8
Upper bound	Lipschitz risk measure	$\exp\left[-\frac{cN}{T}\left[\frac{1}{L}\left[\epsilon - \frac{\gamma^T K}{1-\gamma}\right] - \frac{c'\sqrt{T}}{\sqrt{N}}\right]^2\right]$	$\widetilde{\mathcal{O}}\left(\epsilon^{-2}\ln(\frac{1}{\delta})\right)$	Theorem 4.11
Upper bound	Variance	$\exp\left(-\frac{cN}{T}\left(\epsilon - \frac{2\gamma^T K^2}{(1-\gamma)^2}\right)^2\right)$	$\widetilde{\mathcal{O}}\left(\epsilon^{-2}\ln(\frac{1}{\delta})\right)$	Theorem 4.12

presume that this MCP can be restarted from a state sampled from the initial state distribution as many times as we want. Such a reset mechanism is feasible, e.g., in a simulation optimization setting (Fu, 2015).

We now describe a general risk estimation algorithm: one that can be used to estimate a desired risk measure $\eta(M, f)$.

Definition 2.1 (Risk Estimation Algorithm)**.** A risk estimation algorithm $\mathscr{A}$ is a tuple $(\boldsymbol{r}, \hat{\eta})$ made up of a reset policy $\boldsymbol{r}$ and an estimator $\hat{\eta}$.

Loosely, $\boldsymbol{r}$ specifies the rule to reset the given Markov chain M, i.e., stop its natural evolution under P, and restart it from a state sampled from ν. On the other hand, $\hat{\eta}$ describes how to transform the observations of states, resets, and single-stage costs into an estimate of the given risk measure.

We now formally define the above two terms as in (Metelli et al., 2023). For $n \in \mathbb{Z}_+$ (the set of non-negative integers), let $\mathcal{H}_n := (\mathcal{U} \times \{0,1\})^n$ be the set of all possible histories of states and reset decisions until time n.

Definition 2.2 (Reset Policy)**.** A reset policy $\boldsymbol{r} \equiv (\boldsymbol{r}_n)_{n\in\mathbb{Z}_+}$ on $\mathcal{S}$ is a sequence of functions where $\boldsymbol{r}_n : \mathcal{H}_n \times \mathcal{U} \mapsto \mathscr{P}(\{0,1\})$ maps $H \in \mathcal{H}_n$ and $s \in \mathcal{U}$ to a distribution $\boldsymbol{r}_n(\cdot|H, s) \in \mathscr{P}(\{0,1\})$.

Definition 2.3 (Estimator)**.** An estimator $\hat{\eta} \equiv (\hat{\eta}_n)_{n\in\mathbb{Z}_+}$ is a sequence of functions where $\hat{\eta}_n : \mathcal{H}_n \times \mathscr{B}(\mathcal{S}) \mapsto \mathbb{R}$ maps $H \in \mathcal{H}_n$ and $f \in \mathscr{B}(\mathcal{S})$ to a real number representing the estimate of $\eta(M, f)$. Specifically, $\hat{\eta}_n$ is a function of $s_0, Y_0, f_0, \ldots, s_{n-1}, Y_{n-1}, f_{n-1}$, where $f_i \sim f(s_i)$, i.e., the history of states, resets, and the observed costs.

Next, we describe the evolution dynamics of a Markov chain M under the reset policy $\boldsymbol{r}$.

Definition 2.4 (Resetted Chain)**.** Let $M \equiv (\mathcal{S}, P, \nu)$ be a Markov chain, and $\boldsymbol{r}$ a reset policy. Then, the corresponding resetted chain $M^{\boldsymbol{r}}$ is a stochastic process $(s_n, Y_n)_{n\in\mathbb{Z}_+}$ taking values in $\mathcal{S} \times \{0, 1\}$ such that

1) the initial state s_0 is sampled from ν;

2) for all $n \in \mathbb{Z}_+$, the reset decision $Y_n \in \{0, 1\}$ is drawn from $r_n(\cdot|H_n, s_n)$ and the subsequent state s_{n+1} from $Y_n\nu + (1 - Y_n)P(\cdot|s_n)$, where $H_n := (s_0, Y_0, \ldots, s_{n-1}, Y_{n-1}) \in \mathcal{H}_n$;

and these samples are drawn with independent randomness. The joint distribution of H_n is denoted by $P^n_{M,\boldsymbol{r}}$.

Remark 2.5. The sequence $(s_n)_{n\in\mathbb{Z}_+}$ in $M^{\boldsymbol{r}}$ satisfies

$$\begin{aligned}&\mathbb{P}(s_{n+1} \in \mathcal{B}|H_n, s_n)\\ &= \boldsymbol{r}_n(\{1\}|H_n, s_n)\nu(\mathcal{B}) + \boldsymbol{r}_n(\{0\}|H_n, s_n)P(\mathcal{B}|s_n) \quad (5)\end{aligned}$$

for any $\mathcal{B} \in \mathfrak{F}_{\mathcal{S}}$. As discussed in (Metelli et al., 2023), this process is non-Markovian and non-stationary when the reset distribution $\boldsymbol{r}_n$ depends on the history H_n.

Research goals: Obtain lower and upper bounds on the samples needed to obtain an ϵ-accurate estimate of a risk measure $\eta(M, f)$ related to the unknown underlying MCP (M, f). Formally, our goal is to first obtain bounds on

$$\inf_{\mathscr{A}\equiv(\hat{\eta},\boldsymbol{r})} \sup_{(M,f)\in\mathscr{M}} \mathbb{E}|\hat{\eta}_n(H_n, f) - \eta(M, f)| \quad (6)$$

and

$$\inf_{\mathscr{A}} \sup_{(M,f)} \mathbb{P}\left\{|\hat{\eta}_n(H_n, f) - \eta(M, f)| \geq \epsilon\right\} \quad (7)$$

as a function of n and ϵ, where $\mathbb{E}$ and $\mathbb{P}$ are with respect to $H_n \sim P^n_{M,\boldsymbol{r}}$, and $\eta(M, f)$ is either $\mu(M, f), \mathrm{V}(M, f)$, or one of $v_\alpha(M, f)$ and $c_\alpha(M, f)$ for a given $\alpha \in (0, 1)$. We then aim to use these results to compute the desired sample complexity bounds.

3. Lower Bounds: Risk Estimation Error

We obtain risk estimation lower bounds for two different MCP problem instances: first, with deterministic costs and, second, with stochastic costs. These two cases are discussed in Subsections 3.1 and 3.2, respectively.

3.1. Deterministic Costs

Throughout this subsection, we presume $f : \mathcal{S} \to \mathbb{R}$, i.e., f assigns a fixed real-valued cost to each state.

Theorem 3.1 (Lower bound). *For an MCP $(M, f) \in \mathscr{M}$, let the risk measure $\eta(M, f)$ be either $\mathcal{F}(M, f)$'s VaR $v_\alpha(M, f)$ or CVaR $c_\alpha(M, f)$ at a given $\alpha \in (0, 1)$. Then, for every $n \in \mathbb{N}$, error threshold $\epsilon > 0$, and discount factor $\gamma \in [0, 1)$, we have that*

$$\inf_{\mathscr{A}} \sup_{(M,f)} \mathbb{P}\{|\hat{\eta}_n(H_n, f) - \eta(M, f)| \geq \epsilon\} \geq \exp\left[-n\epsilon^2 \ln\left(\frac{1}{\alpha}\right) \ln\left(\frac{1}{\gamma}\right)\right] \quad (8)$$

and

$$\inf_{\mathscr{A}} \sup_{(M,f)} \mathbb{E}|\hat{\eta}_n(H_n, f) - \eta(M, f)| \geq \frac{\exp\left[-\ln\alpha \ln\gamma\right]}{\sqrt{n}}. \quad (9)$$

Proof. See Section 5. □

Remark 3.2. By substituting $\epsilon = -\ln\delta/\sqrt{n \ln\alpha \ln\gamma}$ in (8) for a $\delta \in (0, 1)$, it follows that

$$\inf_{\mathscr{A}} \sup_{(M,f)} \mathbb{P}\left\{|\hat{\eta}_n(H_n, f) - \eta(M, f)| \geq \frac{-\ln\delta}{\sqrt{n \ln\alpha \ln\gamma}}\right\} \geq \delta.$$

From the above, it is apparent that the number of samples should be at least $\Omega(\epsilon^{-2})$ to guarantee an ϵ-accurate VaR or CVaR estimate with probability $1 - \delta$. Similarly, (9) says that any algorithm would need $\Omega(\epsilon^{-2})$ samples to guarantee an ϵ-accurate VaR or CVaR estimate in an expected sense.

