Accelerated Online Risk-Averse Policy Evaluation in POMDPs with Theoretical Guarantees and Novel CVaR Bounds

YAACOV PARIENTE*, Technion - Israel Institute of Technology, Israel
VADIM INDELMAN, Technion - Israel Institute of Technology, Israel

Background: Risk-averse decision-making under partial observability involves evaluating a risk measure, such as the Conditional Value-at-Risk (CVaR), of the cumulative return, a computationally demanding task in POMDPs.
Objectives: This work introduces a framework for accelerating CVaR value function evaluation by replacing the original POMDP dynamics with a computationally tractable surrogate model while providing formal performance guarantees.
Methods: As the mathematical foundation, we derive bounds on the CVaR of a random variable X using an auxiliary random variable Y under assumptions relating their cumulative distribution functions. We introduce a truncation decomposition that separates CVaR estimation from n samples into a data-driven term—bounded at rate $O(1/\sqrt{n})$ without support knowledge—and a tail correction where the support bound enters only through an $O(1/n)$ term, strictly improving existing concentration bounds. These bounds are further tightened via local CDF concentration parameters that exploit the quantile structure of CVaR. Building on this foundation, we establish upper and lower bounds on the CVaR value function computable from a simplified belief-MDP, accommodating general simplifications of the transition dynamics. We develop sample-based estimators for computing these bounds during online policy evaluation with probabilistic performance guarantees. These bounds enable computational acceleration via action elimination: actions whose bounds indicate suboptimality under the simplified model are safely discarded while ensuring consistency with the original POMDP.
Results: Empirical evaluation across multiple POMDP domains confirms that the bounds reliably distinguish safe from dangerous policies while achieving 9–21× computational speedups on the evaluated benchmarks.
Conclusions: The proposed framework provides a principled and practical approach for accelerating risk-averse policy evaluation in POMDPs, with theoretical guarantees and significant computational gains demonstrated across multiple domains.

JAIR Associate Editor: Nils Jansen

JAIR Reference Format:
Yaacov Pariente and Vadim Indelman. 2026. Accelerated Online Risk-Averse Policy Evaluation in POMDPs with Theoretical Guarantees and Novel CVaR Bounds. *Journal of Artificial Intelligence Research* 87, Article 24 (November 2026), 67 pages. DOI: 10.1613/jair.1.22332

1 Introduction

Autonomous agents have emerged as a cornerstone of modern intelligent systems, with applications spanning healthcare, disaster response, education, industrial automation, and decentralized systems. From autonomous aerial vehicles coordinating firefighting efforts to multi-agent systems executing complex financial transactions, these agents are designed to perceive, reason, and act without continuous human oversight. However, ensuring that such systems behave safely and reliably in uncertain and partially observable environments remains a central

*Corresponding Author.

Authors' Contact Information: Yaacov Pariente, ORCID: 0009-0000-5433-4071, yaacovp@campus.technion.ac.il, Faculty of Mathematics, Technion - Israel Institute of Technology, Haifa, Israel; Vadim Indelman, ORCID: 0000-0002-1863-3442, vadim.indelman@technion.ac.il, Stephen B. Klein Faculty of Aerospace Engineering; Faculty of Data and Decision Sciences, Technion - Israel Institute of Technology, Haifa, Israel.



This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2026 Copyright held by the owner/author(s).
DOI: 10.1613/jair.1.22332

challenge. In these settings, risk-aware decision-making frameworks, particularly those grounded in partially observable Markov decision processes (POMDPs), provide a principled foundation for modeling uncertainty and enforcing safety through explicit risk-sensitive optimization (Ahmadi, Rosolia, et al. 2023; Hou et al. 2016; Marecki and Varakantham 2010).

Exact solution methods for POMDPs become computationally intractable when the state, observation, and action spaces are large (Papadimitriou and Tsitsiklis 1987). To address this fundamental limitation, approximate planning algorithms utilize forward search trees constructed via simulation of future agent trajectories (Silver and Veness 2010; Sunberg and Kochenderfer 2018). The sampling procedure is guided by the underlying reward function, which provides evaluative feedback for action selection. This approach enables practical online planning in partially observable domains by trading off solution optimality for computational tractability.

Risk can be incorporated into planning in several ways, two of which are relevant to this work. One keeps the expected return as the objective and adds a constraint. Chance-constrained planners bound the probability that a policy violates a safety requirement (Moss, Jamgochian, et al. 2024; Ono et al. 2015; Rodrigues Quemel e Assis Santana et al. 2016), and risk-constrained formulations bound the CVaR of the accumulated cost (Chow et al. 2018). Online planners have adopted the latter criterion, with Monte Carlo tree search maximizing the expected return while holding the CVaR of the accumulated cost below a prescribed threshold (Zhang et al. 2025). The other way replaces the expectation itself, evaluating the risk measure on the return, either recursively at every stage (Ahmadi, Ono, et al. 2020; Ruszczyński 2010) or once over the cumulative return. Among risk-averse formulations, this work adopts the approach wherein a risk measure—Conditional Value-at-Risk (CVaR)—is incorporated directly into the value function and computed over the cumulative return (Bäuerle and Ott 2011). CVaR is a coherent risk measure that is widely used (Majumdar and Pavone 2020).

Current state-of-the-art POMDP planning algorithms remain computationally demanding, as they rely on extensive sampling from the original POMDP dynamics. To mitigate this issue, the simplification approach replaces the true POMDP dynamics with a computationally tractable surrogate model, which is then employed during planning while still admitting performance guarantees relative to the original dynamics. Simplification techniques for expectation-based value functions are already available (Barenboim and Indelman 2023; Lev-Yehudi et al. 2024; Zhitnikov, Sztyglic, et al. 2025) and have been shown to accelerate POMDP planning. However, risk-averse simplification has remained largely unexamined; the only existing contribution in this direction employs a non-coherent risk measure (Zhitnikov and Indelman 2022), which offers weaker justification for risk-averse decision making (Majumdar and Pavone 2020).

In several settings, the original observation model is too computationally expensive to plan with, and a simplified one is used in its place. In robotics, the observation model describes the robot's sensor observations and may involve, for example, a large neural network for image processing. A simplified observation model in such a setting could use a less computationally expensive neural network. Another example is replacing a multimodal observation model, such as a Gaussian mixture, with a single Gaussian. In these cases, the surrogate is used only for planning, and the planner uses the simplified observation model to compute bounds on action values under the original model. This setting was studied by Lev-Yehudi et al. (2024) for the expected-return objective and is studied in this work for the CVaR objective.

We begin by establishing the mathematical foundations of this work, deriving several CVaR bounds for a random variable X using an auxiliary random variable Y, under assumptions relating their respective CDFs and PDFs. Although not all of these theoretical results are employed in our practical applications, they form a foundational framework that we expect will support future research on simplification. Building on these foundations, we analyze the relationship between the original value function—defined as the CVaR of the return—and its simplified belief transition-model counterpart. In particular, we establish lower and upper bounds on the original value function in terms of the simplified value function and provide corresponding performance guarantees. We then

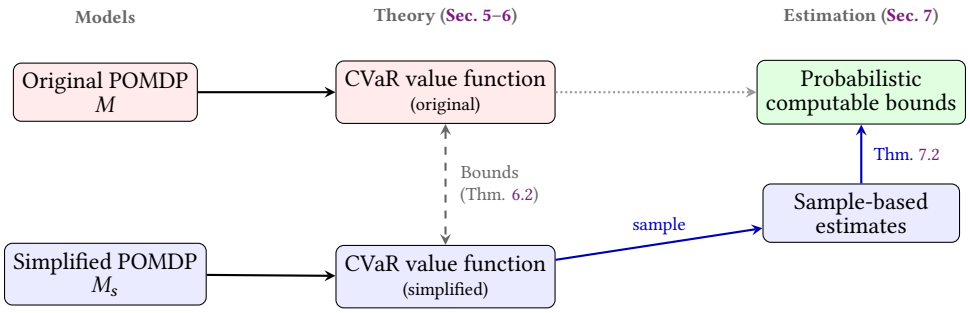


Fig. 1. Overview of the proposed framework. The original POMDP yields the original CVaR value function. The simplified POMDP M_s admits a simplified value function, related to the original via Theorem 6.2. In practice, the simplified value function and distributional discrepancy are estimated from sample trajectories, yielding probabilistic computable bounds on the original value function (Theorem 7.2).

develop estimators for computing these bounds during online planning and derive performance guarantees for these estimators as well. The primary contributions of this work are as follows:

- Derivation of support-relaxed bounds on the CVaR of a random variable X given another random variable Y, where a truncation decomposition confines unknown support bounds to $O(1/n)$ tail corrections. These bounds yield interpretable concentration inequalities that are 56–88% tighter on our benchmark distributions than the state-of-the-art bounds of Thomas and Learned-Miller (2019), establishing the tightest known finite-sample CVaR bounds under bounded support, and converge as the distributional discrepancy vanishes.
- Derivation of lower and upper bounds on the theoretical value function in risk-averse POMDPs expressed through the action-value function from a simplified belief-MDP transition model, with corresponding bounds for simplified observation models.
- Estimators for evaluating the simplified bounds during online planning, with probabilistic performance guarantees on the deviation between the theoretical action-value function and its estimated bounds computed using simplified belief-MDP and observation models within a particle-belief MDP framework.
- Application of the derived bounds to computational acceleration via action elimination, with empirical demonstration of 9–21× speedups and bounds that reliably distinguish safe from dangerous policies across multiple POMDP domains.

An overview of the proposed framework is depicted in Figure 1. The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 introduces the necessary preliminaries on POMDPs, particle belief MDPs, and CVaR estimation. Section 4 formulates the problem of bounding the risk-averse value function under a simplified belief-transition model. Section 5 derives the foundational CVaR bounds using auxiliary random variables, including the truncation decomposition and local concentration refinements. Section 6 establishes value function bounds for the static CVaR value function under general and observation-model simplifications. Section 7 develops online estimators for computing these bounds during planning and provides corresponding performance guarantees. Section 8 discusses the limitations of the framework. Section 9 presents the experimental evaluation across multiple POMDP domains. Section 10 concludes the paper. Proofs and additional experimental details are provided in the appendix.

2 Related Work

Conditional Value at Risk (CVaR) (Rockafellar, Uryasev, et al. 2000) is a principled and extensively studied risk measure with broad applicability across domains involving uncertainty. A central property of CVaR is its dual representation (Artzner et al. 1999), which allows it to be interpreted as the expected loss in the worst-case tail of a cost distribution (Acerbi and Tasche 2002). Unlike Value at Risk (VaR), which identifies only the quantile threshold of extreme losses at a chosen confidence level, CVaR quantifies both the probability and the magnitude of such losses, yielding a more comprehensive assessment of risk. This property makes CVaR especially relevant in safety-critical decision-making contexts, where both the occurrence and severity of adverse outcomes are significant. Moreover, CVaR satisfies the axioms of coherent risk measures—such as subadditivity and positive homogeneity—ensuring consistent and rational aggregation of risk across time and decision stages (Artzner et al. 1999). Practical estimators of CVaR are supported by concentration inequalities and deviation guarantees from the true risk value (Brown 2007; Thomas and Learned-Miller 2019), further reinforcing its reliability in applied decision-making problems.

Risk can be incorporated into planning under uncertainty through chance constraints (Ono et al. 2015), exponential utility (Koenig and Simmons 1994), equivalently the entropic risk measure (Föllmer and Schied 2002), distributionally robust optimization (Cubuktepe et al. 2021; Osogami 2015; Xu and Mannor 2010), and quantile regression (Dabney et al. 2018). Variance-based and prospect-theoretic criteria are also used: Mannor and Tsitsiklis (2011) show that mean–variance optimization of the finite-horizon cumulative reward is NP-hard, Brázdil et al. (2017) synthesize strategies that keep the expected mean payoff above and its variance below given thresholds, which requires both memory and randomization, and Brihaye et al. (2025) give complexity results for cumulative prospect theory, a non-coherent criterion. Among these, distributionally robust formulations align closely with CVaR, as both emphasize resilience to rare but high-impact events (Bäuerle and Glauner 2020). A related line of work optimizes a percentile of the return under model uncertainty in robust MDPs (Behzadian et al. 2021; Delage and Mannor 2010; Lobo et al. 2023). Risk measures must satisfy fundamental axioms of interpretability and consistency (Majumdar and Pavone 2020); coherent risk measures satisfy these axioms and have been employed as objectives in POMDPs, constrained MDPs, and shortest-path problems (Ahmadi, Dixit, et al. 2021; Ahmadi, Ono, et al. 2020; Ahmadi, Rosolia, et al. 2021; Dixit et al. 2023), where CVaR arises as a key special case. These criteria differ in the risk they address. Robust and distributionally robust formulations address errors in the estimated transition model, and chance constraints bound the probability of failure without measuring its severity. The remaining criteria are risk measures on the cost distribution itself, none of them coherent: VaR marks where the worst outcomes begin but ignores their magnitude, variance penalizes deviations in both directions, exponential utility reweights all outcomes through a sensitivity parameter, and prospect theory distorts probabilities to model human perception. CVaR is coherent and directly interpretable: given a risk budget α, it is the expected cost of the worst α-fraction of outcomes. This work adopts its static form on the cumulative cost, applied once rather than stage-wise or as a constraint.

A separate line of work computes such tail criteria exactly, assuming a fully known model. The closest *formulation* to ours is that of Bäuerle and Ott (2011): MDPs whose objective is the CVaR of the cumulative cost, solved by dynamic programming on a state space augmented with the accumulated cost. In probabilistic verification, Křetínský and Meggendorfer (2018) use linear programming to determine whether a policy of a finite MDP can satisfy given CVaR, VaR, and expectation constraints on reachability and mean-payoff objectives, and Meggendorfer (2022) computes the exact static CVaR of the accumulated cost of a stochastic shortest path by linear programming or Pareto-set value iteration. Ummels and Baier (2013) and Randour et al. (2017) compute quantiles of the accumulated cost of finite MDPs and synthesize strategies for percentile constraints on it, by exact unfolding. Jiang and Powell (2018) instead optimize a nested, time-consistent composition of per-stage quantile-based measures by approximate dynamic programming with almost-sure convergence guarantees. All

of these assume a known, typically finite, model and compute exact values or asymptotic approximations; none provides finite-sample bounds on the CVaR value function when planning uses an approximate model—the setting of this work.

Simplification techniques for POMDPs aim to reduce computational complexity while retaining performance guarantees, thereby facilitating real-time deployment of POMDP-based policies in practical systems (Papadimitriou and Tsitsiklis 1987). Simplification refers to replacing one or more POMDP components with computationally tractable approximations while providing formal performance guarantees that ensure decision quality with respect to the original model. For instance, Lev-Yehudi et al. (2024) examined simplification of the observation model by introducing a less expensive surrogate while establishing finite-sample convergence bounds. Similarly, Barenboim and Indelman (2026) studied simplification of the state and observation spaces, providing deterministic performance guarantees and integrating them into state-of-the-art solvers. Furthermore, Zhitnikov and Indelman (2022) examined the simplification of belief-dependent rewards and provided deterministic guarantees under the Value at Risk (VaR) criterion. However, VaR is not a coherent risk measure and therefore lacks several essential properties for principled risk-averse decision making. More broadly, the simplification of risk-averse planning remains largely unexplored, with Zhitnikov and Indelman (2022) being the only prior work in this direction. To the best of our knowledge, no prior work has addressed the problem of bounding the CVaR of a random variable using an auxiliary random variable with a known distributional discrepancy, which constitutes a key contribution of this work.

CVaR concentration bounds provide finite-sample guarantees on CVaR from observed data. Brown (2007) derived the first such bounds using McDiarmid's inequality, requiring two-sided support. Only their bound on underestimation of the CVaR requires a continuous CDF. Thomas and Learned-Miller (2019) improved these, requiring only one-sided support and no continuity assumption. Budde et al. (2025) apply the same DKW machinery to statistical model checking of quantiles and CVaR. Ghosh et al. (2024) developed bounds for optimized certainty equivalents via a variational representation, requiring sub-Gaussianity, with deviations of order $O(1/\sqrt{n})$ in the sample size n. The DKW-based bounds share a structural limitation in how the assumed support bound enters. In Brown (2007) the width of the support multiplies the statistical error term of the bound. In Thomas and Learned-Miller (2019) the support bound is appended to the sorted sample as its largest value, and as we show in Appendix A, it carries a weight of the same order as that error term. In both cases the weight of the support bound decays only at $O(1/\sqrt{n})$, so overestimating the support inflates the bound by the overestimate times this slowly decaying weight, which dominates the estimation error whenever the support is estimated conservatively. Our support-relaxed bounds instead confine the support bound to a tail correction that decays at $O(1/n)$ (Section 5.3).

3 Preliminaries

3.1 Partially Observable Markov Decision Process

A finite-horizon Partially Observable Markov Decision Process (POMDP) is formally defined as the tuple $(\mathcal{X}, A, Z, \mathcal{T}, O, c, b_0, \gamma)$, where $\mathcal{X}$, A, and Z denote the state, action, and observation spaces, respectively. The transition model $\mathcal{T}(x_{t+1} \mid x_t, a_t) \triangleq P(x_{t+1} \mid x_t, a_t)$ specifies the conditional probability of transitioning from state x_t to x_{t+1} given action a_t, while the observation model $O(z_t \mid x_t) \triangleq P(z_t \mid x_t)$ defines the likelihood of observing z_t when the system is in state x_t. The belief space B is the set of all probability distributions over $\mathcal{X}$, and the stage-wise cost function is given by $c : B \times A \to \mathbb{R}$.