Remark 3.3. In Theorem 4.1 of (Metelli et al., 2023), the authors claim a $\Omega(1/\sqrt{n})$ minimax lower bound for estimating the mean of the infinite horizon discounted cost. However, this bound is misleading because the lower bound is derived assuming a certain quantity σ_f^2 is a constant. On closer inspection of their proof, it is apparent that $\sigma_f^2 = \epsilon(1 - \epsilon)$, where ϵ is the estimation accuracy; see (34) in the Appendix there. When this ϵ dependence is factored in, it can be seen that the lower bound is only $\Omega(1/n)$; see our discussion in Remark A.1 in the Appendix for additional details.

Remark 3.4. For a given $\epsilon > 0$, we derive Theorem 3.1 by constructing two MCP instances with states A and B. In the first instance, the single-stage costs at these two states, i.e., $f_0(A)$ and $f_0(B)$, are both 0. In the second, the cost at state A (i.e., $f_1(A)$) is $2\epsilon \exp(1/\epsilon^2)$, while that at state B is 0. We remark that we can get a lower bound of $\exp(-n\epsilon^k c)$, $k > 2$, by choosing $f_1(A)$ appropriately. This is not surprising since the bound now applies to bigger class of cost functions. Note that the bound in (Metelli et al., 2023) applies to the class of bounded costs (w.r.t. ϵ), but see Remark 3.3.

3.2. Stochastic Costs

In Section 3.1, we looked at the case where the cost function f is deterministic. However, to get the $\Omega(\epsilon^{-2})$ sample complexity, we require that a suitable single-stage cost ($f_1(A)$) increase as ϵ decays to 0; see Remark 3.4. In this subsection, we show that, by allowing the single-stage costs to be stochastic, we can get similar lower bounds even when these costs have bounded mean. To derive these bounds, we use a radically novel proof idea to that of Theorem 3.1 and also to the one used in (Metelli et al., 2023).

Theorem 3.5 (Lower bound). *For an MCP $(M, f) \in \mathscr{M}$, let the risk measure $\eta(M, f)$ be either $\mathcal{F}(M, f)$'s VaR $v_\alpha(M, f)$ or CVaR $c_\alpha(M, f)$ at a given $\alpha \in (0, 1)$, or $\mathcal{F}(M, f)$'s mean $\mu(M, f)$ or its variance $\mathrm{V}(M, f)$. Then, for every $n \in \mathbb{N}$ and $\epsilon \in (0, 1/(2\sqrt{Kn}]$,*

$$\inf_{\mathscr{A}} \sup_{(M,f)} \mathbb{P}\left\{|\hat{\eta}_n(H_n, f) - \eta(M, f)| \geq \epsilon\right\} \geq \frac{1}{2}\exp\left(-2\sqrt{Kn\epsilon^2}\right), \quad (10)$$

and

$$\inf_{\mathscr{A}} \sup_{(M,f)} \mathbb{E}|\hat{\eta}_n(H_n, f) - \eta(M, f)| \geq \frac{1}{8\sqrt{Kn}}, \quad (11)$$

where $K = 6$ if $\eta(M, f) = \mathrm{V}(M, f)$, and 2 otherwise.

Proof. See Section 5. □

Remark 3.6. By substituting $\epsilon = \ln[1/(2\delta)]/\sqrt{8n}$ in (10), we have

$$\inf_{\mathscr{A}} \sup_{(M,f)} \mathbb{P}\left\{|\hat{\eta}_n(H_n, f) - \eta(M, f)| \geq \frac{\ln[1/(2\delta)]}{\sqrt{8n}}\right\} \geq \delta. \quad (12)$$

Hence, as in Remark 3.2, the sample complexity in the high-probability and in the expected sense is $\Omega(\epsilon^{-2})$.

4. Upper Bounds for Risk Estimation Error

4.1. Upper Bound: VaR and CVaR

In this section, we first present an estimation scheme for the VaR and CVaR of the cumulative discounted cost in an infinite-horizon Markov chain. Subsequently, we provide an estimation scheme for more general risk measures that satisfy a certain continuity criterion.

We first present the classic estimators for VaR and CVaR of a random variable X using m i.i.d. samples, which are denoted by $\{X^1, \ldots, X^m\}$. Define the empirical distribution function (EDF) $F_m(\cdot)$ as follows:

$$F_m(x) = \frac{1}{n}\sum_{i=1}^{m} \mathbb{I}\left\{X^i \leq x\right\}, \forall x \in \mathbb{R}.$$

Using EDF F_m, VaR and CVaR estimators, $\hat{v}_{m,\alpha}$ and $\hat{c}_{m,\alpha}$, respectively, are formed as follows:

$$\hat{v}_{m,\alpha} = F_m^{-1}(\alpha) = X^{[\lceil m\alpha \rceil]}, \text{ and}$$
$$\hat{c}_{m,\alpha} = \frac{1}{m(1-\alpha)}\sum_{i=1}^{m} X^i \mathbb{I}\left\{X^i \geq \hat{v}_{m,\alpha}\right\}, \quad (13)$$

where $X^{[i]}$ denotes the ith order statistic, for $i = 1, \ldots, m$.

In the infinite-horizon discounted Markov chain setting that we consider in this paper, we estimate the VaR and CVaR of the cumulative discounted cost $D(f)$ (4) using a truncated estimator[1]. More precisely, given a budget of N transitions, we reset after every T steps to obtain $m = \lceil \frac{N}{T} \rceil$ trajectories. Since each trajectory provides an independent truncated sample of the cumulative cost D, we use the m samples obtained from the distribution of D to form the VaR and CVaR estimates using (13). More precisely, let $D^1, \ldots, D^m$ denote the cumulative discounted cost samples obtained from the m trajectories over the finite horizon T. With these samples, we form estimates $\hat{v}_N$ and $\hat{c}_N$ of VaR and CVaR of cumulative discounted cost r.v D.

Theorem 4.1 (Bound in expectation for CVaR). *Assume $|f(s)| \leq K$ for all $s \in \mathcal{S}$. Let $\hat{c}_N$ denote the CVaR estimator formed using (13) with a sample path of N transitions and truncation parameter T. Let $m = \lceil \frac{N}{T} \rceil$. Then, we have*

$$\mathbb{E}\left|\hat{c}_N - c_\alpha(D)\right| \leq \frac{32(1-\gamma^T)^2K^2}{(1-\alpha)(1-\gamma)^2\sqrt{m}} + \frac{\gamma^T K}{1-\gamma}. \quad (14)$$

Proof. See Section 5. □

Remark 4.2. Comparing the bound in expectation with the lower bound in (11), it is apparent that the bounds match in terms of the dependence on the length n of the sample path, modulo an additional factor of $\frac{\gamma^T K}{1-\gamma}$ in the upper bound. The latter can be made small by choosing a larger truncation parameter T.

Remark 4.3. For a given ϵ, choose T such that $\frac{\gamma^T K}{1-\gamma} < \frac{\epsilon}{2}$. Such a T is $O\left(\log(1/\epsilon)\right)$. Separately, using the relation $1 - \gamma^T \leq 1$, it can be seen that the first term on RHS of (14) would fall below $\frac{\epsilon}{2}$ when $m = O\left(\frac{1}{\epsilon^2}\right)$. Now, since $N \leq mT$, we have that the CVaR estimation scheme has a sample complexity $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$, matching the corresponding result in the lower bound up to logarithmic factors.

The result below establishes an exponential concentration result for the CVaR estimate $\hat{c}_N$.

Theorem 4.4 (Concentration bound for CVaR). *Under the assumptions of Theorem 4.1, for every $\epsilon > \frac{\gamma^T K}{1-\gamma}$,*

$$\mathbb{P}\left[|\hat{c}_N - c_\alpha(D)| > \epsilon\right] \leq 6\exp\left(-\frac{m(1-\alpha)(1-\gamma)^2}{11(1-\gamma^T)^2K^2}\left(\epsilon - \frac{\gamma^T K}{1-\gamma}\right)^2\right), \quad (15)$$

where $m = \lceil \frac{N}{T} \rceil$.

Proof. See Section 5. □

[1] For ease of notation, we drop the dependence on the cost function f, and use D to denote the cumulative discounted cost random variable

Remark 4.5. Suppose $\tau := \frac{(1-\alpha)(1-\gamma)^2}{11K^2}$. Using $(1-\gamma^T) \leq 1$, the bound in (15) leads to

$$\mathbb{P}\left[|\hat{c}_N - c_\alpha(D)| > \epsilon\right] \leq 6\exp\left(-m\tau\left(\epsilon - \frac{\gamma^T K}{1-\gamma}\right)^2\right)$$

Choosing T, m, and, hence, N along the lines discussed in Remark 4.3 shows that, for $N = \tilde{O}(\epsilon^{-2})$, we have $\mathbb{P}\left[|\hat{c}_N - c_\alpha(D)| > \epsilon\right] \leq \delta$.