At each time step t, the agent maintains a belief $b_t \in B$, representing the posterior distribution over the latent state conditioned on the history of observations and actions. The history is denoted by $H_t \triangleq \{z_{1:t}, a_{0:t-1}, b_0\}$, and the belief update is defined as $b_t(x_t) \triangleq P(x_t \mid H_t)$ for each $x_t \in \mathcal{X}$. A policy $\pi_t : B \to A$ is a measurable mapping that prescribes the action $a_t = \pi_t(b_t)$ based on the current belief state. The immediate expected cost under belief

b_t and action a_t is

$$c(b_t, a_t) \triangleq \mathbb{E}_{x\sim b_t}[c_x(x, a_t)], \tag{1}$$

where $c_x(x, a_t)$ denotes the state-dependent cost, bounded as $R_{\min} \le c_x(x, a_t) \le R_{\max}$.

The cumulative cost, or *return*, over a finite horizon $T \in \mathbb{N}$ is given by

$$R_{t:T} \triangleq \sum_{\tau=t}^{T} \gamma^{\tau-t} c(b_\tau, a_\tau), \tag{2}$$

which serves as the performance criterion starting at time t. The value function associated with policy π and initial belief b_k is written with the annotated operator $\mathbb{E}^{\pi}_{b_k}$, which carries the initial belief and the policy that generate the trajectory distribution instead of conditioning on them:

$$V^{\pi}(b_k) \triangleq \mathbb{E}^{\pi}_{b_k}[R_{k:T}] = \sum_{t=k}^{T} \gamma^{t-k} \mathbb{E}^{\pi}_{b_k}[c(b_t, a_t)], \tag{3}$$

and the corresponding action-value function is

$$Q^{\pi}(b_k, a_k) \triangleq \mathbb{E}^{\pi}_{b_k, a_k}[c(b_k, a_k) + V^{\pi}(b_{k+1})]\,. \tag{4}$$

A POMDP can equivalently be formulated as a fully observable *belief-MDP* (BMDP), in which the belief b_t serves as the state variable and is a sufficient statistic for the interaction history. In this formulation, the dynamics are governed by a belief transition kernel induced by the latent state-transition and observation models. Formally, the belief-MDP associated with a POMDP M is defined as

$$M_B \triangleq (\mathcal{B}, A, \tau, \rho, \gamma), \tag{5}$$

where $\mathcal{B} \triangleq \Delta(\mathcal{X})$ denotes the space of probability distributions over the latent state space $\mathcal{X}$, A is the action space, τ denotes the belief-transition kernel, ρ is the expected cost, and γ is the discount factor. Given the state-transition and observation models, the belief-transition model of the belief-MDP is given by

$$P(b_{t+1}|b_t, a_t) \triangleq \int_{x_t\in\mathcal{X}} \int_{x_{t+1}\in\mathcal{X}} \int_{z_{t+1}\in Z} P(x_{t+1}|x_t, a_t) P(z_{t+1}|x_{t+1}) b_t(x_t) dx_t dx_{t+1} dz_{t+1}. \tag{6}$$

This BMDP formulation is exact but generally intractable, since the belief space $\mathcal{B}$ is infinite-dimensional, motivating approximate representations such as particle belief MDPs.

3.2 Particle Belief MDP

In order to estimate the theoretical action-value function, we consider a particle belief MDP (PB-MDP) setting. A PB-MDP is the belief-MDP in which each belief is replaced by a weighted set of N_p state particles and the belief transition is realized by sampling: particles are propagated through the state-transition model, an observation is sampled, and the weights are updated by the observation likelihood. It is needed because the exact belief-MDP above cannot be sampled or stored, whereas the particle belief can be, and it is the object that the estimators of Section 7 operate on. Formally, denote by $M_P \triangleq (\Sigma, A, \tau, \rho, \gamma)$ the PB-MDP that is defined with respect to the POMDP M (Lim et al. 2023) and $N_p \in \mathbb{N}$, where

- $\Sigma \triangleq \{\bar{b} : \bar{b} = \{(x_i, w_i)\}_{i=1}^{N_p}, x_i \in \mathcal{X}, \forall i, w_i \ge 0, \exists i \text{ such that } w_i > 0\}$ is the state space over the particle beliefs.
- A is the action space as defined in the POMDP M.
- $\tau(\bar{b}_{t+1}|\bar{b}_t, a)$ is the belief transition probability, for $a \in A, \bar{b} \in \Sigma$.
- $\rho(\bar{b}, a) \triangleq \frac{\sum_{j=1}^{N_p} w_j c(x_j, a)}{\sum_{j=1}^{N_p} w_j}$ is the belief-dependent cost.
- γ as defined in the POMDP M.

This belief estimator maintains a finite set of state particles that serve as an approximation to the theoretical belief, which, in principle, may involve an infinite number of states.

3.3 Conditional Value-at-Risk

Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$, where Ω is the sample space, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and $P : \mathcal{F} \to [0, 1]$ is a probability measure. Throughout, we write $\mathbb{P}[\,\cdot\,]$ for the probability of an event under P, so that the probability operator matches the bracket notation of the expectation $\mathbb{E}[\,\cdot\,]$. The symbols P and P_s are reserved for the original and simplified belief-transition kernels introduced in the problem formulation, and the trajectory measures those kernels induce carry the annotated operators defined there. Assume further that $\mathbb{E}[|X|] < \infty$. We denote the cumulative distribution function (CDF) of the random variable X by $F_X(x) \triangleq \mathbb{P}[X \le x]$. Define its generalized inverse by $F_X^{-1}(u) \triangleq \inf\{x \in \mathbb{R} : F_X(x) \ge u\}$. The essential infimum and essential supremum of X are

$$\operatorname{ess\,inf} X \triangleq \sup\{c \in \mathbb{R} : \mathbb{P}[X \ge c] = 1\}, \qquad \operatorname{ess\,sup} X \triangleq \inf\{c \in \mathbb{R} : \mathbb{P}[X \le c] = 1\}. \tag{7}$$

These are the tightest almost-sure bounds on X, and any looser bound is equivalent to them: for $a, b \in \mathbb{R}$, $\mathbb{P}[X \ge a] = 1$ holds if and only if $a \le \operatorname{ess\,inf} X$, and $\mathbb{P}[X \le b] = 1$ holds if and only if $b \ge \operatorname{ess\,sup} X$. We use the latter form to state almost-sure support hypotheses, which keeps such a hypothesis readable while still admitting any conservative bound. The value at risk at confidence level $\alpha \in (0, 1]$ is defined using this lower-quantile convention:

$$\mathrm{VaR}_\alpha(X) \triangleq F_X^{-1}(1 - \alpha). \tag{8}$$

For simplicity we denote $q_\alpha^X \triangleq \mathrm{VaR}_\alpha(X)$. The conditional value at risk (CVaR) at confidence level α is defined as (Rockafellar, Uryasev, et al. 2000)

$$\mathrm{CVaR}_\alpha(X) \triangleq \inf_{w \in \mathbb{R}} \{w + \frac{1}{\alpha}\mathbb{E}[(X - w)^+] | w \in \mathbb{R}\}, \tag{9}$$

where $(x)^+ \triangleq \max(x, 0)$. Using the generalized inverse defined above, every integrable X satisfies (Pflug 2000)

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{\alpha}\int_{1-\alpha}^{1} F_X^{-1}(v)dv. \tag{10}$$

If F_X is continuous, then $\mathrm{CVaR}_\alpha(X) = \mathbb{E}[X | X > \mathrm{VaR}_\alpha(X)]$. Let $X_i \overset{\text{iid}}{\sim} F$ for $i \in \{1, \ldots, n\}$. Denote by

$$\hat{C}_\alpha(X) \triangleq \hat{C}_\alpha(\{X_i\}_{i=1}^n) \triangleq \inf_{x \in \mathbb{R}} \left\{x + \frac{1}{n\alpha}\sum_{i=1}^{n}(X_i - x)^+\right\} \tag{11}$$

the estimate of $\mathrm{CVaR}_\alpha(X)$ (Brown 2007). Theorem 3.1, that bounds the deviation of the estimated CVaR and the true CVaR, was proved in Brown (2007); we restate it because it is the baseline our bounds are compared against in Section 9.

Theorem 3.1. *If $a \le \operatorname{ess\,inf} X$ and $b \ge \operatorname{ess\,sup} X$, and X has a continuous distribution function, then for any $\delta \in (0, 1]$,*

$$\mathbb{P}\left[\mathrm{CVaR}_\alpha(X) - \hat{C}_\alpha(X) > (b - a)\sqrt{\frac{5\ln(3/\delta)}{\alpha n}}\right] \le \delta, \tag{12}$$

$$\mathbb{P}\left[\mathrm{CVaR}_\alpha(X) - \hat{C}_\alpha(X) < -\frac{(b - a)}{\alpha}\sqrt{\frac{\ln(1/\delta)}{2n}}\right] \le \delta. \tag{13}$$

The estimator in (11) can be expressed as

$$\hat{C}_\alpha(X) = X^{(n)} - \frac{1}{\alpha}\sum_{i=1}^{n-1}(X^{(i)} - X^{(i-1)})\Big(\frac{i}{n} - (1-\alpha)\Big)^+, \tag{14}$$

where $X^{(i)}$ is the ith order statistic of $X_1, \dots, X_n$ in ascending order (Thomas and Learned-Miller 2019). We extend this indexing to the endpoints by writing $X^{(0)} \triangleq a$ and $X^{(n+1)} \triangleq b$ whenever finite bounds $a \le \operatorname{ess\,inf} X_1$ and $b \ge \operatorname{ess\,sup} X_1$ are available, so that Theorems 3.2 and 3.3 below use the same order-statistic notation as (14). The results presented in Theorems 3.2 and 3.3, following the work of Thomas and Learned-Miller (2019), yield tighter bounds on the CVaR compared to those established by (Brown 2007).

Theorem 3.2. *If $X_1, \dots, X_n$ are independent and identically distributed random variables and $b \ge \operatorname{ess\,sup} X_1$ for some finite b, then for any $\delta \in (0, 0.5]$,*

$$\mathbb{P}\left[\mathrm{CVaR}_\alpha(X_1) \le X^{(n+1)} - \frac{1}{\alpha}\sum_{i=1}^{n}(X^{(i+1)} - X^{(i)})\left(\frac{i}{n} - \sqrt{\frac{\ln(1/\delta)}{2n}} - (1-\alpha)\right)^+\right] \ge 1-\delta, \tag{15}$$

where $X^{(1)}, \dots, X^{(n)}$ are the order statistics (i.e., $X_1, \dots, X_n$ sorted in ascending order), $X^{(n+1)} \triangleq b$, and $x^+ \triangleq \max\{0, x\}$ for all $x \in \mathbb{R}$.

Theorem 3.3. *If $X_1, \dots, X_n$ are independent and identically distributed random variables and $a \le \operatorname{ess\,inf} X_1$ for some finite a, then for any $\delta \in (0, 0.5]$,*

$$\mathbb{P}\left[\mathrm{CVaR}_\alpha(X_1) \ge X^{(n)} - \frac{1}{\alpha}\sum_{i=0}^{n-1}(X^{(i+1)} - X^{(i)})\Big(\min\Big\{1, \frac{i}{n} + \sqrt{\frac{\ln(1/\delta)}{2n}}\Big\} - (1-\alpha)\Big)^+\right] \ge 1-\delta, \tag{16}$$

where $X^{(1)}, \dots, X^{(n)}$ are the order statistics (i.e., $X_1, \dots, X_n$ sorted in ascending order), $X^{(0)} \triangleq a$, and $x^+ \triangleq \max\{0, x\}$ for all $x \in \mathbb{R}$.

Notably, Theorems 3.2 and 3.3 require only one-sided boundedness of the support and do not assume continuity of the distribution function, in contrast to Theorem 3.1, which requires both two-sided boundedness and a continuous distribution. The restriction to $\delta \in (0, 0.5]$ is mild in practice, as it corresponds to confidence levels of at least 50%.

4 Problem Formulation

In this work, we adopt a risk-averse formulation in which, instead of optimizing the expected return as in (3) and (4), the value function is defined as the *static* CVaR of the return—that is, the CVaR applied to the cumulative return over the full planning horizon, as opposed to the nested (dynamic) CVaR that applies risk measures recursively at each time step. Most work on CVaR in MDPs, including the foundational formulation of Bäuerle and Ott (2011), uses the static version. The nested version originates in the time-consistent dynamic risk measures of Ruszczyński (2010). Let $\alpha \in (0, 1]$, and denote the original POMDP by M. The value and action-value functions at time k are defined by

$$V_M^\pi(b_k, \alpha) \triangleq \mathrm{CVaR}^\pi_{\alpha, b_k}[\sum_{t=k}^{T} c(b_t, \pi(b_t))], \tag{17}$$

$$Q_M^\pi(b_k, a_k, \alpha) \triangleq \mathrm{CVaR}^\pi_{\alpha, b_k, a_k}[c(b_k, a_k) + \sum_{t=k+1}^{T} c(b_t, \pi(b_t))]. \tag{18}$$

Here $\mathrm{CVaR}^\pi_{\alpha, b_k}$ denotes the CVaR at level α of the return under the trajectory measure $\mathbb{P}^\pi_{b_k}$ induced by the original model M, the policy π, and the initial belief b_k, defined below; the corresponding operator under the simplified

model is written $\text{CVaR}^{\pi,s}_{\alpha,b_k}$ and is taken under $\mathbb{P}^{\pi,s}_{b_k}$. Figure 3a illustrates the difference between the standard expected value function and the CVaR value function.

We consider cases in which the belief-MDP transition model is simplified for computational tractability. Let Ω_B be the sample space of the belief random variable and let $\Omega \triangleq \Omega_B^{T-k+1}$ denote the corresponding product space, endowed with a σ-algebra $\mathcal{F}$. Let P and P_s be the original and simplified belief-transition kernels. An initial belief b_k and a policy π turn each kernel into a probability measure on $(\Omega, \mathcal{F})$, written $\mathbb{P}^{\pi}_{b_k}[\,\cdot\,]$ and $\mathbb{P}^{\pi,s}_{b_k}[\,\cdot\,]$ with expectations $\mathbb{E}^{\pi}_{b_k}$ and $\mathbb{E}^{\pi,s}_{b_k}$, so that b_k and π annotate the operator instead of being conditioned on, which would be undefined whenever the conditioning event has probability zero. Inside the integrals below, $P(\cdot \mid \cdot)$ and $P_s(\cdot \mid \cdot)$ are the kernel densities from which these measures are built, not conditional probabilities of them. Both measures are defined on the same space $(\Omega, \mathcal{F})$ of belief trajectories, because the simplification changes only the kernel that generates a trajectory and not the set of trajectories or the events formed from them; only the probability assigned to each event differs. The CDF of $R_{k:T}$ under the original model is

$$\mathbb{P}^{\pi}_{b_k}[R_{k:T} \leq l] \triangleq \int_{b_{k+1:T} \in B^{T-k}} P(R_{k:T} \leq l | b_{k:T}, \pi) \prod_{i=k+1}^{T} P(b_i | b_{i-1}, \pi) db_{k+1:T}, \tag{19}$$

and the simplified CDF of $R_{k:T}$ is defined using only the simplified belief transition model:

$$\mathbb{P}^{\pi,s}_{b_k}[R_{k:T} \leq l] \triangleq \int_{b_{k+1:T} \in B^{T-k}} P(R_{k:T} \leq l | b_{k:T}, \pi) \prod_{i=k+1}^{T} P_s(b_i | b_{i-1}, \pi) db_{k+1:T}. \tag{20}$$

The belief-MDP transition model is defined using the state transition and observation models. Hence, simplification of the belief-MDP transition model is general enough to include simplifications of state and observation models simultaneously.

Denote the simplified POMDP by M_s. For example, the observation model of M may be an expensive computer-vision neural network, and M_s replaces it with a smaller network that is much cheaper to evaluate at the price of reduced accuracy, so that simulating an observation under M_s costs a fraction of simulating one under M. The value and action-value functions corresponding to M_s are denoted by $V^{\pi}_{M_s}(b_k, \alpha)$ and $Q^{\pi}_{M_s}(b_k, a_k, \alpha)$, respectively. These definitions are identical to those in (17) and (18), except that the return distribution is defined with respect to $\mathbb{P}^{\pi,s}_{b_k}$:

$$V^{\pi}_{M_s}(b_k, \alpha) \triangleq \text{CVaR}^{\pi,s}_{\alpha,b_k}[\sum_{t=k}^{T} c(b_t, \pi(b_t))], \tag{21}$$

$$Q^{\pi}_{M_s}(b_k, a_k, \alpha) \triangleq \text{CVaR}^{\pi,s}_{\alpha,b_k,a_k}[c(b_k, a_k) + \sum_{t=k+1}^{T} c(b_t, \pi(b_t))]. \tag{22}$$

While the simplified model P_s is cheaper to simulate, directly using $Q^{\pi}_{M_s}(b_k, a_k, \alpha)$ as a surrogate for $Q^{\pi}_{M}(b_k, a_k, \alpha)$ may lead to incorrect risk assessments, since the return distributions under P and P_s differ. Our objective is to establish lower and upper bounds, denoted by L_s and U_s, such that $L_s \leq Q^{\pi}_{M}(b_k, a_k, \alpha) \leq U_s$, depending solely on quantities computable from the simplified model (20); this conceptual approach is summarized in Figure 2. These bounds provide formal guarantees on the true value function, enabling reliable action selection during online planning. As illustrated in Figure 3, the key quantity governing these bounds is ϵ, an upper bound on the discrepancy between the CDFs of the return under P and P_s. Bounding this CDF discrepancy allows us to bound the difference in CVaR, as established in subsequent sections.