Remark 4.6. The tail bound in (15) can be inverted to arrive at the following high-confidence form: given $\delta \in (0,1)$, with probability (w.p.) $1-\delta$, we have

$$|\hat{c}_N - c_\alpha(D)| \leq \frac{K}{1-\gamma}\left[(1-\gamma^T)\sqrt{\frac{11\log(\frac{6}{\delta})}{m(1-\alpha)}} + \gamma^T\right]. \tag{16}$$

Remark 4.7. Comparing the high-confidence bound in (16) with the lower bound in (12), it is apparent that the bounds match in terms of the dependence on the length n of the sample path. However, in terms of dependence on δ, it can be seen that there is a gap as the lower bound has a $\log\frac{1}{\delta}$ factor, while the upper bound has $\sqrt{\log\frac{1}{\delta}}$. We believe the lower bound can be improved to close this δ-gap.

The last result of this section is a concentration bound for the VaR estimate $\hat{v}_N$, which is obtained by using a truncation parameter T. In the i.i.d. case, establishing a tail bound for VaR is difficult for distributions that are flat around the VaR, and a minimum growth rate assumption is usually imposed on the underlying distribution to overcome this problem, cf. (Kolla et al., 2019b; Prashanth et al., 2020). In the case of the truncated estimator $\hat{v}_N$, the concentration bound is arrived by first relating the VaR of the truncated random variable D_T and discounted cumulative cost D, followed by an application of an i.i.d. concentration bound for VaR of D_T. For this approach to work, we require a minimum growth assumption on the distribution of D_T, which is formalized below.

(A1) The r.v. D_T is continuous with a density, say f_T satisfying the following for some $\eta, \zeta > 0$: $f_T(x) > \ell$ for all $x \in \left[v_\alpha(D_T) - \frac{\zeta}{2}, v_\alpha(D_T) + \frac{\zeta}{2}\right]$.

Theorem 4.8 (Concentration bound for VaR). *Under (A1) and the assumptions of Theorem 4.1, for every* $\epsilon > \frac{\gamma^T K}{1-\gamma}$, *we have*

$$\mathbb{P}\left[|\hat{v}_N - v_\alpha(D)| > \epsilon\right] \leq 2\exp\left(-2m\ell^2\min\left\{\left(\epsilon - \frac{\gamma^T K}{1-\gamma}\right)^2, \zeta^2\right\}\right),$$

where $m = \lceil \frac{N}{T} \rceil$.

Proof. See Section 5. □

4.2. Upper Bound: Lipschitz Risk Measures

In this section, we extend the truncation-based estimation scheme presented earlier to cover risk measures that satisfy a continuity criterion, which is made precise below.

Definition 4.9. Let $(\mathcal{L}, W_1)$ denote the metric space of distributions with the 1-Wasserstein distance as the metric W_1. A risk measure $\eta(\cdot)$ is said to be Lipschitz-continuous if there exists $L > 0$ such that, for any two distributions $F, G \in \mathcal{L}$, the following holds:

$$|\eta(F) - \eta(G)| \leq LW_1(F,G). \tag{17}$$

In (Prashanth & Bhat, 2022), the authors establish that optimized certainty equivalent (OCE) risk (Ben-Tal & Teboulle, 1986; 2007) that includes CVaR, spectral risk measure (Acerbi, 2002), and utility-based shortfall risk (Föllmer & Schied, 2002) belong to the class of Lipschitz risk measures, see Lemmas 12, 13 and 15 in (Prashanth & Bhat, 2022).

As before, let D denote the cumulative cost, and $D^1, \ldots, D^m$ denote the samples obtained from m truncated trajectories. Let $\eta(D)$ be a Lipschitz risk measure. Using the samples with EDF F_m, we form the following estimate:

$$\hat{\eta}_N = \eta(F_m). \tag{18}$$

For the case of spectral risk measure and utility-based shortfall risk, the estimate defined above coincides with their classic estimators, cf. (Hu & Zhang, 2018; Prashanth & Bhat, 2022). Moreover, the CVaR estimator defined in (13) is a special case of (18).

The result below provides a bound in expectation for the estimator (18) of a Lipschitz risk measure.

Theorem 4.10 (Bound in expectation). *Assume* $|f(s)| \leq K$ *for all* $s \in \mathcal{S}$. *Let* $\eta(\cdot)$ *denote a Lipschitz risk measure, which satisfies the following properties: (i)* $\eta(X+a) = \eta(X) + a$ *for any* $a \in \mathbb{R}$; *and (ii)* $\eta(X) \leq \eta(Y)$ *if* $X \leq Y$ *almost surely. Let* η_N *denote the estimator* (18) *formed using a sample path of* N *transitions, with truncated horizon* T. *Let* $m = \lceil \frac{N}{T} \rceil$. *Then, we have*

$$\mathbb{E}\left|\hat{\eta}_N - \eta(D)\right| \leq \frac{32L(1-\gamma^T)^2K^2}{(1-\gamma)^2\sqrt{m}} + \frac{\gamma^T K}{1-\gamma}.$$

Proof. Follows in a similar manner as the proof of Theorem 4.1. The reader is referred to Appendix C for the details. □

The result below provides a concentration bound for the estimator $\hat{\eta}_N$ defined in (18).

Theorem 4.11 (Concentration bound). *Under conditions of Theorem 4.10, for every* ϵ *such that* $\frac{512K}{(1-\gamma)\sqrt{m}} < \frac{1}{L}\left(\epsilon - \frac{\gamma^T K}{1-\gamma}\right) < \frac{512K}{(1-\gamma)\sqrt{m}} + 16K\sqrt{\mathsf{e}}$, *with* e *denoting the*

Euler constant and $m = \lceil \frac{N}{T} \rceil$*, we have*

$$\mathbb{P}\left[|\hat{\eta}_N - \eta(D)| > \epsilon\right] \leq 2\exp\left[-\frac{2m(1-\gamma)^2}{256(1-\gamma^T)^2K^2\mathsf{e}} \times \left[\frac{1}{L}\left[\epsilon - \frac{\gamma^T K}{1-\gamma}\right] - \frac{512K}{(1-\gamma)\sqrt{m}}\right]^2\right].$$

Proof. Follows in a similar manner as the proof of Theorem 4.4. The reader is referred to Appendix D for the details. □

4.3. Upper Bound: Variance

Recall the truncation-based scheme from Section 4.1, where $m = \lceil \frac{N}{T} \rceil$ independent trajectories were obtained with the truncation parameter T. From these trajectories, we obtain samples $D^1, \ldots, D^m$ of the truncated discounted cost $D_T = \sum_{t=0}^{T-1} \gamma^t f(s_t)$. Using these samples, we form the estimate $\hat{\mathrm{V}}_N$ of the variance $\mathrm{V}(D)$ of the cumulative discounted cost as follows:

$$\hat{\mathrm{V}}_N = \frac{1}{m-1}\sum_{i=1}^{m}\left(D^i - \bar{\kappa}_m\right)^2, \text{ where } \bar{\kappa}_m = \frac{1}{m}\sum_{i=1}^{m} D^i. \tag{19}$$

Notice that variance is not a Lipschitz risk measures in the sense of Definition 4.9. Thus, the result in Theorem 4.11 does not apply for variance. However, using the same proof idea, we can infer a concentration bound for the variance estimator in (19). This result is presented below.

Theorem 4.12 (Concentration Bound). *Assume* $|f(s)| \leq K$ *for all* $s \in \mathcal{S}$*. Let* $\hat{\mathrm{V}}_N$ *denote the variance estimator formed using* (19)*. Let* $m = \lceil \frac{N}{T} \rceil$ *and suppose* $m \geq 2$*. Then,*

$$\mathbb{E}|\hat{\mathrm{V}}_N - V(D)| \leq \frac{8(1-\gamma^T)^2K^2}{\sqrt{m}(1-\gamma)^2} + \frac{4\gamma^T K^2}{(1-\gamma)^2}. \tag{20}$$

In addition, for every $\epsilon > \frac{4\gamma^T K^2}{(1-\gamma)^2}$*, we have*

$$\mathbb{P}\left[|\hat{\mathrm{V}}_N - \mathrm{V}(D)| > \epsilon\right] \leq 2\exp\left(-\frac{m(1-\gamma)^4}{32(1-\gamma^T)^4K^4}\left(\epsilon - \frac{4\gamma^T K^2}{(1-\gamma)^2}\right)^2\right). \tag{21}$$

Proof. See Appendix E. □

5. Proofs and Proof Sketches

Proof Outline of Theorem 3.1. We give a brief sketch of our proof, and refer the reader to Appendix A for the details.