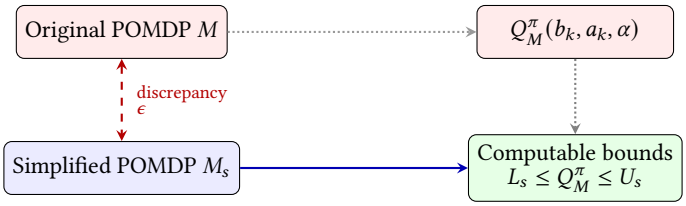
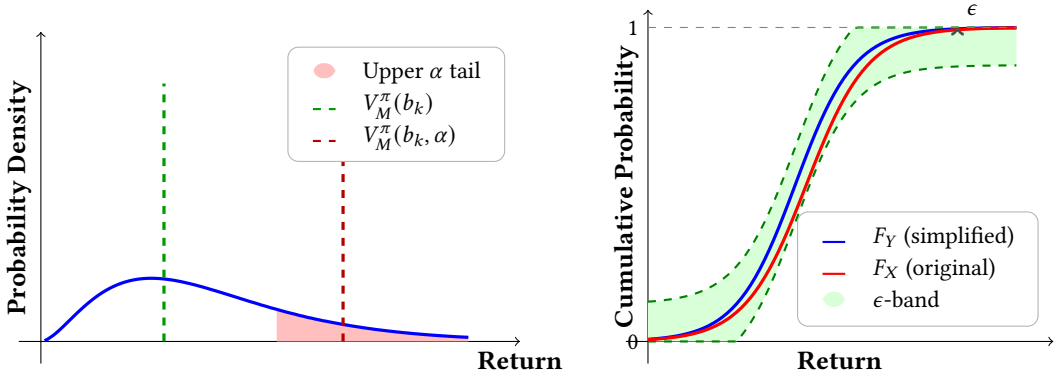


Fig. 2. Conceptual overview. Direct evaluation of Q_M^π under the original model M is computationally expensive. Instead, the simplified model M_s is used to compute bounds on Q_M^π, where the bound quality is governed by the distributional discrepancy ϵ between their belief transition models.



(a) Comparison of expected value $V_M^\pi(b_k)$ and CVaR value $V_M^\pi(b_k, \alpha)$. The shaded region represents the worst α fraction of outcomes.

(b) CDF discrepancy between original (F_X) and simplified (F_Y) models. The ϵ-band bounds the discrepancy, enabling CVaR bounds.

Fig. 3. Problem formulation illustrations. (a) The CVaR value function focuses on the tail of the return distribution, unlike the expected value. (b) The simplified model P_s yields a computable CDF F_Y, while the original CDF F_X is intractable. When $|F_X(r) - F_Y(r)| \le \epsilon$, bounds on $\text{CVaR}_\alpha(X)$ can be derived from Y.

5 CVaR Bounds

In the previous section, we introduced a simplified belief-transition model with distribution P_s as a computationally tractable approximation of the original belief-transition model P. This approximation induces two corresponding random variables: X, representing the return under the original model P, and Y, representing the return under the simplified model P_s. While direct evaluation of risk-sensitive criteria for X is often computationally infeasible, sampling and analysis of Y can be carried out efficiently.

The goal of this section is to characterize how $\text{CVaR}_\alpha(X)$ can be bounded and estimated using information derived from Y. The section is organized as follows. Section 5.1 establishes deterministic CVaR bounds relating X and Y through their CDF discrepancy. Section 5.2 introduces a truncation decomposition that confines the support dependence to an $O(1/n)$ tail correction. Section 5.3 assembles these components with the DKW inequality to present the main result of this section—Theorem 5.6, the support-relaxed CVaR bounds—where the unknown support enters only through a fast-decaying term. All proofs in this section are deferred to Appendices B and D.

5.1 Theoretical CVaR Bounds

In this section, we establish bounds for the CVaR of the random variable X by leveraging an auxiliary random variable Y, under the condition that their CDFs differ by at most a known uniform bound ϵ. These deterministic bounds provide the theoretical basis for the finite-sample concentration inequalities developed in Section 5.3.

Theorem 5.1 bounds $\text{CVaR}_\alpha(X)$ in terms of $\text{CVaR}_\alpha(Y)$ under the sole condition that the cumulative distribution functions of X and Y differ by at most a known uniform bound. This representation enhances the interpretability of the bound and constitutes a novel aspect of the result, made possible by framing the bounding problem in terms of distributional discrepancies. In Section 5.3, we demonstrate that the bounds established in Thomas and Learned-Miller (2019) (Theorems 3.2 and 3.3) arise as a special case of Theorem 5.1, thereby providing an interpretation for existing CVaR bounds that are otherwise difficult to interpret.

Theorem 5.1. *Let X and Y be random variables and $\epsilon \in [0, 1]$.*

(1) ***Upper Bound:*** *assume that finite $b_X, b_Y \in \mathbb{R}$ satisfy $b_X \geq \operatorname{ess\,sup} X$ and $b_Y \geq \operatorname{ess\,sup} Y$, and $\forall z \in \mathbb{R}, F_Y(z) - F_X(z) \leq \epsilon$, then*
   (a) *If $\alpha > \epsilon$,*
   $$\text{CVaR}_\alpha(X) \leq \frac{\epsilon}{\alpha} \max(b_X, b_Y) + (1 - \frac{\epsilon}{\alpha}) \text{CVaR}_{\alpha-\epsilon}(Y). \tag{23}$$
   (b) *If $\alpha \leq \epsilon$, then $\text{CVaR}_\alpha(X) \leq \max(b_X, b_Y)$.*

(2) ***Lower Bound:*** *If finite $a_X, a_Y \in \mathbb{R}$ satisfy $a_X \leq \operatorname{ess\,inf} X$ and $a_Y \leq \operatorname{ess\,inf} Y$, and $\forall z \in \mathbb{R}, F_X(z) - F_Y(z) \leq \epsilon$, then*
   (a) *If $\alpha + \epsilon < 1$,*
   $$\text{CVaR}_\alpha(X) \geq (1 + \frac{\epsilon}{\alpha}) \text{CVaR}_{\alpha+\epsilon}(Y) - \frac{\epsilon}{\alpha} \text{CVaR}_\epsilon(Y) \tag{24}$$
   (b) *If $\alpha + \epsilon \geq 1$ and $a_{min} = \min(a_X, a_Y)$,*
   $$\text{CVaR}_\alpha(X) \geq \frac{1}{\alpha}\big(\mathbb{E}[Y] - \epsilon \,\text{CVaR}_\epsilon(Y) + (\alpha + \epsilon - 1)a_{min}\big) \tag{25}$$

Proof. The proof is available in Appendix B. □

Remark 5.2 (Interpretation of ϵ). *The parameter $\epsilon \in [0, 1]$ quantifies the CDF discrepancy between X and Y. When $\epsilon = 0$, the CDFs are identical and the bounds are exact: $\text{CVaR}_\alpha(X) = \text{CVaR}_\alpha(Y)$. When $\epsilon = 1$, the bounds reduce to the trivial support bounds. For the bounds to be informative, we need $\alpha > \epsilon$ (upper) and $\alpha+\epsilon \leq 1$ (lower). Under these conditions, the upper bound (23) is a weighted average of $\text{CVaR}_{\alpha-\epsilon}(Y)$ and $\max(b_X, b_Y)$ with weights proportional to ϵ/α, while the lower bound (24) equals $\text{CVaR}_{\alpha+\epsilon}(Y)$ plus a correction $(\epsilon/\alpha)(\text{CVaR}_{\alpha+\epsilon}(Y) - \text{CVaR}_\epsilon(Y))$. Both bounds converge to $\text{CVaR}_\alpha(X)$ as $\epsilon \to 0$ (Theorem 5.3).*

Theorem 5.3. *Under the definitions of X, Y, ϵ, and α specified in Theorem 5.1, the lower and upper bounds established therein converge to $\text{CVaR}_\alpha(X)$ as $\epsilon \to 0$.*

Proof. The proof is available in Appendix B. □

The uniform discrepancy ϵ in Theorem 5.1 can be conservative when the actual CDF gap varies with x. When a tighter pointwise bound $F_X(x) \leq F_Y(x) + g(x)$ is available for a non-negative function g, sharper CVaR bounds can be obtained via a stochastic dominance construction. Since these pointwise refinements are not used in subsequent sections, we defer them to Appendix C (Theorems C.1 and C.2).

5.2 Truncation Decomposition

The deterministic bounds in Theorem 5.1 require knowledge of the support of X. When combined with the DKW inequality for finite-sample estimation, the support enters at $O(1/\sqrt{n})$: in Theorem 5.1 the support bound $\max(b_X, b_Y)$ carries the weight ϵ/α, and replacing ϵ by the DKW radius $\eta_{\mathrm{DKW}} = \sqrt{\ln(1/\delta_1)/(2n)}$ makes this weight $\sqrt{\ln(1/\delta_1)/(2n)}/\alpha = O(1/\sqrt{n})$, so an overestimate of the support by Δb widens the bound by $\Delta b \cdot O(1/\sqrt{n})$, rendering the bounds loose when the support is conservatively overestimated. We now introduce a truncation decomposition that confines the support dependence to an $O(1/n)$ tail correction.

Proposition 5.4 (Upper Truncation Decomposition). *Let X be a random variable and $b \in \mathbb{R}$ with $b \geq \operatorname{ess\,sup} X$, and let $\kappa \in \mathbb{R}$. Define $X_\kappa \triangleq \min(X, \kappa)$ and $p \triangleq \mathbb{P}[X > \kappa]$. Then:*

$$\mathrm{CVaR}_\alpha(X) = \mathrm{CVaR}_\alpha(X_\kappa) + \frac{1}{\alpha}\int_{1-\alpha}^{1} \left[F_X^{-1}(v) - \kappa\right]^+ dv. \tag{26}$$

Consequently,

$$\mathrm{CVaR}_\alpha(X_\kappa) \leq \mathrm{CVaR}_\alpha(X) \leq \mathrm{CVaR}_\alpha(X_\kappa) + \frac{\min(p,\, \alpha)}{\alpha}\,(b - \kappa). \tag{27}$$

Proof. The proof is available in Appendix D. □

Proposition 5.5 (Lower Truncation Decomposition). *Let X be a random variable and $a \in \mathbb{R}$ with $a \leq \operatorname{ess\,inf} X$, and let $\kappa \in \mathbb{R}$. Define $X^\kappa \triangleq \max(X, \kappa)$ and $q \triangleq \mathbb{P}[X < \kappa]$. Then:*

$$\mathrm{CVaR}_\alpha(X^\kappa) - \frac{\min(q,\, \alpha)}{\alpha}\,(\kappa - a) \leq \mathrm{CVaR}_\alpha(X) \leq \mathrm{CVaR}_\alpha(X^\kappa). \tag{28}$$

Proof. The proof is available in Appendix D. □

The key insight is that $\mathrm{CVaR}_\alpha(X_\kappa)$ satisfies $\mathbb{P}[X_\kappa \leq \kappa] = 1$ by construction, so existing bounds apply with no conservatism. When κ is taken as the sample maximum, the tail correction $p = O(1/n)$ via order statistics.

5.3 Support-Relaxed CVaR Concentration Bounds

The deterministic bounds in Theorem 5.1 involve the true CVaR of Y. In practice, we observe only a finite sample $Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} F_Y$. The ECDF based on these samples is defined by $\hat{F}_Y(y) \triangleq \frac{1}{n}\sum_{i=1}^{n} 1_{Y_i \leq y}$, for $y \in \mathbb{R}$, and we write $\mathrm{CVaR}_\alpha^{\hat{F}_Y}$ for the CVaR computed under the empirical distribution $\hat{F}_Y$ (analogous notation applies for any random variable).

The DKW-based concentration bounds (Appendix A) require known lower and upper bounds on the support. In each concentration bound, one of these support bounds is multiplied by a term of order $O(1/\sqrt{n})$. A conservative lower or upper support bound can therefore produce a wide CVaR interval. Our support-relaxed bounds replace the lower and upper support bounds in the main CVaR terms with the sample minimum and sample maximum, respectively. The lower and upper support bounds are used only to account for values beyond the corresponding sample extreme. When $\epsilon = 0$, the probability of the relevant unobserved tail is at most $1 - \delta_2^{1/n} = O(1/n)$, so the contribution of the support bound decreases at rate $O(1/n)$.

Theorem 5.6 (Support-Relaxed CVaR Bounds via Auxiliary Distributions). *Let X and Y be random variables. Let $Y_1, \ldots, Y_n \overset{\mathrm{iid}}{\sim} F_Y$ and let $\delta_1, \delta_2 \in (0, 0.5)$, $\alpha \in (0, 1]$, $\epsilon \geq 0$. Define $\hat{b}_Y \triangleq \max_i Y_i$, $\hat{a}_Y \triangleq \min_i Y_i$. Let $\eta \geq 0$. Define the effective discrepancy $\epsilon' \triangleq \min(\epsilon + \eta,\, 1)$ and the tail factors $\hat{p} \triangleq \hat{q} \triangleq 1 - \delta_2^{1/n} + \epsilon$.*

Upper Bound. *Assume* $\sup_z(F_Y(z) - F_X(z)) \le \epsilon$, $\mathbb{P}[\sup_z(\hat{F}_Y(z) - F_Y(z)) \le \eta] > 1 - \delta_1$, *suppose* $b_X \ge \operatorname{ess\,sup} X$, *and let* $\bar{b} \ge b_X$. *If* $\epsilon' < \alpha$:

$$U \triangleq \underbrace{\left(1 - \frac{\epsilon'}{\alpha}\right) \mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\epsilon'} + \frac{\epsilon'}{\alpha}\,\hat{b}_Y}_{U_{\mathrm{trunc}}} \;+\; \underbrace{\frac{\hat{p}}{\alpha}\,(\bar{b} - \hat{b}_Y)^+}_{\Delta_U}. \tag{29}$$

Then $\mathbb{P}[\mathrm{CVaR}_\alpha(X) \le U] > 1 - \delta_1 - \delta_2$.

Lower Bound. *Assume* $\sup_z(F_X(z) - F_Y(z)) \le \epsilon$, $\mathbb{P}[\sup_z(F_Y(z) - \hat{F}_Y(z)) \le \eta] > 1 - \delta_1$, F_Y *continuous,* $a_X \le \operatorname{ess\,inf} X$, *and* $\underline{a} \le a_X$. *If* $\epsilon' + \alpha < 1$:

$$L \triangleq \underbrace{\left(1 + \frac{\epsilon'}{\alpha}\right) \mathrm{CVaR}^{\hat{F}_Y}_{\alpha+\epsilon'} - \frac{\epsilon'}{\alpha}\,\mathrm{CVaR}^{\hat{F}_Y}_{\epsilon'}}_{L_{\mathrm{trunc}}} \;-\; \underbrace{\frac{\hat{q}}{\alpha}\,(\hat{a}_Y - \underline{a})^+}_{\Delta_L}. \tag{30}$$

Then $\mathbb{P}[\mathrm{CVaR}_\alpha(X) \ge L] > 1 - \delta_1 - \delta_2$.

Proof sketch. We truncate X at the largest sample of Y, $\hat{b}_Y$, and split $\mathrm{CVaR}_\alpha(X)$ by the truncation decomposition (Proposition 5.4) into two terms: the CVaR of the truncated $X_{\hat{b}_Y}$, which Theorem 5.1 bounds by the empirical CVaR of Y, and a remainder for the tail of X above $\hat{b}_Y$. With $p \triangleq \mathbb{P}[X > \hat{b}_Y]$, $\kappa = \hat{b}_Y$, $b = \bar{b}$, and $\min(p, \alpha) \le p$,

$$\mathrm{CVaR}_\alpha(X) \le \mathrm{CVaR}_\alpha(X_{\hat{b}_Y}) + \frac{p}{\alpha}(\bar{b} - \hat{b}_Y)^+ \quad \text{(Prop. 5.4)}. \tag{31}$$

Using Theorem 5.1, we bound $\mathrm{CVaR}_\alpha(X_{\hat{b}_Y})$ and obtain

$$\mathrm{CVaR}_\alpha(X_{\hat{b}_Y}) \le \frac{\epsilon'}{\alpha}\hat{b}_Y + \left(1 - \frac{\epsilon'}{\alpha}\right) \mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\epsilon'} = U_{\mathrm{trunc}}. \tag{32}$$

The lower bound is proved similarly. *The full proof is in Appendix D.* □

The bound separates two sources of error: ϵ quantifies the distributional discrepancy between X and Y, while η quantifies the finite-sample error in estimating F_Y from n samples. The bound decomposes into a *data-driven term* (U_{trunc} or L_{trunc}) that captures the bulk of the CVaR estimate without any support assumption, converging at $O(1/\sqrt{n})$; and a *tail correction* (Δ_U or Δ_L) where the support bound $\bar{b}$ (or $\underline{a}$) is confined to the tail correction, isolated from the dominant data-driven term. The tail factor $\hat{p} \triangleq 1 - \delta_2^{1/n} + \epsilon$ comprises an $O(1/n)$ statistical term and the discrepancy ϵ; when $\epsilon = 0$ the entire tail correction decays at $O(1/n)$ (see Corollary 5.7). Any CDF concentration inequality yields a valid η; the canonical choice is the DKW inequality, giving $\eta_{\mathrm{DKW}} \triangleq \sqrt{\ln(1/\delta_1)/(2n)}$. Tighter local alternatives apply as well; see Remark 5.9.

Setting $Y = X$ and $\epsilon = 0$ in Theorem 5.6 yields standalone CVaR concentration bounds from i.i.d. samples alone:

Corollary 5.7 (Support-Relaxed CVaR Concentration). *Let* $X_1, \ldots, X_n \overset{\mathrm{iid}}{\sim} F_X$ *and let* $\delta_1, \delta_2 \in (0, 0.5)$, $\alpha \in (0, 1]$. *Define* $\eta \triangleq \sqrt{\ln(1/\delta_1)/(2n)}$, $\hat{b} \triangleq \max_i X_i$, $\hat{a} \triangleq \min_i X_i$, *and* $\hat{p} \triangleq \hat{q} \triangleq 1 - \delta_2^{1/n}$.

(1) ***Upper Bound.*** *If* $\eta < \alpha$, $b_X \ge \operatorname{ess\,sup} X$ *for some finite* b_X *and* $\bar{b} \ge b_X$:

$$\mathbb{P}\left[\mathrm{CVaR}_\alpha(X) \le \left(1 - \frac{\eta}{\alpha}\right) \mathrm{CVaR}^{\hat{F}_X}_{\alpha-\eta} + \frac{\eta}{\alpha}\,\hat{b} + \frac{\hat{p}}{\alpha}(\bar{b} - \hat{b})^+\right] \ge 1 - \delta_1 - \delta_2. \tag{33}$$

(2) ***Lower Bound.*** *If* $\alpha + \eta < 1$, F_X *is continuous,* $a_X \le \operatorname{ess\,inf} X$ *for some finite* a_X *and* $\underline{a} \le a_X$:

$$\mathbb{P}\left[\mathrm{CVaR}_\alpha(X) \ge \left(1 + \frac{\eta}{\alpha}\right) \mathrm{CVaR}^{\hat{F}_X}_{\alpha+\eta} - \frac{\eta}{\alpha}\,\mathrm{CVaR}^{\hat{F}_X}_{\eta} - \frac{\hat{q}}{\alpha}(\hat{a} - \underline{a})^+\right] \ge 1 - \delta_1 - \delta_2. \tag{34}$$

With $\epsilon = 0$ the tail factor reduces to $\hat{p} \triangleq \hat{q} \triangleq 1 - \delta_2^{1/n}$, which is $O(1/n)$ at every finite n: applying $xe^{-x} \leq 1 - e^{-x} \leq x$ for $x \geq 0$ with $x = \ln(1/\delta_2)/n$ gives

$$\delta_2^{1/n} \frac{\ln(1/\delta_2)}{n} \;\leq\; 1 - \delta_2^{1/n} \;\leq\; \frac{\ln(1/\delta_2)}{n}, \tag{35}$$

so the tail correction in Corollary 5.7 satisfies $\Delta_U \leq \ln(1/\delta_2)\,(\bar{b} - \hat{b})^+/(\alpha n)$ for every $n \geq 1$, and the conservative support bound $\bar{b}$ (or $\underline{a}$) enters only through an $O(1/n)$ tail correction. In contrast, the equivalent CVaR formulation of the bounds of Thomas and Learned-Miller (2019) (Theorem A.1) has the form $(1 - \frac{\epsilon}{\alpha})\,\mathrm{CVaR}_{\alpha-\epsilon}^{\hat{F}_X} + \frac{\epsilon}{\alpha}\,b$ with $\epsilon = \sqrt{\ln(1/\delta)/(2n)} = O(1/\sqrt{n})$, so the support bound b is multiplied by $\epsilon/\alpha = O(1/\sqrt{n})$. Corollary 5.7 therefore strictly improves upon these bounds when the support is overestimated, which is the common case: the returns of a given policy typically fall in a much narrower range than the support bound $[TR_{\min}, TR_{\max}]$ implied by the per-step cost range and the horizon.

Extension to estimated discrepancy. Theorem 5.6 assumes that the CDF discrepancy ϵ between X and Y is known. In practice—for instance, in the POMDP simplification setting of Section 7—ϵ must be estimated from data. The following theorem extends Theorem 5.6 to this setting by introducing a third confidence budget for the discrepancy estimation.

Theorem 5.8 (Support-Relaxed CVaR Bounds with Estimated Discrepancy). *Let X and Y be random variables, $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} F_Y$, $\delta_1, \delta_2, \delta_3 \in (0, 0.5)$, $\alpha \in (0, 1]$, and $\eta \geq 0$. Define $\hat{b}_Y \triangleq \max_i Y_i$ and $\hat{a}_Y \triangleq \min_i Y_i$.*

Upper Bound. *Define $\epsilon \triangleq \sup_z (F_Y(z) - F_X(z))$, and let $\hat{\epsilon} \geq 0$ be an estimator of ϵ. Define $\hat{\epsilon}' \triangleq \min(\hat{\epsilon} + \eta, 1)$ and $\hat{p} \triangleq 1 - \delta_2^{1/n} + \hat{\epsilon}$. Assume*

$$\mathbb{P}[\hat{\epsilon} \geq \epsilon] \geq 1 - \delta_3, \qquad \mathbb{P}\big[\sup_z (\hat{F}_Y(z) - F_Y(z)) \leq \eta\big] > 1 - \delta_1, \qquad \bar{b} \geq b_X \geq \operatorname{ess\,sup} X. \tag{36}$$

If $\hat{\epsilon}' < \alpha$:

$$U \triangleq \underbrace{\left(1 - \frac{\hat{\epsilon}'}{\alpha}\right) \mathrm{CVaR}_{\alpha-\hat{\epsilon}'}^{\hat{F}_Y} + \frac{\hat{\epsilon}'}{\alpha}\,\hat{b}_Y}_{U_{\text{trunc}}} \;+\; \underbrace{\frac{\hat{p}}{\alpha}\,(\bar{b} - \hat{b}_Y)^+}_{\Delta_U}. \tag{37}$$

If $\hat{\epsilon}' \geq \alpha$, define $U \triangleq \hat{b}_Y + \frac{\hat{p}}{\alpha}(\bar{b} - \hat{b}_Y)^+$. Then $\mathbb{P}[\mathrm{CVaR}_\alpha(X) \leq U] > 1 - \delta_1 - \delta_2 - \delta_3$.

Lower Bound. *Define $\epsilon \triangleq \sup_z (F_X(z) - F_Y(z))$, and let $\hat{\epsilon} \geq 0$ be an estimator of ϵ. Define $\hat{\epsilon}' \triangleq \min(\hat{\epsilon} + \eta, 1)$ and $\hat{q} \triangleq 1 - \delta_2^{1/n} + \hat{\epsilon}$. Assume F_Y is continuous and*

$$\mathbb{P}[\hat{\epsilon} \geq \epsilon] \geq 1 - \delta_3, \qquad \mathbb{P}\big[\sup_z (F_Y(z) - \hat{F}_Y(z)) \leq \eta\big] > 1 - \delta_1, \qquad \underline{a} \leq a_X \leq \operatorname{ess\,inf} X. \tag{38}$$

If $\hat{\epsilon}' + \alpha < 1$:

$$L \triangleq \underbrace{\left(1 + \frac{\hat{\epsilon}'}{\alpha}\right) \mathrm{CVaR}_{\alpha+\hat{\epsilon}'}^{\hat{F}_Y} - \frac{\hat{\epsilon}'}{\alpha}\,\mathrm{CVaR}_{\hat{\epsilon}'}^{\hat{F}_Y}}_{L_{\text{trunc}}} \;-\; \underbrace{\frac{\hat{q}}{\alpha}\,(\hat{a}_Y - \underline{a})^+}_{\Delta_L}. \tag{39}$$

If $\hat{\epsilon}' + \alpha \geq 1$, define $L \triangleq \underline{a}$. Then $\mathbb{P}[\mathrm{CVaR}_\alpha(X) \geq L] > 1 - \delta_1 - \delta_2 - \delta_3$.

Proof. The proof is available in Appendix D. □

Theorem 5.8 has the same structure as Theorem 5.6, with ϵ replaced by $\hat{\epsilon}$ and an additional δ_3 in the confidence budget. On the event $\{\hat{\epsilon} \geq \epsilon\}$, which holds with probability $\geq 1 - \delta_3$, the effective discrepancy $\hat{\epsilon}' \geq \epsilon' = \epsilon + \eta$ is a valid upper bound on the true effective discrepancy, and all steps of the proof of Theorem 5.6 apply. The monotonicity of the bounds in ϵ' (the upper bound is non-decreasing in ϵ' and the lower bound is non-increasing)

ensures that overestimating ϵ produces valid but possibly looser bounds, while the $O(1/n)$ tail correction structure is preserved.

Remark 5.9 (Local concentration refinements). *Since* CVaR_α *depends on quantiles only in* $[1-\alpha, 1]$*, the global DKW inequality can be replaced by tighter* local *alternatives. The canonical DKW choice gives* $\eta_{\mathrm{DKW}} \triangleq \sqrt{\ln(1/\delta_1)/(2n)}$*. Two tighter options are: (i) a* restricted Kolmogorov–Smirnov (KS) statistic *(*η_{rKS}*) that restricts the supremum to* $[1-\alpha, 1]$*, yielding an asymptotically valid but tighter parameter; and (ii) an* exact local DKW *parameter (*η_{lDKW}*) (Maillard 2021) that provides a finite-sample-exact concentration adapted to the CVaR quantile region, yielding 56–88% tighter bounds on the distributions evaluated in Section 9. These refinements apply directly to Theorem 5.6 and Corollary 5.7 by substituting* η_{lDKW} *or* η_{rKS} *for* η_{DKW} *in the effective discrepancy* $\epsilon' = \epsilon + \eta$*. Full derivations and closed-form expressions for all three instantiations are provided in Appendix H.*

6 Value Function Bounds for Static CVaR

In this section, we leverage the CVaR bounds from the previous section to establish bounds between the simplified and original value functions as defined in (17), where the simplification is considered in a general form with respect to the belief transition model. We then derive bounds for the specific case in which the simplification applies to the observation model. The implications of these bounds for policy evaluation and open-loop POMDP planning will be examined in the Experiments section.

6.1 Bounds for a General Belief Transition Model Simplification

In this section, we establish bounds on the difference between the CDFs of the returns, computed with respect to the simplified and original belief transition models. These bounds are then used to bound the original value function in terms of the simplified value function.

Theorem 6.1 demonstrates that the difference between the CDFs of the returns, computed with respect to the simplified and original belief transition models, can be characterized in terms of the difference between the simplified and original belief transition models.

Theorem 6.1. *Let* $\mathbb{P}^{\pi}_{b_k,a_k}$ *and* $\mathbb{P}^{\pi,s}_{b_k,a_k}$ *denote the trajectory measures induced by the original and simplified belief-transition kernels, and let* $\mathbb{E}^{\pi,s}_{b_k,a_k}$ *denote the expectation under the simplified one. Then, the following holds:*

$$\sup_{l\in\mathbb{R}} |\mathbb{P}^{\pi}_{b_k,a_k}[R_{k:T} \le l] - \mathbb{P}^{\pi,s}_{b_k,a_k}[R_{k:T} \le l]| \le \sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_t, a_t)], \tag{40}$$

where Δ^s *is the TV distance that is defined by*

$$\Delta^s(b_{t-1}, a_{t-1}) \triangleq \tfrac{1}{2} \int_{b_t \in B} |P(b_t|b_{t-1}, a_{t-1}) - P_s(b_t|b_{t-1}, a_{t-1})| db_t. \tag{41}$$

Proof. The proof is available in Appendix E. □

The expression $\mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_t, a_t)]$ quantifies the expected one-step distributional discrepancy at time t between the transition dynamics of the simplified and original belief-MDPs, conditioned on the planning process being initiated at time k. Summing this quantity over the horizon accumulates the total distributional discrepancy, leading to the bound in (40), which in turn bounds the cumulative distribution function of the return across the horizon. This accumulation mechanism is illustrated in Figure 4.

By leveraging Theorem 5.1 and combining it with the supremum norm bounds for the simplified and original return CDFs in (40), we derive the following bounds on the value functions (Theorem 6.2). Finite deterministic bounds d_{min} and d_{max} exist under both models because the cost is bounded, $R_{\min} \le c_x(x, a_t) \le R_{\max}$, and the

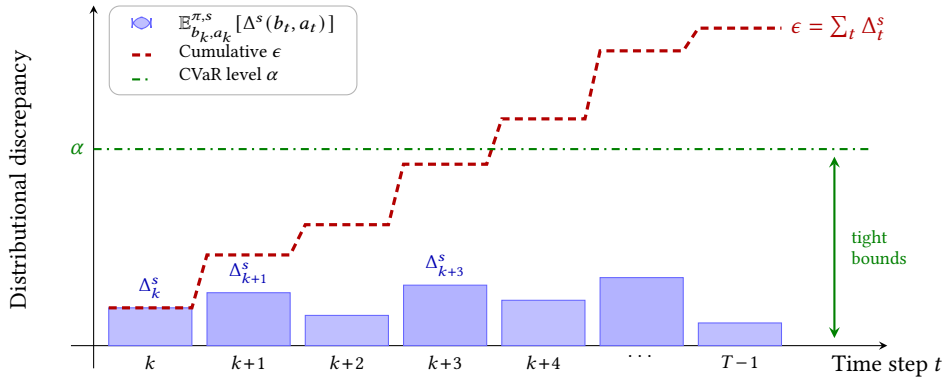


Fig. 4. Accumulation of distributional discrepancy over the planning horizon. Each bar represents the expected one-step TV distance $\mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_t,a_t)]$ at time t. The cumulative sum $\epsilon = \sum_{t=k}^{T-1} \Delta^s_t$ (dashed) governs the bound quality: when $\epsilon < \alpha$ (below the dash-dotted line), the CVaR bounds from Theorem 6.2 remain informative; when $\epsilon \geq \alpha$, the bounds reduce to trivial extrema.

horizon is finite. For nonnegative discount weights, $d_{min} = R_{\min} \sum_{j=0}^{T-k} \gamma^j$ and $d_{max} = R_{\max} \sum_{j=0}^{T-k} \gamma^j$ are sufficient choices.

THEOREM 6.2. *Let $d_{min}, d_{max} \in \mathbb{R}$ be deterministic bounds such that $d_{min} \leq R_{k:T} \leq d_{max}$ almost surely under both $\mathbb{P}^{\pi}_{b_k,a_k}$ and $\mathbb{P}^{\pi,s}_{b_k,a_k}$, and denote*

$$\epsilon \triangleq \epsilon(b_k, a_k) \triangleq \min\Big(\sum_{i=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_i,a_i)], 1\Big), \tag{42}$$

(1) ***Upper Bound:***
   (a) *If $\epsilon < \alpha$, then $U_s \triangleq (1 - \frac{\epsilon}{\alpha}) Q^{\pi}_{M_s}(b_k, a_k, \alpha - \epsilon) + \frac{\epsilon}{\alpha} d_{max}$*
   (b) *If $\epsilon \geq \alpha$ then $U_s \triangleq d_{max}$*

(2) ***Lower Bound:***
   (a) *If $\epsilon + \alpha < 1$ then $L_s \triangleq (1 + \frac{\epsilon}{\alpha}) Q^{\pi}_{M_s}(b_k, a_k, \alpha + \epsilon) - \frac{\epsilon}{\alpha} Q^{\pi}_{M_s}(b_k, a_k, \epsilon)$*
   (b) *If $\epsilon + \alpha \geq 1$ then $L_s \triangleq \frac{1}{\alpha}[(\alpha + \epsilon - 1) d_{min} + \mathbb{E}^{\pi,s}_{b_k,a_k}[R_{k:T}] - \epsilon Q^{\pi}_{M_s}(b_k, a_k, \epsilon)]$*

where $\mathbb{E}^{\pi,s}_{b_k,a_k}[R_{k:T}] = Q^{\pi}_{M_s}(b_k, a_k, 1)$ denotes the expected return under the simplified model. For $\epsilon = 0$, interpret $\epsilon Q^{\pi}_{M_s}(b_k, a_k, \epsilon)$ as zero. Then $L_s \leq Q^{\pi}_{M}(b_k, a_k, \alpha) \leq U_s$ deterministically.

PROOF. The proof is available in Appendix E. □

The bounds established in Theorem 6.2 provide upper and lower bounds on the original value function in terms of the simplified value function, evaluated at an adjusted confidence level that accounts for the distributional discrepancy ϵ between the simplified and original belief transition models.

6.2 Bounds for Observation Model Simplification

Although observation model simplification is a special case of belief-transition simplification, computing the belief-transition kernel directly remains intractable in general. Therefore, the bounds from Theorem 6.2 must be specialized to this setting.

Let q_z denote the simplified observation model associated with the POMDP M_s, and let p_z denote the original observation model associated with the POMDP M. The bounds for observation model simplification are exhibited in Theorem 6.3.

THEOREM 6.3. *Both models update the belief with the same kernel given the new observation, $P(b_{t+1}|b_t, a_t, z_{t+1}) = P_s(b_{t+1}|b_t, a_t, z_{t+1})$, and differ only in the observation model: p_z under M, q_z under M_s. In the case where the simplified and original observation models are denoted by q_z and p_z respectively, it holds that*

$$\sup_{l\in\mathbb{R}} |\mathbb{P}^{\pi}_{b_k,a_k}[R_{k:T} \leq l] - \mathbb{P}^{\pi,s}_{b_k,a_k}[R_{k:T} \leq l]| \leq \sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s_x(x_{t+1})], \tag{43}$$

where Δ^s_x is the state-level TV distance defined by

$$\Delta^s_x(x_t) \triangleq \tfrac{1}{2} \int_{z_t\in Z} |p_z(z_t|x_t) - q_z(z_t|x_t)| dz_t, \tag{44}$$

and x_{t+1} is drawn by $x_t \sim b_t$, $x_{t+1} \sim P(\cdot|x_t, a_t)$, with (b_t, a_t) from the simplified trajectory started at (b_k, a_k). Note that Δ^s_x differs from the belief-level TV distance Δ^s in (41): the former measures discrepancy at a single state, while the latter integrates over belief-weighted transitions.