We only show how we derive our lower bounds for $\eta(M, f) = v_\alpha(M, f)$. This proof is later extended to cover CVaR. Our proof for the VaR result has three main steps.

1. **MCP construction**: Let A and B be two arbitrary but distinct elements of $\mathcal{U}$. The MCP we construct is (M_*, f_*), where $M_* \equiv (\mathcal{S}_*, P_*, \nu_*)$ is the two-state Markov chain with $\mathcal{S}_* = \{A, B\}$, the transition matrix

$$P_* = \begin{pmatrix} p & 1-p \\ p & 1-p \end{pmatrix}$$

for some $p \in [0, 1]$, and the initial state distribution $\nu_* \equiv (q, 1-q)$ for some $q \in [0, 1]$; this Markov chain is shown in Figure 1 in the appendix. We let the single-stage cost function f_* be either f_0 or f_1, where $f_0(A) = f_0(B) = f_1(B) = 0$, while $f_1(A) = 2\epsilon \exp\left(\frac{1}{\epsilon^2}\right)$. Clearly, for $\alpha \in [0, 1]$, $v_\alpha(M_*, f_0) = 0$, while $v_\alpha(M_*, f_1) \geq 0$.

2. **Lower bounding the minimax error**: Let $\mathscr{A} \equiv (\eta, \boldsymbol{r})$ be any VaR estimation algorithm. Then, for any p and q such that $v_\alpha(M_*, f_1) = v_\alpha(M_*, f_1) - v_\alpha(M_*, f_0) \geq 2\epsilon$, we show that

$$\sup_{M,f} \mathbb{P}_{H_n \sim P^n_{M,\boldsymbol{r}}}\left(|\hat{\eta}(H_n, f) - v_\alpha(M, f)| \geq \epsilon\right) \geq \frac{1}{2}\min\{1-q, 1-p\}^n. \tag{22}$$

3. **Tightening the bound**: Since (22) is true for any p and q such that $v_\alpha(M_*, f_1) \geq 2\epsilon$, we can now optimize over these values. This leads to following optimization problem:

$$\begin{aligned} \max_{p,q} \quad & \min\{1-q, 1-p\} \\ \text{s.t.} \quad & 0 \leq p \leq 1, \quad 0 \leq q \leq 1, v_\alpha(M_*, f_1) \geq 2\epsilon \end{aligned} \tag{23}$$

While steps 1 and 2 above are similar to those employed for deriving lower bounds for the expected value objective in (Metelli et al., 2023), step 3 involves significant deviations owing to the inequality constraint on the VaR in (23). Such a constraint is not present in the case of the expected value, making the solution of the corresponding optimization problem simpler. Also, unlike the expected value case, the optimal p and q cannot be inferred in closed form, and our proof involves a non-trivial argument using the VaR statistic to arrive at values for p and q that suitably lower bound the max-min optimization problem in (23) above. □

Proof of Theorem 3.5. We give the full proof here. We consider two variants of the problem instance given by the MCP (M_0, f), where

1) the Markov chain is $M_0 \equiv (\mathcal{S}, P, \nu)$ with $\mathcal{S}$ being an arbitrary but fixed singleton subset of $\mathcal{U}$ (say $\{s\}$), implying that P and ν are trivial distributions; and

2) the single-stage cost function is

$$f(s) \sim \mathcal{N}((1-\gamma)\mu, (1-\gamma^2)\sigma^2), \tag{24}$$

for some unknown $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.

In general, for any estimation algorithm $\mathscr{A} \equiv (\hat{\eta}, \boldsymbol{r})$, its estimate at time $n \in \mathbb{Z}_+$ depends fully on the history $H_n \equiv (s_0, Y_0, \ldots, s_{n-1}, Y_{n-1})$ and the single-stage costs $f_0, \ldots, f_{n-1}$, where f_i has the same distribution as $f(s_i)$ in (24). However, in the case of (M_0, f), the state-space $\mathcal{S}$ underlying M_0 is trivial. Therefore, $s_0 = \cdots = s_{n-1} = s$. Similarly, whatever be the reset decision at any time n, the algorithm gets to only see state s and the associated random cost (which is independent of everything else) at n. Hence, the estimate $\hat{\eta}(H_n, f)$ is essentially a function of $f_0, \ldots, f_{n-1}$ i.e.,

$$\hat{\eta}(H_n, f) \equiv \hat{\eta}(f_0, \ldots, f_{n-1}). \tag{25}$$

Separately, the structure of (M_0, f) implies that $f_0, \ldots, f_{n-1}$ are IID samples with the distribution in (24).

We break the rest of our proof into three parts. First, we show that the infinite-horizon cumulative discounted cost's CDF $\mathcal{F}(M_0, f)$ (see (4)) exists for the above MCP and obtain expressions for its mean, variance, VaR, and CVaR. Second, we establish a relationship between $\mathcal{F}(M_0, f)$ and the CDF of $f(s)$ given in (24). Third, we build upon the discussions in (Duchi, 2024) to obtain the stated lower bounds.

Clearly, the n-horizon CDF $\mathcal{F}_n(M_0, f)$, $n \in \mathbb{N}$, equals the CDF of $\sum_{t=0}^{n-1} \gamma^t X_t$, where (X_t) is an IID sequence of random variables having the same distribution as $f(s)$. However, the CDF of $\sum_{t=0}^{n-1} \gamma^t X_t$ is that of $\mathcal{N}((1-\gamma^n)\mu, (1-\gamma^{2n})\sigma^2)$. Further, its limit $\mathcal{F}(M_0, f)$, as $n \to \infty$, exists and is the CDF of $\mathcal{N}(\mu, \sigma^2)$. Hence,

$$\begin{aligned} \mu(M_0, f) &= \mu, \mathrm{V}(M_0, f) = \sigma^2, \\ v_\alpha(M_0, f) &= \mu + \sigma \Phi^{-1}(\alpha), \\ c_\alpha(M_0, f) &= \mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}, \end{aligned} \tag{26}$$

where ϕ and Φ are the PDF and CDF of the standard normal distribution, respectively.

We next discuss a relationship between the samples of $\mathcal{F}(M_0, f)$ and the distribution of $f(s)$. Let

$$\begin{aligned} a &= \frac{1-\gamma+\sqrt{(1-\gamma)(1+3\gamma)}}{2} \text{ and} \\ b &= \frac{1-\gamma-\sqrt{(1-\gamma)(1+3\gamma)}}{2}. \end{aligned} \tag{27}$$

Then, for any two independent samples Z and Z' of $\mathcal{F}(M_0, f)$, i.e., of $\mathcal{N}(\mu, \sigma^2)$, the sum

$$aZ + bZ' \sim \mathcal{N}((1-\gamma)\mu, (1-\gamma^2)\sigma^2), \tag{28}$$

i.e., it has the same distribution as $f(s)$; see (24).

Thus, any function $\hat{\eta}$ that uses the n single-stage costs $f_0, \ldots, f_{n-1}$ to estimate $\eta(M, f)$ gives rise to a hypothetical function $\hat{\eta}'$ that operates on $2n$ IID samples $Z_0, \ldots, Z_{2n-1}$ of the limiting distribution $\mathcal{F}(M, f)$, i.e., of $\mathcal{N}(\mu, \sigma^2)$, through the following two steps:

1) combine every successive pair Z_{2i} and Z_{2i+1}, $i \in \{0, \ldots, n-1\}$, to get $X_i = aZ_{2i} + bZ_{2i+1}$, where a and b are as in (27);

2) use $\hat{\eta}(X_0, \ldots, X_{n-1})$ to get the final estimate.