PROOF. The proof is available in Appendix E. □

Theorem 6.3 establishes a state-dependent upper bound on the sup-norm difference between the CDFs of the return. This bound is amenable to practical computation. The state x_{t+1} appearing in (43) can be sampled from the distribution $P(\cdot \mid x_t, \pi_t) \prod_{i=k+1}^{t-1} P(x_{i+1} \mid x_i, \pi_i) P(x_k \mid b_k)$, where π_i denotes the action selected by policy π at time i.

By leveraging Theorem 6.2 with Theorem 6.3, we obtain Corollary 6.4, which provides computable CVaR bounds based on state-dependent total variation distance.

COROLLARY 6.4. *The bound from Theorem 6.2 holds for $\epsilon(b_k, a_k) = \min(\sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s_x(x_{t+1})], 1)$.*

PROOF. The proof is available in Appendix E.1. □

7 Online Bound Estimation

In this section, we derive practical estimators corresponding to the theoretical bounds presented in Section 6, and describe how these bounds can be efficiently computed within the context of online policy evaluation and open-loop planning. All estimators are constructed from samples generated by particle-based belief estimators drawn under the simplified belief-transition distribution. Specifically, we develop estimators for the action–value function, the distributional discrepancy ϵ appearing in Corollary 6.4, and the minimum and maximum values of the return, thereby enabling computation of the complete bound stated in Corollary 6.4. Moreover, we provide performance guarantees for these estimators, establishing their reliability when computed from a finite number of samples.

7.1 Action-Value Function Estimator

The action-value function is defined as the CVaR of the return $R_{k:T} \triangleq \sum_{t=k}^{T} \gamma^{t-k} c(b_t, a_t)$, and can be estimated in a straightforward manner by generating sample trajectories of the return and computing the empirical CVaR from these samples.

To generate return samples, it is necessary to simulate belief trajectories from time k to time T. This is achieved using a particle-based belief estimator, where the belief is represented by a weighted set of $N_p \in \mathbb{N}$ particles. Given a particle belief $\bar{b}_k$, provided by the user to represent the agent's belief at time k, the agent generates N_b

simulated particle-belief trajectories $\{\bar{b}_t^i\}$, where $i \in \{1, \dots, N_b\}$ indexes the trajectory and $t \in \{k, \dots, T\}$ denotes the time step. For each simulated trajectory, a return sample is computed as $\bar{R}_{k:T}^i \triangleq \sum_{t=k}^{T} c(\bar{b}_t^i, \pi(\bar{b}_t^i))$, resulting in a collection of return samples $\{\bar{R}_{k:T}^i\}_{i=1}^{N_b}$. Let $\bar{b}_t^i \triangleq \{(x_t^{i,j}, w_t^{i,j})\}_{j=1}^{N_p}$ denote the i-th simulated particle belief at time t, where j indexes the particles. Define the normalized weights by $\tilde{w}_t^{i,j} \triangleq w_t^{i,j} / \sum_{j=1}^{N_p} w_t^{i,j}$. Assuming the cost function is state-dependent, the cost associated with belief $\bar{b}_t^i$ and action a is given by $c(\bar{b}_t^i, a) \triangleq \sum_{j=1}^{N_p} \tilde{w}_t^{i,j} c(x_t^{i,j}, a)$. By estimating CVaR according to (14), we get the action-value function estimator

$$\hat{Q}_{M_P}^{\pi}(\bar{b}_k, a, \alpha) \triangleq \hat{C}(\{R^i\}_{i=1}^{N_b}) \quad = R^{(N_b)} - \frac{1}{\alpha} \sum_{i=1}^{N_b} (R^{(i)} - R^{(i-1)}) \Big(\frac{i}{N_b} - (1-\alpha) \Big)^+. \tag{45}$$

Detailed pseudo-code for the estimation of the action–value function is presented in Algorithm 1.

7.2 Online CDF Bound Estimation

The bound on the difference between return simplified and original CDFs, established in Theorem 6.3 and denoted by ϵ in Corollary 6.4, plays a central role in bounding the discrepancy between the simplified and original action-value functions. Notably, it is the only term in the bound that explicitly captures the impact of the difference between the simplified and original observation models. A challenge arises from the fact that the quantity ϵ cannot be directly estimated during planning, as its computation requires access to the full observation model p_z, which is typically too expensive to evaluate online. For the purpose of online computation of the bound, we adopt the methodology proposed in Lev-Yehudi et al. (2024), which enables evaluation without real-time access to the original observation model. This is achieved by decoupling the offline sampling phase—conducted using the original observation model—from the online sampling phase, which relies solely on the simplified observation model. Specifically, the offline–online decoupling strategy and the state-level importance-sampling estimator in (46) were introduced by Lev-Yehudi et al. (2024); the belief-level aggregation, horizon-level accumulation, and the probabilistic guarantees developed in the remainder of this section are novel contributions of the present work. This offline–online decoupling procedure is illustrated in Figure 5.

Let Q_0 be a distribution over the state space, and denote the TV-distance expectation given the previous state and action by m. (46) shows how the one-step distributional discrepancy bound m can be computed using importance sampling.

$$m(x_t, a) \triangleq \mathbb{E}_{x_{t+1} \sim P(\cdot|x_t,a)} [\Delta_x^s(x_{t+1})] \quad = \mathbb{E}_{x_{t+1} \sim Q_0} [\frac{P(x_{t+1}|x_t, a)}{Q_0(x_{t+1})} \Delta_x^s(x_{t+1})]. \tag{46}$$

Define the belief-dependent one-step distributional discrepancy m, and the belief-dependent total distributional discrepancy bound of the CDFs difference at time t by ϵ_t.

$$m(b_t, a) \triangleq \mathbb{E}_{x_t \sim b_t} [m(x_t, a)], \tag{47}$$

$$\epsilon_t(b_t, a) \triangleq \sum_{\tau=t}^{T-1} \mathbb{E}_{b_t,a}^{\pi,s} [\Delta_x^s(x_{\tau+1})] = \sum_{\tau=t}^{T-1} \mathbb{E}_{b_t,a}^{\pi,s} [m(b_\tau, a)]. \tag{48}$$

In practice, a set of states $\{x_n^\Delta\}_{n=1}^{N^\Delta} \overset{\text{i.i.d.}}{\sim} Q_0$ is drawn from a user-specified distribution Q_0, and the TV-distance corresponding to each sampled state is computed offline. During online planning, the precomputed TV distance estimates are reweighted via importance sampling, thereby adapting the offline-computed distributional discrepancy to reflect the actual distributional discrepancy encountered by the agent during execution, as described

Algorithm 1 CVaR Value Function Estimation for Policy π

Global Parameters: generative model G (returns next state and reward given state and action), observation likelihood $Z(z \mid a, x')$

Input: Particle belief $\bar{b}$, planning horizon T, policy π, action a, risk level α, number of belief trajectories N_b and number of belief particles N_p.

Output: Estimated value $\hat{V}^{\pi}(\bar{b})$ or $\hat{Q}^{\pi}(\bar{b}, a)$

procedure GENPF($\bar{b}, a$)
  Sample x_0 from $\bar{b}$ with prob. $w_i / \sum_i w_i$
  $z \leftarrow G(x_0, a)$
  for $i = 1$ **to** N_p **do**
    $(x_i', r_i) \leftarrow G(x_i, a)$
    $w_i' \leftarrow w_i \cdot Z(z \mid a, x_i')$
  end for
  $\bar{b}' \leftarrow \{(x_i', w_i')\}_{i=1}^{N_p}$
  $\rho \leftarrow \frac{\sum_i w_i r_i}{\sum_i w_i}$
  return $\bar{b}', \rho$
end procedure
procedure SAMPLENEXTBELIEFS(beliefs, π)
  NextBeliefs $\leftarrow list()$
  for $j = 1$ **to** N_b **do**
    NextBelief $\leftarrow$ GENPF(beliefs[j], π(beliefs[j]))
    append NextBelief to NextBeliefs
  end for
  return NextBeliefs
end procedure
procedure GENBELIEFTRAJECTORIES($\bar{b}, \pi, T$)
  trajectories $\leftarrow list()$
  for $i = 1$ **to** N_b **do**
    append $\bar{b}$ to trajectories
  end for
  for $t = 0$ **to** $T - 1$ **do**
    NextBeliefs $\leftarrow$ SampleNextBeliefs(trajectories[t], π)
    append NextBeliefs to trajectories
  end for
  return trajectories
end procedure

procedure GENERATERETURNSAMPLE($\bar{b}, a, \pi$)
  trajectories $\leftarrow$ GenBeliefTrajectories($\bar{b}, \pi, T$)
  ReturnSample $\leftarrow$ list()
  for $i = 1$ **to** N_b **do**
    Return $\leftarrow$ c(trajectories[0][i], a)
    for $t = 1$ **to** T **do**
      Return += c(trajectories[t][i], π(trajectories[t][i]))
    end for
    append Return to ReturnSample
  end for
  return ReturnSample
end procedure
procedure CVARESTIMATE(samp, α)
  Sort samp in ascending order as $z_1, \ldots, z_n$ with $z_0 = R_{\min} \cdot T$
  return $z_n - \frac{1}{\alpha} \sum_{i=0}^{n-1} (z_{i+1} - z_i) \left(\frac{i}{n} - (1 - \alpha)\right)^+$
end procedure
procedure ESTIMATE$V^{\pi}(\bar{b}, \alpha)$
  return ESTIMATEQ($\bar{b}, \pi(\bar{b}), \alpha$)
end procedure
procedure ESTIMATEQ($\bar{b}, a, \alpha$)
  samp $\leftarrow$ GenerateReturnSample($\bar{b}, a, \pi$)
  return CVARESTIMATE(samp,α)
end procedure

in (49). By utilizing $\{x_n^{\Delta}\}$, (46) and (47) can be estimated as follows

$$\hat{m}(x_t, a) \triangleq \frac{1}{N^{\Delta}} \sum_{n=1}^{N^{\Delta}} \frac{P(x_n^{\Delta} | x_t, a)}{Q_0(x_n^{\Delta})} \Delta_x^s(x_n^{\Delta}), \tag{49}$$

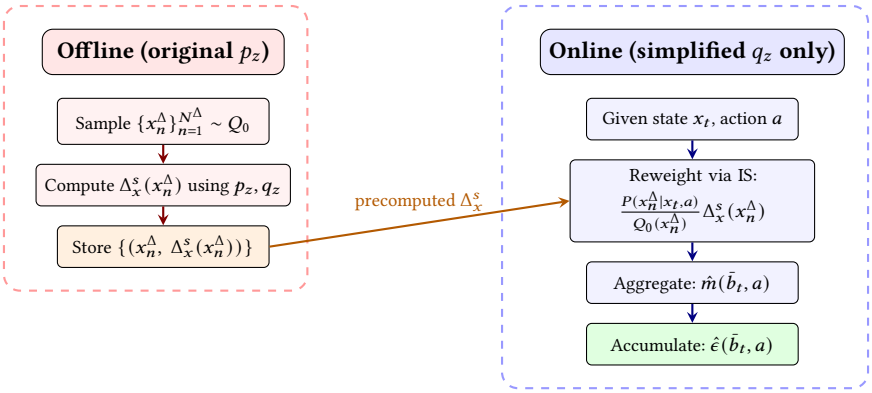


Fig. 5. Offline–online decoupling for distributional discrepancy estimation. In the offline phase, states are sampled from a proposal distribution Q_0 and the state-level TV distances Δ_x^s are computed using the original observation model p_z. During online planning, the precomputed Δ_x^s values are reweighted via importance sampling using only the state-transition model, eliminating the need to evaluate p_z in real time.

$$\hat{m}(\bar{b}_t, a) \triangleq \sum_{i=1}^{N_p} \hat{m}(x_t^i, a)\tilde{w}_t^i, \tag{50}$$

for $\bar{b}_t \triangleq \{(x_t^i, w_t^i)\}$ where $\tilde{w}_t^i$ are the normalized weights. Using the particle-belief trajectories $\{\bar{b}_\tau^i\}_{i=1}^{N_b}$ introduced in Section 7.1, the TV-distance at time $\tau + 1$ is estimated by using $\{\bar{b}_\tau^i\}_{i=1}^{N_b}$ as samples from the conditional distribution of $\bar{b}_\tau$ given $\bar{b}_t$, as described below.

$$\hat{\mathbb{E}}^{\pi,s}_{\bar{b}_t,a}[m(b_\tau, a)] \triangleq \frac{1}{N_b}\sum_{i=1}^{N_b} \hat{m}(\bar{b}_\tau^i, \pi(\bar{b}_\tau^i)). \tag{51}$$

The estimate of ϵ is then computed directly by

$$\hat{\epsilon}(\bar{b}_t, a) \triangleq \sum_{\tau=t}^{T-1} \hat{\mathbb{E}}^{\pi,s}_{\bar{b}_t,a}[m(\bar{b}_\tau, \pi_\tau)]. \tag{52}$$

Theorem 7.1 establishes performance guarantees for the distributional discrepancy estimator defined in (52). Increasing N_b concentrates the estimate around its conditional mean given the offline sample; accuracy relative to the true discrepancy ϵ_t also requires control of the offline error, which shrinks as N^Δ increases.

Theorem 7.1. *Let $S \triangleq \{(x_n^\Delta, \hat{\Delta}_x^s(x_n^\Delta))\}_{n=1}^{N^\Delta}$ denote the offline sample, with $\hat{\Delta}_x^s$ unbiased for Δ_x^s given x_n^Δ. Assume $\bar{b}_t$ and a are fixed independently of S, the online trajectories $\{\bar{b}_\tau^i\}_{i=1}^{N_b}$ are i.i.d. and independent of S, and the pairs $(x_n^\Delta, \hat{\Delta}_x^s(x_n^\Delta))_{n=1}^{N^\Delta}$ are i.i.d. If $T > t$, then for any $u, v > 0$,*

$$\mathbb{P}^{\pi,s}_{\bar{b}_t,a}[|\hat{\epsilon}_t(\bar{b}_t, a) - \epsilon_t(\bar{b}_t, a)| \geq u + v] \leq 2\exp\left(-\frac{2N_b u^2}{D^2(T-t)^2}\right) + 2\exp\left(-\frac{2N^\Delta v^2}{D^2(T-t)^2}\right), \tag{53}$$

for $D = \max_{i\in\{t,\ldots,T-1\}} D_i$ where $D_i = \sup_{x_{i+1}} P(x_{i+1}|x_i, a_i)/Q_0(x_{i+1})$. If $T = t$ the failure probability is zero.

Proof. The proof is available in Appendix E. □

7.3 Performance Guarantees

In this section we provide performance guarantees for the bounds exhibited in previous sections, making them reliable in practice. Combining Theorem 5.8 with the value function bounding framework yields the following result, in which all quantities are computable from samples generated under the simplified model during online planning.

Theorem 7.2 (Support-Relaxed Guarantees with Estimated Discrepancy). *Let b_k be an initial belief and $R^1, \dots, R^{N_b} \sim \mathbb{P}^{\pi,s}_{b_k}$. Let $\hat{\epsilon} \triangleq \hat{\epsilon}(b_k, a_k)$ be the importance-sampling estimator from (52), and assume that $\mathbb{P}[\hat{\epsilon} \geq \epsilon] \geq 1 - \delta_3$ where ϵ is the true distributional discrepancy as defined in (42). Let $\delta_1, \delta_2, \delta_3 \in (0, 0.5)$, and let $\eta \geq 0$. Define $\hat{\epsilon}' \triangleq \min(\hat{\epsilon} + \eta, 1)$, $\hat{p} \triangleq 1 - \delta_2^{1/N_b} + \hat{\epsilon}$, $\hat{q} \triangleq 1 - \delta_2^{1/N_b} + \hat{\epsilon}$, and let $\bar{d}_{\max} \triangleq R_{\max} \sum_{i=0}^{T-k} \gamma^i$, $\underline{d}_{\min} \triangleq R_{\min} \sum_{i=0}^{T-k} \gamma^i$.*

(1) ***Upper Bound:*** *Assume $\mathbb{P}[\sup_z (\hat{F}_{R_s}(z) - F_{R_s}(z)) \leq \eta] > 1 - \delta_1$. If $\hat{\epsilon}' < \alpha$,*

$$U_s^{\mathrm{tr}} \triangleq \left(1 - \frac{\hat{\epsilon}'}{\alpha}\right) \hat{Q}^{\pi}_{M_s}(b_k, a_k, \alpha - \hat{\epsilon}') + \frac{\hat{\epsilon}'}{\alpha} \hat{d}^{\pi}_{\max}(b_k) + \frac{\hat{p}}{\alpha} \left(\bar{d}_{\max} - \hat{d}^{\pi}_{\max}(b_k)\right)^+. \tag{54}$$

If $\hat{\epsilon}' \geq \alpha$, define $U_s^{\mathrm{tr}} \triangleq \hat{d}^{\pi}_{\max}(b_k) + \frac{\hat{p}}{\alpha}(\bar{d}_{\max} - \hat{d}^{\pi}_{\max}(b_k))^+$.

(2) ***Lower Bound:*** *Assume $\mathbb{P}[\sup_z (F_{R_s}(z) - \hat{F}_{R_s}(z)) \leq \eta] > 1 - \delta_1$. If $\hat{\epsilon}' + \alpha < 1$ and the CDF of $R_{k:T}$ under $\mathbb{P}^{\pi,s}_{b_k}$ is continuous,*

$$L_s^{\mathrm{tr}} \triangleq \left(1 + \frac{\hat{\epsilon}'}{\alpha}\right) \hat{Q}^{\pi}_{M_s}(b_k, a_k, \alpha + \hat{\epsilon}') - \frac{\hat{\epsilon}'}{\alpha} \hat{Q}^{\pi}_{M_s}(b_k, a_k, \hat{\epsilon}') - \frac{\hat{q}}{\alpha} \left(\hat{d}^{\pi}_{\min}(b_k) - \underline{d}_{\min}\right)^+. \tag{55}$$

Then $\mathbb{P}^{\pi}_{b_k,a_k}[Q^{\pi}_M(b_k, a_k, \alpha) \leq U_s^{\mathrm{tr}}] > 1 - \delta_1 - \delta_2 - \delta_3$ and $\mathbb{P}^{\pi}_{b_k,a_k}[L_s^{\mathrm{tr}} \leq Q^{\pi}_M(b_k, a_k, \alpha)] > 1 - \delta_1 - \delta_2 - \delta_3$.

Proof. The proof is available in Appendix E. □

Theorem 7.2 combines the truncation decomposition (support enters at $O(1/N_b)$), local CDF concentration (tighter η via Remark 5.9), and estimated distributional discrepancy (no access to the true ϵ required). All quantities—$\hat{Q}^{\pi}_{M_s}$, $\hat{d}^{\pi}_{\max}$, $\hat{d}^{\pi}_{\min}$, and $\hat{\epsilon}$—are computable from samples generated under the simplified model during online planning. The confidence budget is split three ways: δ_1 for CDF concentration, δ_2 for tail estimation, and δ_3 for discrepancy estimation. In the POMDP setting, the condition $\mathbb{P}[\hat{\epsilon} \geq \epsilon] \geq 1 - \delta_3$ can be ensured by constructing $\hat{\epsilon}$ as a high-probability upper bound on the true discrepancy, for example via concentration inequalities on the importance-sampling estimator (52). By Corollary 6.4, the guarantees also apply when only the observation model is simplified.

8 Limitations

The guarantees of this work rest on three assumptions: (i) finite support bounds for the return; (ii) a discrepancy estimate $\hat{\epsilon}$ such that $\hat{\epsilon}/\alpha$ is tight enough, since a loose estimate consumes the risk budget α and loosens the bounds; and (iii) a simplified model for which such a tight estimate is obtainable. The bounds in Theorem 6.2 depend on the accumulation of one-step distributional discrepancies over the planning horizon, which may limit their effectiveness for long-horizon problems.

Importantly, this accumulation does not necessarily grow linearly with the horizon and is highly problem dependent. For example, in a Light–Dark environment, the agent receives observations only in light regions and relies exclusively on the state-transition model in dark regions. As a result, no distributional discrepancy is accumulated in dark regions, since the observation model is not invoked. Consequently, the effective distributional discrepancy can grow sublinearly with the horizon. This behavior is examined empirically in the experiments section, where we also consider a more challenging variant of the Light–Dark environment in which distributional discrepancy accumulates even in dark regions.

Finally, the proposed simplification framework yields computational gains only when sampling from the state-transition model is cheaper than sampling from the observation model. The bounds in Theorem 6.2 require samples from both the simplified observation model and the state-transition model to compensate for the absence of the original observation model. Consequently, if the state-transition model is computationally more expensive than the original observation model, the simplification approach would increase rather than decrease planning time. This limitation is shared by other observation model simplification frameworks (Lev-Yehudi et al. 2024).

This work addresses the static CVaR objective, defined as the CVaR of the cumulative return over the planning horizon. An alternative formulation is the dynamic (nested) CVaR, which applies risk measures recursively at each time step and satisfies time-consistency properties (Ruszczyński 2010). While dynamic CVaR avoids the well-known limitations of static CVaR in sequential settings—such as the inability to decompose the risk measure across time steps in a Bellman-like fashion—it leads to overly conservative policies in practice and requires a fundamentally different algorithmic treatment. Extending the surrogate-model framework to dynamic CVaR objectives is an interesting direction for future work.

9 Experiments

The experiments evaluate the bounds of Section 5 on their own and then the bounds of Sections 6–7 inside a planner. Sections 9.1 and 9.6 compare the support-relaxed bounds with the bounds of Thomas and Learned-Miller (2019) and check their sensitivity with respect to ϵ. Sections 9.4, 9.5, and 9.7 test whether the bounds computed under a simplified observation model separate safe from dangerous action sequences and policies. Sections 9.4 and 9.5 also measure the speedup of bound evaluation under the simplified model relative to the original one. Section 9.3 explains how the distributional discrepancy is measured.

All simulations were conducted using the POMDPPlanners package (Pariente and Indelman 2026). Detailed simulation configurations are provided in Appendix K.4, and hardware specifications in Appendix N.

9.1 Theoretical CVaR Bounds

Specifically, we compare a truncated Gaussian Mixture Model (GMM) and its corresponding truncated Normal approximation for CVaR estimation under bounded support. Such a setting arises, for example, in POMDP planning, where a computationally expensive model used by the agent during online planning is replaced with a more tractable surrogate model, thereby improving the agent's decision-making speed (Lev-Yehudi et al. 2024). In these settings, the agent's decisions may rely on bounds for the original value function that are derived from the tractable surrogate model (Barenboim and Indelman 2022).

The GMM consists of five components with means $\mu_1 = 0.2$, $\mu_2 = -0.2$, $\mu_3 = -0.5$, $\mu_4 = 0.5$, $\mu_5 = 0.0$, variances $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.2$, $\sigma_3^2 = 0.1$, $\sigma_4^2 = 0.1$, $\sigma_5^2 = 0.3$, and weights $w_1 = 0.3$, $w_2 = 0.2$, $w_3 = 0.05$, $w_4 = 0.05$, $w_5 = 0.4$. The Normal approximation matches the GMM's mean and variance but cannot reproduce its multi-modal structure or outlier effects. All samples are truncated to the interval $[-1, 1]$, which reshapes the tails and slightly distorts boundary components. We compute CVaR at the 20% quantile for sample sizes ranging from 100 to 10,000, with 100 independent repetitions per setting to ensure statistical reliability. The distributional discrepancy (ϵ in Theorem A.3) between the truncated GMM and the truncated Normal distribution is assessed via simulation, by estimating their respective cumulative distribution functions over a common set of bins and computing the maximum difference across all bins. This setup enables a direct assessment of the trade-off between computational efficiency and statistical accuracy when approximating a complex truncated mixture by a single truncated Normal distribution.

Figure 6c presents the sampling time comparison between the truncated GMM and the truncated Normal distribution, with an observed average time ratio of approximately 3.7 in favor of the Normal distribution. Figure 6d reports a total computational speedup of approximately 3.1 when comparing the process of sampling from the

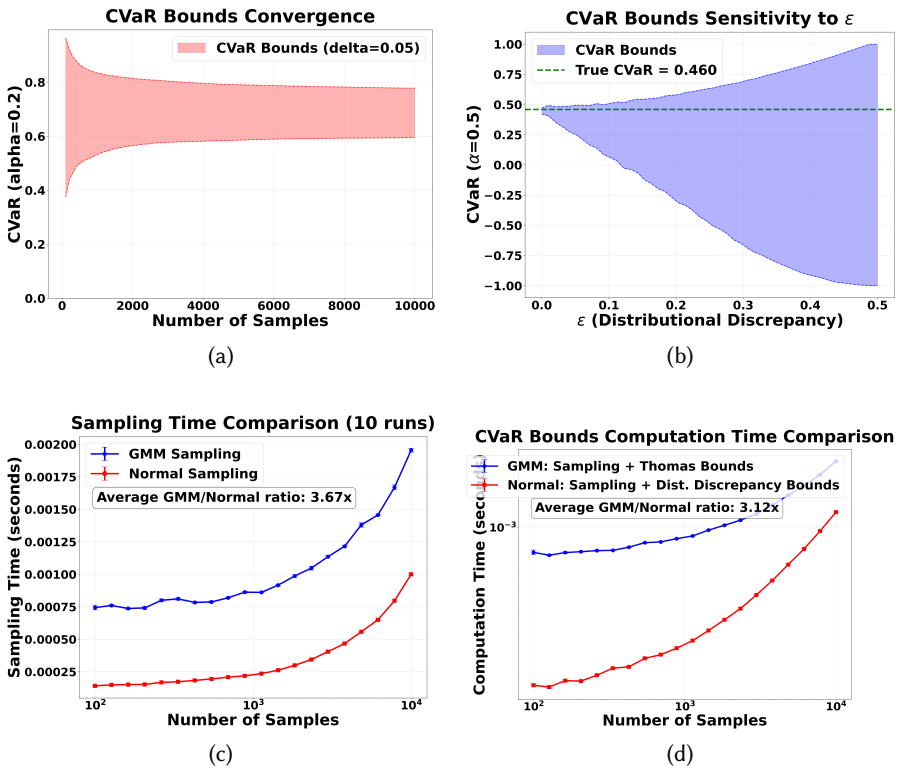


Fig. 6. Figure 6a shows the bounds established in Theorem A.3, demonstrating how a simple truncated Normal distribution can be employed to bound the CVaR of a truncated GMM. Figure 6b illustrates the sensitivity of the CVaR bounds to the distributional discrepancy ϵ, showing how the bound interval widens as ϵ increases, while the true CVaR remains contained within the bounds. When ϵ exceeds the risk level α, the bounds become trivial, spanning the entire support; this is expected, as a large distributional discrepancy implies that the surrogate random variable Y provides negligible information about the target random variable X. Figure 6c presents a comparison of the sampling times for each distribution, while Figure 6d reports the total computation time required to sample from the truncated GMM and estimate its CVaR, in contrast to sampling from the truncated Normal surrogate and computing the associated concentration bounds. Figures 6a, 6c, and 6d were configured with confidence level $\alpha = 0.2$ and probability of error $\delta = 0.05$. Figure 6b uses $\alpha = 0.5$. Figures 6c and 6d display 95% confidence intervals around the empirical means.

truncated GMM and estimating its CVaR to that of sampling from the truncated Normal and computing both upper and lower bounds as given in Theorem A.3. Figure 6a illustrates the convergence behavior of the CVaR bounds as a function of the number of samples, comparing estimates obtained from the truncated GMM with bounds derived from the truncated Normal surrogate. It is important to note that these bounds are obtained without requiring full knowledge of the underlying GMM distribution. Instead, they rely solely on the discrepancy between the corresponding CDFs. Figure 6b depicts the sensitivity of the CVaR bounds to the distributional discrepancy ϵ. In this experiment, the base distribution F is a truncated Normal on $[-1, 1]$ with mean 0 and standard deviation 1. For each value of $\epsilon \in [0, 0.5]$, a perturbed distribution G is constructed such that its CDF satisfies $G(x) = \min(F(x) + \epsilon, 1)$, yielding a Kolmogorov–Smirnov distance of exactly ϵ between F and

G by construction. Samples from G are obtained via inverse transform sampling. The resulting bounds grow monotonically with ϵ while consistently enclosing the true CVaR of F. When ϵ exceeds the risk level α, the bounds become trivial, spanning the full support; this is expected, since a large distributional discrepancy implies that Y provides negligible information about X.

We evaluated the concentration bounds established by Thomas and Learned-Miller (2019) alongside our proposed bounds on a set of probability distributions matching those used in the original study by Thomas and Learned-Miller (2019). As shown in Figure 7, our bounds with the exact local DKW parameter η_{lDKW} exhibit faster convergence than the bounds of Thomas and Learned-Miller (2019). While the bounds of Thomas and Learned-Miller (2019) may be tighter at small sample sizes for some distributions, our bounds ultimately produce tighter intervals as the number of samples increases.

9.2 POMDP Environments

We evaluate our approach on three continuous navigation POMDP environments from the POMDPPlanners package (Pariente and Indelman 2026): 2D Light-Dark Navigation, Laser Tag, and Push. Each environment contains dangerous areas where the likelihood of incurring high penalties is elevated, requiring the agent to balance risk-aware decision-making with task completion. In the 2D Light-Dark POMDP, an agent must navigate from a start position to a goal region while avoiding these dangerous areas. The observation noise is position-dependent, with lower uncertainty near designated beacon locations, requiring the agent to balance information gathering with goal-directed movement. In Laser Tag POMDP, the agent must locate and tag an opponent in a partially observable environment using noisy range-bearing measurements from multiple sensors, resulting in an 8-dimensional observation space, while navigating around dangerous regions. The Push POMDP involves manipulating an object to a target location while managing partial observability about the object's position through noisy 2-dimensional observations and avoiding dangerous areas. For each domain, we construct two environment variants: an original environment with a complex observation model represented by Gaussian Mixture Models (GMMs) capturing multi-modal observation distributions, and a simplified environment where this complex distribution is approximated using a single Gaussian distribution. This approximation introduces a controlled model mismatch, allowing us to evaluate the robustness of planning algorithms when the true observation dynamics differ from the assumed model. The GMM-based observation models enable realistic representation of sensor fusion, occlusions, and multi-hypothesis tracking, while the Gaussian approximation provides computational tractability at the cost of distributional fidelity. Full environment specifications are provided in Appendix K.

9.3 Online Distributional Discrepancy Estimation

To compute the distributional discrepancy ϵ appearing in Corollary 6.4, we define Q_0 as the uniform distribution over the state region of each environment: the 7×7 grid in Light-Dark, the $[0, 5]^2$ square for both the robot and the object in Push, and the arena interior for the robot and the opponent in Laser Tag. We choose the uniform distribution Q_0 to sample the state region without favoring any location, since we make no assumptions about which states the planner will visit. Samples x_n^Δ are drawn from the mixture distribution $(p_Z + q_Z)/2$; that is, with probability 0.5, a sample is drawn from the original GMM observation model, and with probability 0.5, from the simplified observation model. The estimator $\hat{\Delta}^s(x_n^\Delta)$ is then defined as

$$\hat{\Delta}_x^s(x_n^\Delta) \triangleq \frac{1}{N_z} \sum_{j=1}^{N_z} \frac{\left| p_z(z_j^n \mid x_n^\Delta) - q_z(z_j^n \mid x_n^\Delta) \right|}{p_z(z_j^n \mid x_n^\Delta) + q_z(z_j^n \mid x_n^\Delta)}, \tag{56}$$

and is evaluated over $N^\Delta = 100$ presampled states. Here, $N_z = 2000$ denotes the number of sampled observations used in the estimation procedure; since the samples come from the mixture $(p_Z + q_Z)/2$, the average is the

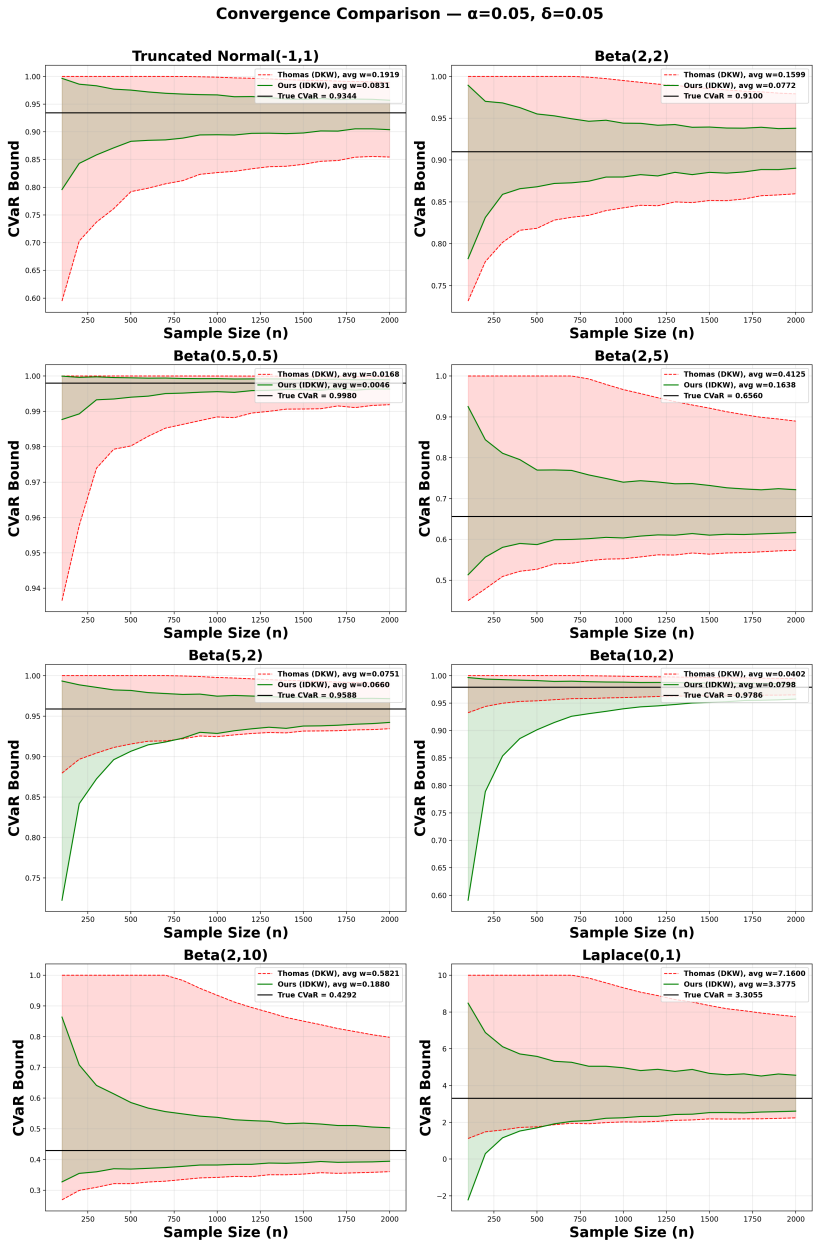


Fig. 7. Convergence comparison across eight distributions ($\alpha = 0.05$, $\delta = 0.05$). Red dashed: Thomas bounds (global DKW). Green solid: our bounds with exact local DKW (η_{lDKW}). Black line: true $\text{CVaR}_{0.05}$. Average interval widths are shown in each legend. Our bounds exhibit faster convergence and produce tighter intervals as the number of samples increases, with the largest gains on left-skewed distributions (Beta(2,5), Beta(2,10)) and heavy-tailed distributions (Laplace).

importance-sampling estimate of Δ_x^s in (44). N_z was set conservatively: Figure 18 in Appendix L shows that the estimate stops changing well before $N_z = 2000$ in all three environments. N^Δ was not tuned, since the estimate was already stable at this value. Given equation (56), the belief-dependent one-step distributional discrepancy $\hat{m}(\bar{b}_t, a)$ can be estimated accordingly. Since the state-transition model in (49) assigns exponentially small probability to states that are far from the current state, we employ a K-Nearest Neighbors (KNN) approach with $K = 10$ to select a subset of the samples $\{x_n^\Delta\}$ that lie in the vicinity of a given state x. The estimation is then performed over this localized subset, rather than the entire collection $\{x_n^\Delta\}$.