In other words,

$$\hat{\eta}'(Z_0, \ldots, Z_{2n-1}) = \hat{\eta}(X_0, \ldots, X_{n-1}). \tag{29}$$

Finally, we derive our stated lower bound. The proofs for mean, VaR, and CVaR are similar. Hence, we discuss them together now. Consider two different variants (M_0, f^{+1}) and (M_0, f^{-1}) of the MCP (M_0, f), where $f^\nu(s)$, $\nu \in \{-1, +1\}$, is as in (24), but with μ replaced by $\mu_\nu := \nu\epsilon$ for some $\epsilon > 0$, and $\sigma^2 = 1$. It follows from (26) that, for any of our quantities of interest, i.e., mean, VaR, or CVaR,

$$\eta(M_0, f^{+1}) - \eta(M_0, f^{-1}) = 2\epsilon. \tag{30}$$

For $\nu \in \{-1, +1\}$, let P_ν denote the distribution $\mathcal{N}(\mu_\nu, \sigma^2)$. Further, let P_ν^k denote the joint distribution of $Z_0, \ldots, Z_{k-1}$, sampled independently from $\mathcal{N}(\mu_\nu, \sigma^2)$. Also, let $\|\cdot\|_{\mathrm{TV}}$ denote the total variation distance and D_{kl} the KL-divergence. Then, for every $n \in \mathbb{N}$ and $\epsilon \in (0, 1/\sqrt{8n}]$, we have

$$\begin{aligned} &\inf_{\mathscr{A}} \sup_{M,f} \mathbb{P}\{|\hat{\eta}(H_n, f) - \eta(M, f)| \geq \epsilon\} && (31) \\ &\geq \inf_{\mathscr{A}} \sup_{\nu=-1,1} \mathbb{P}\{|\hat{\eta}(H_n, f^\nu) - \eta(M_0, f^\nu)| \geq \epsilon\} && (32) \\ &= \inf_{\hat{\eta}} \sup_{\nu=-1,1} \mathbb{P}\{|\hat{\eta}(f_0^\nu, \ldots, f_{n-1}^\nu) - \eta(M_0, f^\nu)| \geq \epsilon\} && (33) \\ &\geq \inf_{\hat{\eta}'} \sup_{\nu=-1,1} \mathbb{P}\{|\hat{\eta}'(Z_0, \ldots, Z_{2n-1}) - \eta(M_0, f^\nu)| \geq \epsilon\} && (34) \\ &= \frac{1}{2}\left[1 - \|P_{+1}^{2n} - P_{-1}^{2n}\|_{\mathrm{TV}}\right] && (35) \\ &\geq \frac{1}{2}\left[1 - \sqrt{\frac{1}{2} D_{\mathrm{kl}}(P_{+1}^{2n}, P_{-1}^{2n})}\right] && (36) \\ &\geq \frac{1}{2}\left[1 - \sqrt{n D_{\mathrm{kl}}(P_{+1}, P_{-1})}\right] && (37) \\ &= \frac{1}{2}[1 - \sqrt{2n\epsilon^2}] && (38) \\ &\geq \frac{1}{2} \exp\left(-\frac{\sqrt{2n\epsilon^2}}{1-\sqrt{2n\epsilon^2}}\right) && (39) \\ &\geq \frac{1}{2} \exp\left(-2\sqrt{2n\epsilon^2}\right), && (40) \end{aligned}$$

which gives the relation in (10), as desired. Above, (32) holds since we only consider two specific problem instances, (33) holds due to (25), (34) follows from (29) by noting that the set of hypothetical estimators η' is a superset of the set of all practically realizable estimators η, (35) follows from

Eqs. (8.2.1) and (8.3.1) of (Duchi, 2024), (36) follows from Pinsker's inequality, (37) follows from the independence of $Z_0, \ldots, Z_{2N-1}$, (38) follows using the KL divergence formula for two Gaussian distributions, (39) follows since $2n\epsilon^2 \leq 1$ and $1 - x \geq e^{-x/(1-x)}$ for every $x \leq 1$, while (40) holds due to the constraint on ϵ. Also, note that the $\mathbb{P}$ in (31) is with respect to $H_n \sim P^n_{M,\boldsymbol{r}}$, while it is with respect to $P^n_{M_0,\boldsymbol{r}}$ in (32). Similarly, the $\mathbb{P}$ in (34) is with respect to $Z_0, \ldots, Z_{2n-1}$ sampled in an IID fashion from $\mathcal{N}(\mu_\nu, 1)$.

We now prove (11). Since, for any random variable X,

$$\mathbb{E}|X| \geq \mathbb{E}\big[|X|\, \mathbb{1}\{|X| \geq \epsilon\}\big] \geq \epsilon\, \mathbb{P}\{|X| \geq \epsilon\},$$

it follows from (38) that

$$\inf_{\mathscr{A}} \sup_{(M,f)} \mathbb{E}|\hat{\eta}_n(H_n, f) - \eta(M, f)| \geq \frac{\epsilon}{2}[1 - \sqrt{2n\epsilon^2}] \quad (41)$$

for every $\epsilon \in (0, 1/\sqrt{2n}]$. By substituting $\epsilon = 1/\sqrt{8n}$, the desired relation in (11) follows.

The lower bounds for $\eta(M, f) = \mathrm{V}(M, f)$ similarly follow. The details are given in Section B in the appendix. □

Proof of Theorem 4.1. Recall $D_T = \sum_{t=0}^{T-1} \gamma^t f(s_t)$ and $D = \sum_{t=0}^{\infty} \gamma^t f(s_t)$. Since $|f(\cdot)| \leq K$, we obtain

$$-\frac{\gamma^T K}{1-\gamma} \leq D - D_T \leq \frac{\gamma^T K}{1-\gamma}. \quad (42)$$

It is well-known that CVaR is a coherent risk measure (Rockafellar & Uryasev, 2000), in particular, satisfying the following two properties for any two random variables X and Y: (i) $c_\alpha(X + a) = c_\alpha(X) + a$ for every $a \in \mathbb{R}$; and (ii) $c_\alpha(X) \leq c_\alpha(Y)$ if $X \leq Y$ almost surely.

Using these properties in conjunction with (42), we obtain

$$c_\alpha(D) - \frac{\gamma^T K}{1-\gamma} \leq c_\alpha(D_T) \leq c_\alpha(D) + \frac{\gamma^T K}{1-\gamma}. \quad (43)$$

Using the bound above, we have

$$\begin{aligned}\mathbb{E}\,|\hat{c}_N - c_\alpha(D)| &\leq \mathbb{E}\,|\hat{c}_N - c_\alpha(D_T)| + \mathbb{E}\,|c_\alpha(D_T) - c_\alpha(D)| \\ &\leq \frac{32(1-\gamma^T)^2 K^2}{(1-\alpha)(1-\gamma)^2\sqrt{m}} + \frac{\gamma^T K}{1-\gamma},\end{aligned}$$

where the final inequality uses (43) to bound the second term on the RHS of the first inequality, while the first term there is bounded using a special case of (Prashanth & Bhat, 2022, Corollary 20), which is stated below for the sake of completeness:

Lemma 5.1. *Let $X^1, \ldots, X^m$ be drawn i.i.d. from the distribution of a random variable X, satisfying $|X| \leq B$ a.s. Let $\hat{c}_{m,\alpha}$ be formed using (13). Then,*

$$\mathbb{E}\,|\hat{c}_{m,\alpha} - c_\alpha(X)| \leq \frac{32B^2}{(1-\alpha)\sqrt{m}}. \quad (44)$$

The application of the result above to bound $\mathbb{E}\,|\hat{c}_N - c_\alpha(D_T)|$ is valid since the truncated trajectories are independent, and $|D_T| \leq \frac{(1-\gamma^T)K}{1-\gamma}$ a.s. □

Proof of Theorem 4.4. The initial passage in the proof of Theorem 4.1 leading up to (43) holds. Using the inequalities in (43), we arrive at the main claim as follows:

$$\begin{aligned}&\mathbb{P}\left[|\hat{c}_N - c_\alpha(D)| > \epsilon\right] \\ &= \mathbb{P}\left[|\hat{c}_N - c_\alpha(D_T) + c_\alpha(D_T) - c_\alpha(D)| > \epsilon\right] \\ &\leq \mathbb{P}\left[|\hat{c}_N - c_\alpha(D_T)| > \epsilon - \frac{\gamma^T K}{1-\gamma}\right] \\ &\leq 6\exp\left(-\frac{m(1-\alpha)(1-\gamma)^2}{44(1-\gamma^T)^2K^2}\left(\epsilon - \frac{\gamma^T K}{1-\gamma}\right)^2\right),\end{aligned}$$

where the penultimate inequality used (43), while the final inequality follows from the CVaR concentration bound in Theorem 3.1 of (Wang & Gao, 2010). □

Proof of Theorem 4.8. As in the CVaR case, we have (i) $v_\alpha(X + a) = v_\alpha(X) + a$ for every $a \in \mathbb{R}$; and (ii) $v_\alpha(X) \leq v_\alpha(Y)$ if $X \leq Y$ almost surely. Using these facts, we obtain

$$v_\alpha(D) - \frac{\gamma^T K}{1-\gamma} \leq v_\alpha(D_T) \leq v_\alpha(D) + \frac{\gamma^T K}{1-\gamma}. \quad (45)$$

For the sake of completeness, we provide a VaR concentration bound for the i.i.d. case from (Prashanth et al., 2020):

Lemma 5.2. *Let $X^1, \ldots, X^m$ be drawn i.i.d. from the distribution of a random variable X, satisfying (A1). Let $\hat{v}_{m,\alpha}$ be formed using (13). Then, for any $\epsilon > 0$, we have*

$$\mathbb{P}\left[|\hat{v}_{m,\alpha} - v_\alpha(X)| \geq \epsilon\right] \leq 2\exp\left(-2n\ell^2 \min\left(\epsilon^2, \zeta^2\right)\right),$$

where ℓ, ζ are specified in (A1).