9.4 Empirical Static CVaR Bound Evaluation: Open-Loop Policies

To evaluate the bounds we define for each environment two action-sequences - one that dictates a safe path and the other that dictates a dangerous path. Figure 8 shows that bounds based on the simplified observation models distinguish between the two action sequences and would eliminate the dangerous path under action elimination. The practical relevance of the distributional discrepancy ϵ is governed by the ratio ϵ/α, which determines the fraction of the risk budget consumed by the model simplification and directly controls bound tightness (see Theorem A.3). At $\alpha = 0.2$, this ratio ranges from 3.5% (Push, safe path) to 16.6% (Laser Tag), meaning that up to a sixth of the risk budget is consumed by the observation model approximation. Despite this non-trivial cost, the bounds successfully discriminate between the two action sequences across all three environments. Moreover, the alpha sensitivity analysis (Figure 10, bottom row) shows that the bound intervals remain separated for risk levels well below $\alpha = 0.2$, even as the ratio ϵ/α increases—demonstrating that the bounds retain discriminative power in the more risk-averse regime where tail risk quantification matters most. The raw distributional discrepancy values are reported in Appendix L.

Comparing the bounds computation time using the original observation models to our bounds that use the simplified observation models, we gain approximately 21× acceleration in Light-Dark POMDP, 9× in Laser Tag POMDP, and 14× in Push POMDP. Figure 9 illustrates how the acceleration remains stable when the number of return samples and horizon changes.

Figure 10 presents the evolution of the CVaR bounds as a function of the planning horizon for the same safe and dangerous action sequences depicted in Figure 8, where each action sequence is decomposed into sub-sequences of increasing horizon length. In the Light-Dark environment, the bounds are separated from the earliest horizon, since the agent's start state is close to the dangerous area and the dangerous path immediately traverses the high-cost region. In Laser Tag and Push, the bounds of both paths are similar at early horizons, before the dangerous path enters the hazardous region, confirming that the bound separation is driven by the actual risk difference rather than by an artifact of the bounding mechanism. In Laser Tag, the separation is most pronounced: both paths have comparable bounds up to horizon 3, after which the dangerous path's bounds increase sharply as it enters the dangerous area, while the safe path remains at low values throughout. In Push, the two paths overlap closely up to horizon 5, and the bounds diverge around horizon 6 when the dangerous path approaches the hazardous region. The bottom row of Figure 10 further examines the sensitivity of the bounds to the risk level α at a fixed horizon. As α decreases, the CVaR focuses on worse-case outcomes, causing both the bound values and the bound intervals to increase; nevertheless, the bounds maintain a clear separation between the safe and dangerous paths for $\alpha \geq 0.1$ (Light-Dark) and $\alpha \geq 0.15$ (Push, Laser Tag).

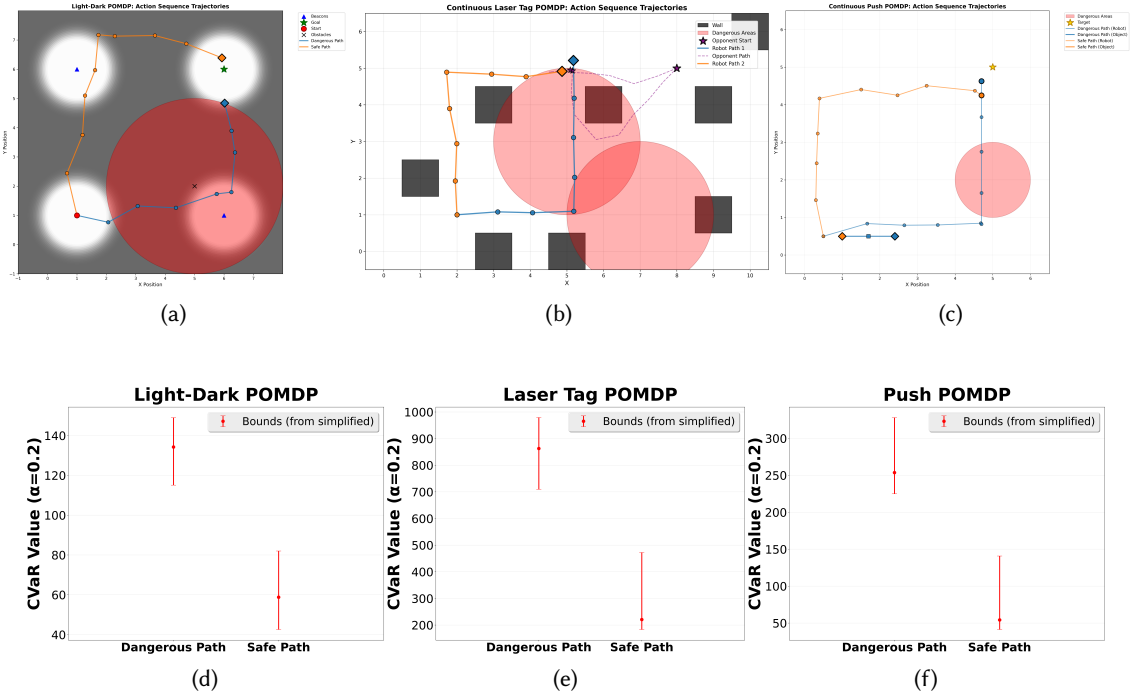


Fig. 8. Figures 8a, 8b, and 8c illustrate pairs of agent trajectories, consisting of a safe (orange) and a dangerous (blue) path. In Figure 8a, the agent moves from the lower-left to the upper-right goal location; the dangerous trajectory incurs higher cost. Figure 8b depicts trajectories in the Laser Tag environment, where the dangerous path traverses a high-cost region indicated in red. Figure 8c shows a similar pair of trajectories, with the dangerous path entering a hazardous region that carries a large per-step penalty. Across all three environments, the theoretical bounds on the values of the corresponding action sequences successfully discriminate between the safe and dangerous trajectories. Bounds are computed with probability of error $\delta = 0.05$.

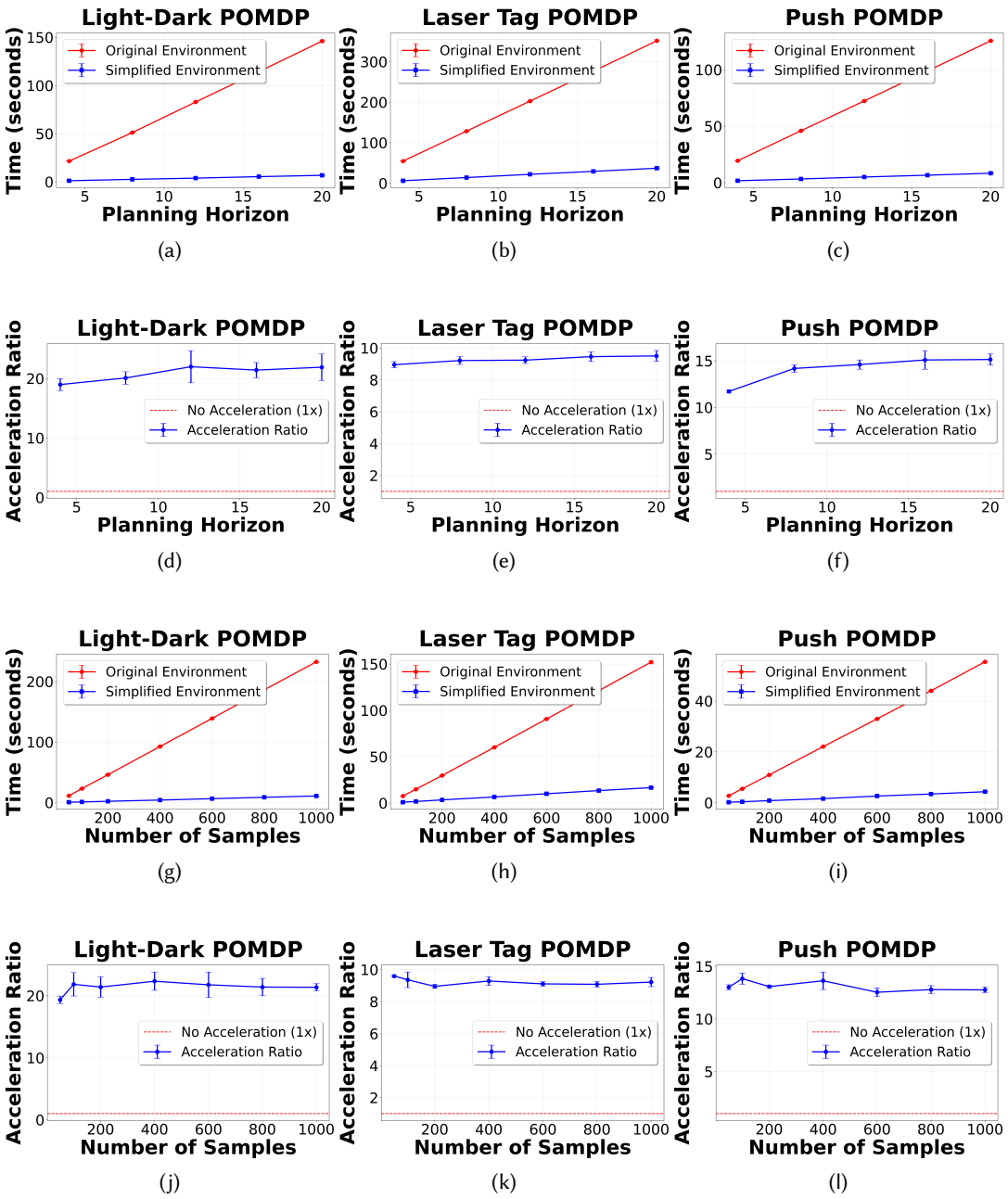


Fig. 9. Sensitivity analysis of the computational acceleration with respect to the planning horizon and number of return samples, for Light-Dark (left), Laser Tag (center), and Push (right), with $\alpha = 0.2$. Rows 1–2: computation time and acceleration ratio as a function of the planning horizon. Rows 3–4: computation time and acceleration ratio as a function of the number of return samples. Error bars indicate 95% confidence intervals.

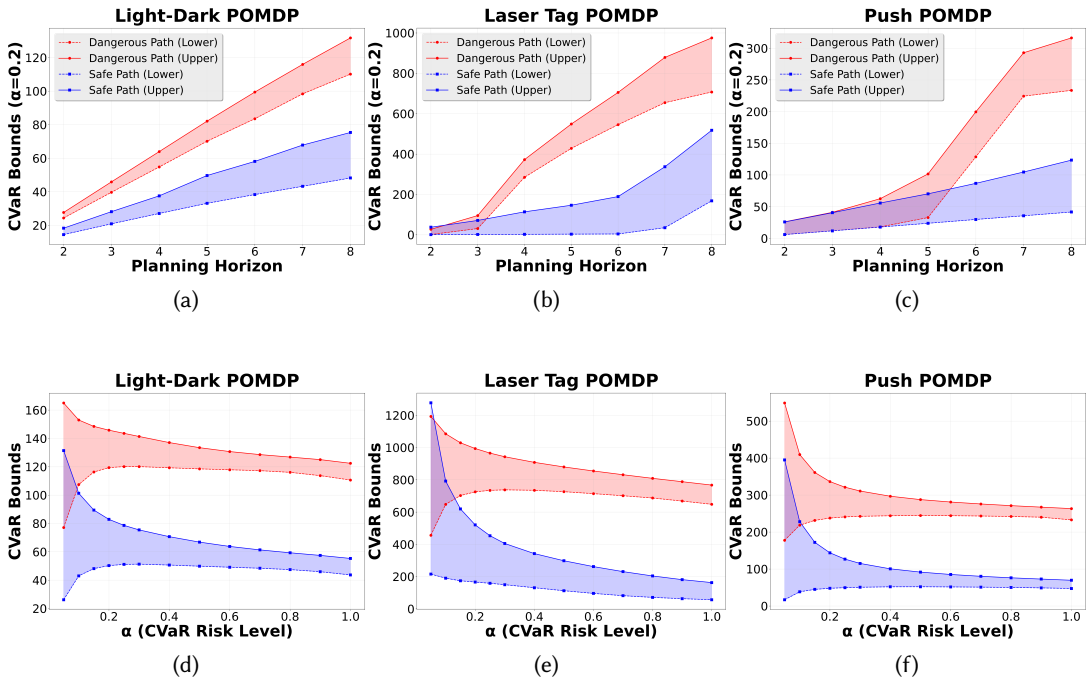


Fig. 10. Sensitivity of the CVaR bounds for the Light-Dark (left), Laser Tag (center), and Push (right) environments, using the complete safe (blue) and dangerous (red) action sequences depicted in Figure 8; solid and dashed lines indicate the upper and lower bounds, respectively. Row 1: bounds as a function of the planning horizon with $\alpha = 0.2$. Row 2: bounds as a function of the risk level α at a fixed horizon. As the horizon increases, the accumulated distributional discrepancy grows, causing the bounds to widen. As α decreases (more risk-averse), the bounds increase and the bound intervals widen, yet they remain well separated between the safe and dangerous paths for $\alpha \geq 0.1$ (Light-Dark) and $\alpha \geq 0.15$ (Push, Laser Tag). Bounds are computed with probability of error $\delta = 0.05$.

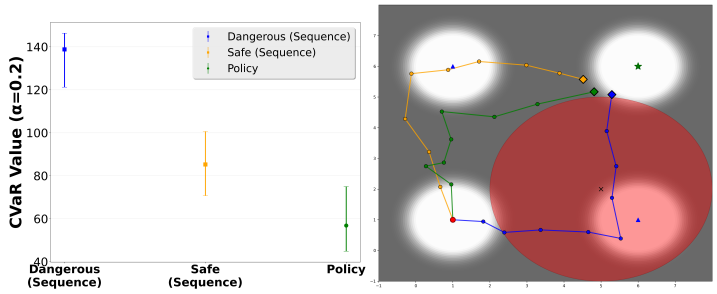
9.5 Empirical Static CVaR Bound Evaluation: Closed-Loop Policy

The preceding evaluation assessed bounds on fixed open-loop action sequences. We now demonstrate that the framework extends to closed-loop planners, using the Light-Dark POMDP as a test case. Specifically, we use the neural network trained by BetaZero (Moss, Corso, et al. 2024) as a deterministic policy: BetaZero learns offline approximations of the optimal policy and value function for POMDPs and combines them with online Monte Carlo tree search at test time. Here, we extract a deterministic closed-loop policy from the trained network by selecting at each belief state the action with minimum predicted cost. The BetaZero configuration used in our experiments is detailed in Appendix M.

We evaluate two objectives. First, we test whether the bounds computed under the simplified observation model successfully distinguish between policies of differing risk levels. We compare the BetaZero policy against the same safe and dangerous action sequences from the preceding subsection. As shown in Figures 11 and 12, the bounds clearly separate the dangerous sequence from both the safe sequence and the BetaZero policy, with non-overlapping bound intervals. The trajectory plot confirms this ordering: the dangerous sequence traverses the high-cost region, while the BetaZero policy navigates through the illuminated safe zones. This separation persists

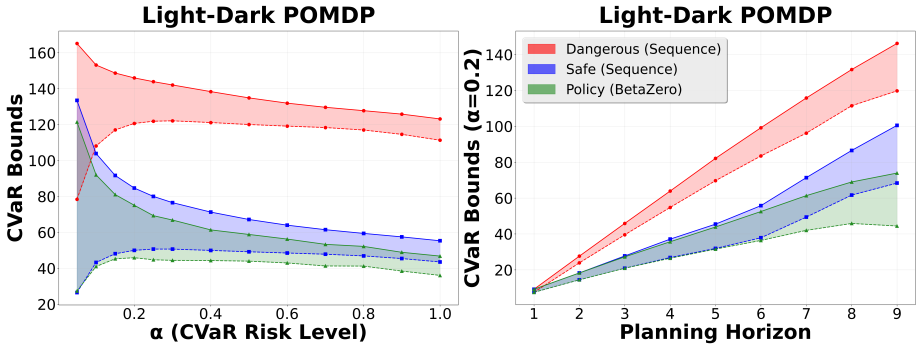
for $\alpha \geq 0.1$ and for all planning horizons from 2 on; at $\alpha = 0.05$ the three intervals overlap. The distributional discrepancy for the closed-loop BetaZero policy evaluation is reported in Appendix M.