The claim follows follows by using the inequality in (45) in conjunction with Lemma 5.2. □

6. Conclusions and Future Work

We studied estimation of risk measures such as VaR, CVaR, mean and variance in an infinite-horizon discounted MCP. We provided minimax lower bounds for estimating an ϵ-accurate solution, both in a high-probability and expected sense. We have also proposed a truncation-based estimator for the aforementioned risk measures, and obtained an upper bound on its sample complexity.

As future work, it would be interesting to extend the upper bounds to cover sub-Gaussian distributions. An orthogonal future research direction is to derive lower bounds that capture the dependence on the mixing time of the underlying Markov chain.

Acknowledgements

We express our gratitude to the anonymous reviewers for taking their time and providing us with valuable comments that improved the quality of the paper. Gugan Thoppe's work was supported in part by DST-SERB's Core Research Grant CRG/2021/008330, CEFIPRA's Indo-French Grant 7102-1, the Walmart Center for Tech Excellence, the Kotak-IISc AI/ML Centre, and by the Pratiksha Trust Young Investigator Award.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Acerbi, C. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518, 2002.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

Ben-Tal, A. and Teboulle, M. Expected utility, penalty functions, and duality in stochastic nonlinear programming. *Management Science*, 32(11):1445–1466, November 1986. ISSN 0025-1909.

Ben-Tal, A. and Teboulle, M. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17:449–476, 02 2007.

Bertsekas, D. P. and Tsitsiklis, J. N. *Neuro-Dynamic Programming*. Athena Scientific, 1996.

Bhat, S. P. and Prashanth, L. A. Concentration of risk measures: A Wasserstein distance approach. In *Advances in Neural Information Processing Systems*, pp. 11739–11748, 2019.

Brown, D. B. Large deviations bounds for estimating conditional value-at-risk. *Operations Research Letters*, 35(6): 722–730, 2007.

Duchi, J. Lectures notes on statistics and information theory. https://web.stanford.edu/class/stats311/lecture-notes.pdf, 2024. Accessed: 2024-02-02.

Föllmer, H. and Schied, A. Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4):429–447, 2002.

Fu, M. C. (ed.). *Handbook of Simulation Optimization*. Springer, 2015.

Hu, Z. and Zhang, D. Utility-based shortfall risk: Efficient computations via monte carlo. *Naval Research Logistics (NRL)*, 65(5):378–392, 2018.

Kolla, R. K., Prashanth, L., Bhat, S. P., and Jagannathan, K. Concentration bounds for empirical conditional value-at-risk: The unbounded case. *Operations Research Letters*, 47(1):16 – 20, 2019a.

Kolla, R. K., Prashanth, L. A., Bhat, S. P., and Jagannathan, K. P. Concentration bounds for empirical conditional value-at-risk: The unbounded case. *Operations Research Letters*, 47(1):16–20, 2019b.

Metelli, A. M., Mutti, M., and Restelli, M. A tale of sampling and estimation in discounted reinforcement learning. In *International Conference on Artificial Intelligence and Statistics*, pp. 4575–4601. PMLR, 2023.

Prashanth, L. and Bhat, S. P. A Wasserstein Distance Approach for Concentration of Empirical Risk Estimates. *Journal of Machine Learning Research*, 23(238):1–61, 2022.

Prashanth, L. A. and Fu, M. C. Risk-sensitive reinforcement learning via policy gradient search. *Foundations and Trends® in Machine Learning*, 15(5):537–693, 2022.

Prashanth, L. A., Jagannathan, K., and Kolla, R. K. Concentration bounds for CVaR estimation: The cases of light-tailed and heavy-tailed distributions. In *International Conference on Machine Learning*, volume 119, pp. 5577–5586, 2020.

Rockafellar, R. T. and Uryasev, S. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

Sutton, R. S. and Barto, A. G. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 2nd ed. edition, 2018.

Thomas, P. and Learned-Miller, E. Concentration inequalities for conditional value at risk. In *International Conference on Machine Learning*, pp. 6225–6233, 2019.

Wang, Y. and Gao, F. Deviation inequalities for an estimator of the conditional value-at-risk. *Operations Research Letters*, 38(3):236–239, 2010.

A. Proof of Theorem 3.1

Proof. We first derive the two lower bounds for $\eta(M, f) = v_\alpha(M, f)$. We later show how a similar proof tactic works in the CVaR case as well. For notational convenience, let $\hat{\eta}(H_n, f) = \hat{\eta}_n(H_n, f)$.

Our proof of the VaR result involves three main steps: i.) constructing a suitable two-state MCP, ii.) deriving a lower bound on the minimax error in terms of the MCP parameters, and iii.) tightening the lower bound by optimizing over the various choices for these parameters. We now provide the details of these three steps.

1. **MCP construction**: Let A and B be two arbitrary but distinct elements of $\mathcal{U}$. The MCP we construct is

$$(M_*, f_*), \tag{46}$$

where $M_* \equiv (\mathcal{S}_*, P_*, \nu_*)$ is the two-state Markov chain with $\mathcal{S}_* = \{A, B\}$, the transition matrix

$$P_* = \begin{pmatrix} p & 1-p \\ p & 1-p \end{pmatrix}$$

for some $p \in [0, 1]$, and the initial state distribution $\nu_* \equiv (q, 1-q)$ for some $q \in [0, 1]$; this Markov chain is shown in Figure 1. We let the single-stage cost function f_* be either f_0 or f_1, where $f_0(A) = f_0(B) = f_1(B) = 0$, while

$$f_1(A) = 2\epsilon \exp\left(\frac{1}{\epsilon^2}\right). \tag{47}$$

Clearly, for $\alpha \in [0, 1]$, $v_\alpha(M_*, f_0) = 0$, while $v_\alpha(M_*, f_1) \geq 0$.

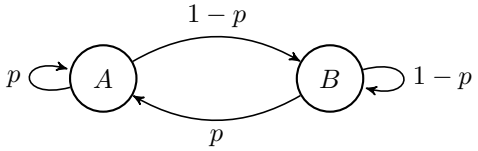

Figure 1. A two state Markov chain

2. **Lower bounding the minimax error**: Let $\mathscr{A} \equiv (\eta, \boldsymbol{r})$ be any VaR estimation algorithm, and $\mathbb{P}_*$ the probability distribution of H_n under $P^n_{M_*, \boldsymbol{r}}$. Then, for any p and q such that

$$v_\alpha(M_*, f_1) = v_\alpha(M_*, f_1) - v_\alpha(M_*, f_0) \geq 2\epsilon, \tag{48}$$

we have

$$\begin{aligned}
&\sup_{M, f} \mathop{\mathbb{P}}_{H_n \sim P^n_{M, \boldsymbol{r}}} \left(|\hat{\eta}(H_n, f) - v_\alpha(M, f)| \geq \epsilon\right) \\
&\qquad \geq \max_{f \in \{f_0, f_1\}} \mathbb{P}_* \left(|\hat{\eta}(H_n, f) - v_\alpha(M, f)| \geq \epsilon\right) && (49) \\
&\qquad \geq \frac{1}{2} \sum_{i=0}^{1} \mathbb{P}_* \left(|\hat{\eta}(H_n, f_i) - v_\alpha(M_*, f_i)| \geq \epsilon\right) && (50) \\
&\qquad \geq \frac{1}{2} \mathbb{P}_* \big(\{|\hat{\eta}(H_n, f_0) - v_\alpha(M_*, f_0)| \geq \epsilon\} \\
&\qquad\qquad \cup \{|\hat{\eta}(H_n, f_1) - v_\alpha(M_*, f_1)| \geq \epsilon\} \big) && (51) \\
&\qquad \geq \frac{1}{2} \mathbb{P}_* \big(\hat{\eta}(H_n, f_0) = \hat{\eta}(H_n, f_1) \big) && (52) \\
&\qquad \geq \frac{1}{2} \mathbb{P}_* \big(s_t = B \; \forall t \in \{0, 1, \ldots, n-1\} \big) && (53) \\
&\qquad \geq \frac{1}{2} \nu_*(B) \min\{\nu_*(B), P_*(B|B)\}^{n-1} && (54)
\end{aligned}$$

$$\geq \frac{1}{2}\min\{1-q, 1-p\}^n. \tag{55}$$

Above, (50) follows since $\max\{x, y\} \geq (x+y)/2$ for any $x, y \in \mathbb{R}$, while (51) is due to an union bound. The relation in (52) holds because

$$\{\hat{\eta}(H_n, f_0) = \hat{\eta}(H_n, f_1)\} \subseteq \{|\hat{\eta}(H_n, f_0) - v_\alpha(M_*, f_0)| \geq \epsilon\} \cup \{|\hat{\eta}(H_n, f_1) - v_\alpha(M_*, f_1)| \geq \epsilon\}$$

which itself is implied by (48). Further, (53) holds because the estimator $\hat{\eta}$ cannot differentiate between f_0 and f_1 when $s_t = B \; \forall t \in \{0, \ldots, n-1\}$, while (54) holds due to (5). Finally, (55) follows by using the fact that $\nu_*(B) = 1-q$ and $P_*(B|B) = 1-p$.

3. **Tightening the bound**: Since (55) is true for any p and q such that (48) holds, we can now optimize over these values. This leads to following optimization problem:

$$\begin{aligned} &\max_{p,q} \;\; \min\{1-q, 1-p\} \\ &\text{s.t. } \; 0 \leq p \leq 1, \;\; 0 \leq q \leq 1, v_\alpha(M_*, f_1) \geq 2\epsilon \end{aligned} \tag{56}$$

Claim: Let $i := \left\lfloor \frac{1}{\epsilon^2 \ln\left(\frac{1}{\gamma}\right)} \right\rfloor$. Then, $p = q = 1 - \alpha^{1/i}$ lies in the constraint set of (56).

We now verify this claim. Clearly, the inequality $v_\alpha(M_*, f_1) \geq 2\epsilon$ is implied by

$$\alpha \geq \mathcal{F}_1(2\epsilon), \tag{57}$$

where $\mathcal{F}_1$ is a shorthand for the CDF $\mathcal{F}(M_*, f_1)$. Therefore, to establish the claim, it suffices to show that (57) holds for the given choice of p and q.

We have

$$\begin{aligned} \mathcal{F}_1(2\epsilon) &= \lim_{t\to\infty} \mathbb{P}\left(\sum_{n=0}^{t} \gamma^n f_1(s_n) \leq 2\epsilon\right) \\ &\leq \sup_{t \geq i-1} \mathbb{P}\left(\sum_{n=0}^{t} \gamma^n f_1(s_n) \leq 2\epsilon\right) \\ &\leq \mathbb{P}\left(s_n = B \;\; \forall n \in \{0, 1, \ldots, i-1\}\right) \qquad (58) \\ &\leq \max\{1-q, 1-p\}^i \qquad (59) \end{aligned}$$

where (58) holds because $s_n = A$ for even one $n \in \{0, \ldots, i-1\}$ would imply that $\sum_{n=0}^{t} \gamma^n f(s_n) > 2\epsilon$ for any $t \geq i-1$, which itself is implied by the fact that, $\forall n \in \{0, \ldots, i-1\}$,

$$\gamma^n f_1(A) \geq \gamma^{i-1} f_1(A) > \gamma^i f_1(A) \geq 2\epsilon;$$

while (59) follows as in (55). Substituting the value of p and q from the claim in (59) shows that $\mathcal{F}_1(2\epsilon) \leq \alpha$, as desired.

Using the claim, it now follows that the optimal value of (56) has the lower bound $\alpha^{1/i}$. Combining this observation with (55) and substituting the value of i gives

$$\sup_{M,f} \; \mathbb{P}_{H_n \sim P^n_{M,\mathbf{r}}} \left(|\hat{\eta}(H_n, f) - v_\alpha(M, f)| \geq \epsilon\right) \geq \alpha^{n/i} \geq \exp\left[-n\epsilon^2 \ln\left(\frac{1}{\alpha}\right) \ln\left(\frac{1}{\gamma}\right)\right]. \tag{60}$$

Since algorithm $\mathscr{A}$ was arbitrary, (8) follows.

We now derive (9). Clearly, for any random variable X,

$$\mathbb{E}|X| \geq \epsilon \mathbb{P}\{|X| \geq \epsilon\}.$$

Hence, for any VaR estimation algorithm $\mathscr{A}$,

$$\begin{aligned}\sup_{M,f} \mathop{\mathbb{E}}_{H_n \sim P^n_{M,\boldsymbol{r}}} &|\hat{\eta}(H_n, f) - v_\alpha(M, f)| \\ &\geq \epsilon \exp\left(-n\epsilon^2 \ln\left(\frac{1}{\alpha}\right) \ln\left(\frac{1}{\gamma}\right)\right) \\ &\geq \frac{e^{-\ln\alpha \ln\gamma}}{\sqrt{n}}, \end{aligned} \tag{61}$$

where the last relation follows by picking $\epsilon = n^{-1/2}$. Again, since $\mathscr{A}$ was arbitrary, (9) follows.

It now remains to derive the lower bounds for the CVaR case. The above proof works in more or less the same way, except for some minor modifications. In the CVaR case, consider the optimization problem that is analogous to (56). Due to (2), any (p, q)-pair that is feasible for (56) is also feasible for this new optimization problem. Hence, the VaR lower bounds hold for CVaR estimation as well. □

Remark A.1 (***On the lower bound of (Metelli et al., 2023)'s work***). For the special case of mean estimation in a MCP, (Metelli et al., 2023) claim a lower bound of order $\Omega\left(1/\sqrt{n}\right)$. However, a closer inspection of their proof reveals that their lower bound is $\Omega\left(1/n\right)$. We justify this claim below. For the special case of η being the mean, by combining the relations in (26) and (34), it follows that

$$\inf_{\mathscr{A}} \sup_{M,f} \mathbb{P}\{|\hat{\eta}(H_n, f) - \eta(M, f)| \geq \epsilon\} \geq \begin{cases} \exp\left(-\frac{\epsilon^2 n(1-\beta\gamma}{\sigma_\gamma^2 f}\right), & \epsilon \in [0, \frac{1-\beta}{1-\beta\gamma}], \\ \exp\left(-\frac{n\epsilon^2}{\sigma_\gamma^2 f}\right), & \text{otherwise.} \end{cases} \tag{62}$$

Importantly, as shown below (34) there, we have $\sigma_\gamma^2 f = \epsilon(1-\epsilon)$. This implies $\epsilon^2/(\sigma_\gamma^2 f) = \epsilon/(1-\epsilon)$. For the more interesting low regime ϵ-case, we obtain

$$\frac{\epsilon}{1-\epsilon} \leq \frac{\epsilon(1-\beta\gamma)}{\beta(1-\gamma)}.$$

Thus,

$$\inf_{\mathscr{A}} \sup_{M,f} \mathbb{P}\{|\hat{\eta}(H_n, f) - \eta(M, f)| \geq \epsilon\} \geq \exp\left(-\frac{\epsilon n(1-\beta\gamma)^2}{\beta(1-\gamma)}\right).$$

By setting the RHS equal to δ, it follows that the estimation error is $\Omega(1/n)$ with probability at least δ. In contrast, in our lower bound in (8), the estimation error is $\Omega(1/\sqrt{n})$.

B. Proof of the Variance Lower Bound in Theorem 3.5

Proof. We now discuss the lower bound proof for the case $\eta(M, f) = \mathrm{V}(M, f)$. We again consider two variants (M_0, f^{-1}) and (M_0, f^{+1}) of the MCP (M_0, f), where $f^\nu(s)$, $\nu \in \{-1, +1\}$, is as in (4), but with $\mu = 0$ and σ^2 replaced by $\sigma_\nu^2 := 1 + (1+\nu)\epsilon$ for some $\epsilon > 0$. We then have $\mathrm{V}(M_0, f^{+1}) - \mathrm{V}(M_0, f^{-0}) = 2\epsilon$.

Arguing as in (31)–(37) we have

$$\inf_{\mathscr{A}} \sup_{M,f} \mathbb{P}\{|\hat{\eta}(H_n, f) - \eta(M, f)| \geq \epsilon\} \tag{63}$$

$$\geq \frac{1}{2}\left[1 - \sqrt{n D_{\mathrm{kl}}(P_{+1}, P_{-1})}\right] \tag{64}$$

$$= \frac{1}{2}\left[1 - \sqrt{n(2\epsilon - \ln(1+2\epsilon) + 2\epsilon^2)}\right] \tag{65}$$

$$\geq \frac{1}{2}\left[1 - \sqrt{6n\epsilon^2}\right], \tag{66}$$

where (65) follows from KL-divergence formula for Gaussian random variables, while (66) holds since $2\epsilon - \ln(1+2\epsilon) \leq 4\epsilon^2$ for any $\epsilon > 0$. The rest of the proof follows as in (38)–(41), mutatis mutandis. □

C. Proof of Theorem 4.10

Proof. The initial passage in the proof of Theorem 4.1 leading up to (43) holds for a Lipschitz risk measure $\eta(\cdot)$ satisfying properties (i) and (ii) from the theorem statement. Thus,

$$\eta(D) - \frac{\gamma^T K}{1-\gamma} \leq \eta(D_T) \leq \eta(D) + \frac{\gamma^T K}{1-\gamma}. \tag{67}$$

Using the bound above, we have

$$\begin{aligned} &\mathbb{E}\,|\hat{\eta}_N - \eta(D)| \\ &\leq \mathbb{E}\,|\hat{\eta}_N - \eta(D_T)| + \mathbb{E}\,|\eta(D_T) - \eta(D)| \\ &\leq \frac{32L(1-\gamma^T)^2K^2}{(1-\gamma)^2\sqrt{m}} + \frac{\gamma^T K}{1-\gamma}, \end{aligned} \tag{68}$$

where the final inequality uses (67) to bound the second term in (68), while the first term there is bounded using a special case of the result from (Prashanth & Bhat, 2022, Theorem 19), which is given below.

Lemma C.1. *Let $X^1, \ldots, X^m$ be drawn i.i.d. from the distribution of a random variable X, satisfying $|X| \leq B$ a.s. Suppose $\eta(\cdot)$ is a Lipschitz risk measure with constant L. Let $\hat{\eta}_m = \eta(F_m)$, where F_m is the EDF. Then,*

$$\mathbb{E}\,|\hat{\eta}_m - \eta(X)| \leq \frac{32LB^2}{\sqrt{m}}. \tag{69}$$

The application of the result above to bound the first term in (68) is valid for reasons listed at the end of the proof of Theorem 4.1. □

D. Proof of Theorem 4.11

Proof. The initial passage in the proof of Theorem 4.10 leading up to (67) holds here. Using the inequalities in (67), we derive the main concentration result as follows:

$$\begin{aligned} &\mathbb{P}\left[|\hat{\eta}_N - \eta(D)| > \epsilon\right] \\ &= \mathbb{P}\left[|\hat{\eta}_N - \eta(D_T) + \eta(D_T) - \eta(D)| > \epsilon\right] \\ &\leq \mathbb{P}\left[|\hat{\eta}_N - \eta(D_T)| > \epsilon - \frac{\gamma^T K}{1-\gamma}\right] \\ &\leq 2\exp\left(-\frac{m(1-\gamma)^2}{256(1-\gamma^T)^2K^2\mathsf{e}}\left[\frac{1}{L}\left[\epsilon - \frac{\gamma^T K}{1-\gamma}\right] - \frac{512K}{(1-\gamma)\sqrt{m}}\right]^2\right), \end{aligned}$$

where the penultimate inequality used (67), while the final inequality follows by applying Theorem 27 in (Prashanth & Bhat, 2022) after observing that $|D_T| \leq \frac{(1-\gamma^T)K}{1-\gamma}$. □

E. Proof of Theorem 4.12

Proof. Follows in a similar manner as the proof of Theorem 4.4 after observing that

$$\begin{aligned} |\mathrm{V}(D) - \mathrm{V}(D_T)| &\leq |\mathbb{E}(D^2 - D_T^2)| + |(\mathbb{E}D)^2 - (\mathbb{E}D_T)^2| \\ &\leq |\mathbb{E}\left[(D - D_T)(D + D_T)\right]| + |(\mathbb{E}D - \mathbb{E}D_T)(\mathbb{E}D + \mathbb{E}D_T)| \\ &\leq \frac{2\gamma^T K^2}{(1-\gamma)^2} + \frac{2\gamma^T K^2}{(1-\gamma)^2} = \frac{4\gamma^T K^2}{(1-\gamma)^2}, \end{aligned} \tag{70}$$

where we used the fact that $|D + D_T| \leq \frac{2K}{1-\gamma}$ and $|D - D_T| \leq \frac{\gamma^T K}{1-\gamma}$.

Notice that

$$\mathbb{E}|\hat{\mathrm{V}}_N - V(D)| \leq \mathbb{E}|\hat{\mathrm{V}}_N - \mathrm{V}(D_T)| + |\mathrm{V}(D_T) - \mathrm{V}(D)|. \tag{71}$$

We bound the first term on the RHS as follows: Letting $\Lambda_i = (D^i - \bar{\kappa}_m)^2$,

$$\hat{\mathrm{V}}_N - \mathrm{V}(D_T) = \frac{1}{m-1}\sum_{i=1}^{m}\Lambda_i - \mathrm{V}(D_T) \tag{72}$$

$$= \frac{1}{m-1}\sum_{i=1}^{m}\big[\Lambda_i - \frac{m-1}{m}\mathrm{V}(D_T)\big]. \tag{73}$$

Therefore, using the fact the summands $\{\Lambda_i\}$ are zero mean and i.i.d., we get

$$\mathbb{E}\big(\hat{\mathrm{V}}_N - \mathrm{V}(D_T)\big)^2 = \frac{1}{(m-1)^2}\sum_{i=1}^{m}\mathbb{E}\big[\Lambda_i - \frac{m-1}{m}\mathrm{V}(D_T)\big]^2.$$

Using $|\Lambda_i| \leq \frac{4(1-\gamma^T)^2K^2}{(1-\gamma)^2}$ and Jensen's inequality, we then obtain

$$\begin{aligned}\mathbb{E}|\hat{\mathrm{V}}_N - \mathrm{V}(D_T)| &\leq \sqrt{\mathbb{E}(\hat{\mathrm{V}}_N - \mathrm{V}(D_T))^2} \\ &\leq \frac{\sqrt{m}}{m-1}\frac{4(1-\gamma^T)^2K^2}{(1-\gamma)^2} \leq \frac{2}{\sqrt{m}}\frac{4(1-\gamma^T)^2K^2}{(1-\gamma)^2}, \text{ for } m \geq 2.\end{aligned}$$

Substituting the bound above and (70) into (71), we obtain

$$\mathbb{E}|\hat{\mathrm{V}}_N - \mathrm{V}(D)| \leq \frac{8(1-\gamma^T)^2K^2}{\sqrt{m}(1-\gamma)^2} + \frac{4\gamma^TK^2}{(1-\gamma)^2}. \tag{74}$$

We now turn to proving the tail bound in the theorem statement. Notice that

$$\begin{aligned}\mathbb{P}\left[|\hat{\mathrm{V}}_N - \mathrm{V}(D)| > \epsilon\right] &= \mathbb{P}\left[|\hat{\mathrm{V}}_N - \mathrm{V}(D_T) + \mathrm{V}(D_T) - \mathrm{V}(D)| > \epsilon\right] \\ &\leq \mathbb{P}\left[|\hat{\mathrm{V}}_N - \mathrm{V}(D_T)| > \epsilon - \frac{4\gamma^TK^2}{(1-\gamma)^2}\right] \\ &\leq \mathbb{P}\left[\left|\frac{1}{m-1}\sum_{i=1}^{m}\big[\Lambda_i - \frac{m-1}{m}\mathrm{V}(D_T)\big]\right| > \epsilon - \frac{4\gamma^TK^2}{(1-\gamma)^2}\right] \\ &= \mathbb{P}\left[\left|\frac{1}{m}\sum_{i=1}^{m}\big[\Lambda_i - \frac{m-1}{m}\mathrm{V}(D_T)\big]\right| > \frac{m-1}{m}\left(\epsilon - \frac{4\gamma^TK^2}{(1-\gamma)^2}\right)\right] \\ &= \mathbb{P}\left[\left|\frac{1}{m}\sum_{i=1}^{m}\big[\Lambda_i - \frac{m-1}{m}\mathrm{V}(D_T)\big]\right| > \frac{1}{2}\left(\epsilon - \frac{4\gamma^TK^2}{(1-\gamma)^2}\right)\right] \\ &\leq 2\exp\left(-\frac{m(1-\gamma)^4}{32(1-\gamma^T)^4K^4}\left(\epsilon - \frac{4\gamma^TK^2}{(1-\gamma)^2}\right)^2\right),\end{aligned}$$

where the final step follows by an application of Hoeffding's inequality. □