Second, we show that evaluating the bounds under the simplified observation model is substantially faster than under the original model. Figure 13 shows a consistent 4.4–4.8× speedup, stable across planning horizons from 4 to 12 and return sample counts from 100 to 1000.



(a) (b)

Fig. 11. Evaluation of the BetaZero neural network policy against the safe and dangerous action sequences in the Light-Dark POMDP ($\alpha = 0.2$, $H = 9$), with bounds computed under the simplified observation model. Figure 11a: CVaR bounds for the dangerous sequence (blue), safe sequence (orange), and BetaZero policy (green). The bound intervals of the policy and dangerous sequence do not overlap, demonstrating that the bounds correctly identify the policy as less risky. Bounds are computed with probability of error $\delta = 0.05$. Figure 11b: agent trajectories; the BetaZero policy (green) and safe sequence (orange) navigate through the illuminated safe zones, while the dangerous sequence (blue) traverses the high-cost region.



(a) (b)

Fig. 12. Sensitivity of the CVaR bounds (computed under the simplified observation model) for the BetaZero policy, safe sequence, and dangerous sequence in the Light-Dark POMDP; solid and dashed lines denote upper and lower bounds, respectively. Figure 12a: bounds as a function of the risk level α at horizon $H = 9$. Figure 12b: bounds as a function of the planning horizon at $\alpha = 0.2$. The dangerous sequence remains clearly separated from the policy and safe sequence for $\alpha \geq 0.1$ and for horizons of two steps or more. Bounds are computed with probability of error $\delta = 0.05$.

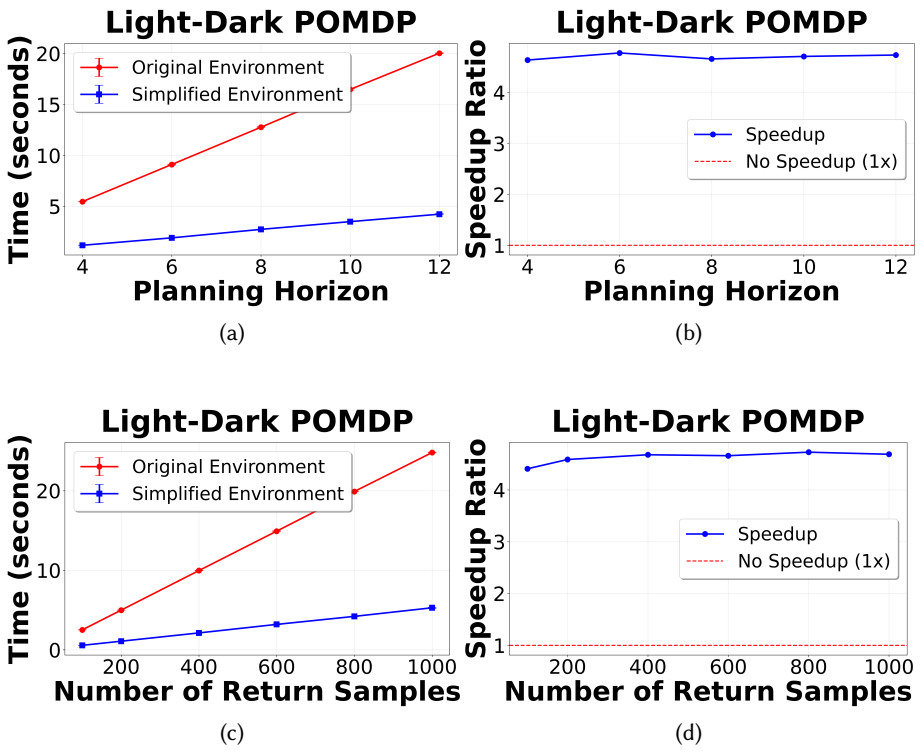


Fig. 13. Computational speedup from evaluating CVaR bounds under the simplified observation model relative to the original model for the BetaZero policy in the Light-Dark POMDP ($\alpha = 0.2$). Row 1: computation time and speedup ratio as a function of the planning horizon. Row 2: computation time and speedup ratio as a function of the number of return samples. The simplified model achieves a consistent 4.4–4.8× speedup across all conditions. Error bars indicate 95% confidence intervals.

9.6 Improved Bounds via Truncation and Local DKW

Figure 14 shows the alpha sensitivity across three POMDP environments. Our truncation-based bounds adapt to the sample range, producing substantially tighter intervals than the Thomas bounds—especially for safe policies, where the environment-level support bounds are most conservative.

Figure 15 presents the full convergence comparison across three POMDP environments, including all three concentration parameters. The truncation decomposition and local DKW refinements (Theorem 5.6 with η_{lDKW}) achieve 56–88% tighter intervals on these distributions at the largest sample size than the bounds of Thomas and Learned-Miller (2019), with 100% empirical coverage across all tested distributions and sample sizes.

9.7 POMDP Policy Evaluation with Improved Bounds

We evaluate the bounds on cumulative cost samples from the three POMDP environments, each with a safe and a dangerous action sequence. Cost samples are generated via belief-weighted Monte Carlo rollouts ($N = 5000$ pool); each rollout independently samples an initial state from the belief and simulates a fresh trajectory, ensuring the resulting cumulative costs are i.i.d. as required by the theorems. Support bounds are set to $[\underline{a}, \bar{b}] = [T \cdot c_{\min}, T \cdot c_{\max}]$

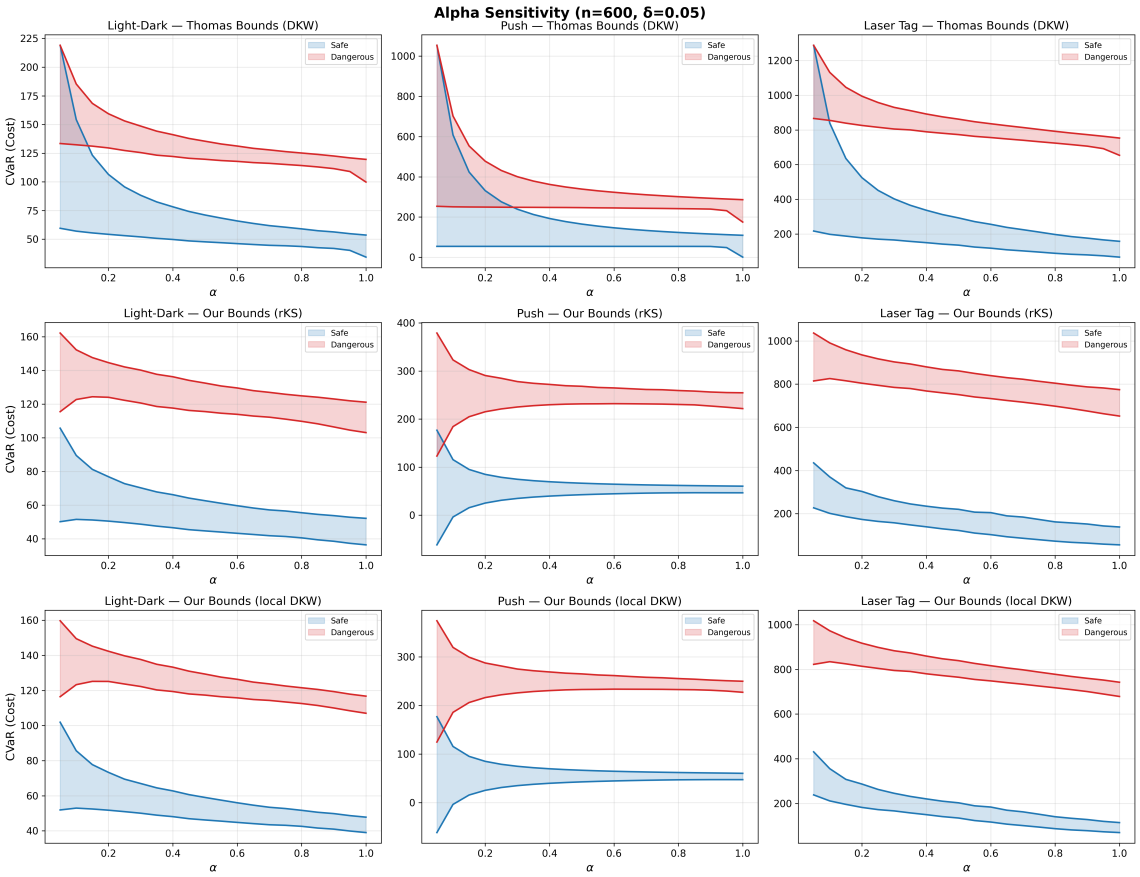


Fig. 14. Alpha sensitivity across three POMDP environments ($n = 600, \delta = 0.05$). **Top:** Thomas bounds use the full environment support $[T \cdot c_{\min}, T \cdot c_{\max}]$, producing wide intervals—especially for the safe policy (blue). **Middle:** Our bounds with asymptotic restricted KS (η_{rKS}) adapt to the sample range via truncation. **Bottom:** Our bounds with exact local DKW (η_{lDKW}) yield the tightest intervals, with the largest gains at small α.

using each environment's per-step cost range—the same bounds for both safe and dangerous sequences, as required when the effective support is unknown.

When the per-step cost range is used as a conservative support estimate, the Thomas bounds produce wide intervals that fail to separate safe and dangerous policies—especially for the safe policy, where the actual return distribution is far from the environment-level support bounds. Our truncation-based bounds confine the support dependence to the $O(1/n)$ tail correction, enabling policy discrimination at moderate sample sizes, with the largest gains at small α (Figure 14).

10 Conclusions

In this work we introduced a belief-simplification framework for risk-averse policy evaluation in POMDPs under a static CVaR objective, with guaranteed bounds on the resulting action-value function estimators, enabling accelerated online policy evaluation. Our approach establishes bounds on the true action-value function by

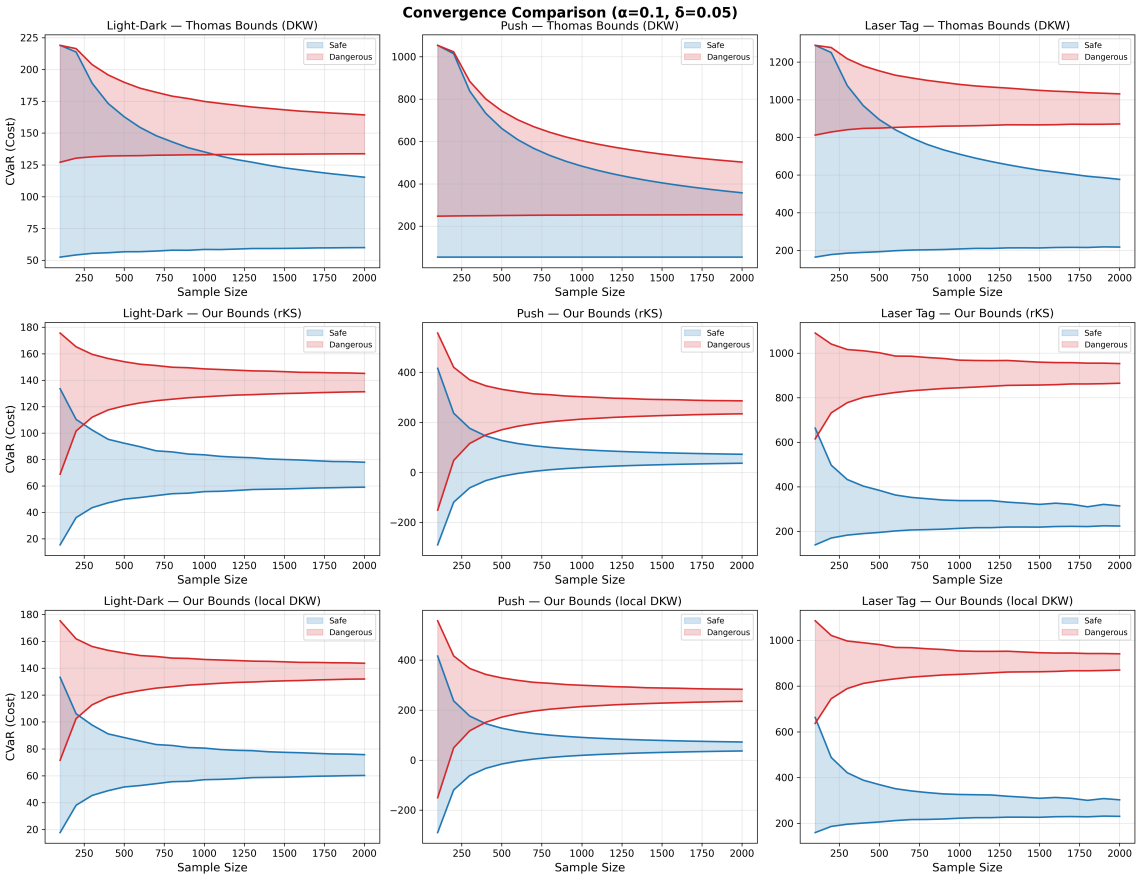


Fig. 15. Full convergence comparison ($\alpha = 0.1$, $\delta = 0.05$) including all three concentration parameters. **Top:** Thomas bounds (global DKW, rigorous). **Middle:** Our bounds with restricted KS (η_{rKS}, asymptotic). **Bottom:** Our bounds with exact local DKW (η_{lDKW}, rigorous).

leveraging an approximate action-value function computed under a simplified belief-transition model. We further derived specialized bounds for the case of a simplified observation model and showed how these bounds can be employed to accelerate online planning. To support these results, we developed the requisite mathematical foundations for bounding the CVaR of a random variable X through an auxiliary random variable Y, under assumptions relating their respective cumulative and density functions.

The mathematical foundations include novel CVaR bounds via auxiliary random variables, improved by a truncation decomposition that confines support dependence to $O(1/n)$ and local concentration refinements yielding 56–88% tighter intervals on the evaluated distributions. These theoretical results are independent of the POMDP-planning setting and provide a general framework for risk-sensitive analysis.

Our empirical evaluation, focused on observation model simplification, demonstrates the practical utility of the framework in both open-loop and closed-loop settings. In the open-loop case, the bounds correctly distinguish action sequences of differing risk levels across multiple environments and risk parameters, achieving 9–21× computational speedups. In the closed-loop case, using a BetaZero neural-network policy, the bound interval of

the dangerous sequence is disjoint from those of the safe sequence and of the policy for $\alpha \geq 0.1$, at a 4.4–4.8× speedup over evaluation under the original model. The theoretical framework accommodates general belief-transition model simplifications, including state-transition models, whose empirical investigation we leave for future work.

This work was supported by the Israel Ministry of Innovation, Science and Technology.

References

C. Acerbi and D. Tasche. 2002. "On the coherence of expected shortfall." *Journal of banking & finance*, 26, 7, 1487–1503.

M. Ahmadi, A. Dixit, J. W. Burdick, and A. D. Ames. 2021. "Risk-averse stochastic shortest path planning." In: *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 5199–5204.

M. Ahmadi, M. Ono, M. D. Ingham, R. M. Murray, and A. D. Ames. 2020. "Risk-averse planning under uncertainty." In: *2020 American Control Conference (ACC)*. IEEE, 3305–3312.

M. Ahmadi, U. Rosolia, M. D. Ingham, R. M. Murray, and A. D. Ames. 2021. "Constrained risk-averse Markov decision processes." In: *Proceedings of the AAAI Conference on Artificial Intelligence* 13. Vol. 35, 11718–11725.

M. Ahmadi, U. Rosolia, M. D. Ingham, R. M. Murray, and A. D. Ames. 2023. "Risk-averse decision making under uncertainty." *IEEE Transactions on Automatic Control*, 69, 1, 55–68.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. 1999. "Coherent measures of risk." *Mathematical finance*, 9, 3, 203–228.

M. Barenboim and V. Indelman. July 2022. "Adaptive Information Belief Space Planning." In: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*. Ed. by L. D. Raedt. Main Track. International Joint Conferences on Artificial Intelligence Organization, (July 2022), 4588–4596. doi:10.24963/ijcai.2022/637.

M. Barenboim and V. Indelman. 2023. "Online POMDP Planning with Anytime Deterministic Guarantees." In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine. Vol. 36. Curran Associates, Inc., 79886–79902. https://proceedings.neurips.cc/paper_files/paper/2023/file/fc6bd0eef19459655d5b097af783661d-Paper-Conference.pdf.

M. Barenboim and V. Indelman. 2026. "Online POMDP planning with anytime deterministic optimality guarantees." *Artificial Intelligence*, 350, 104442. doi:https://doi.org/10.1016/j.artint.2025.104442.

N. Bäuerle and A. Glauner. 2020. "Distributionally robust Markov decision processes and their connection to risk measures." *Mathematics of Operations Research*, 47, 3, 1757–1780.

N. Bäuerle and J. Ott. 2011. "Markov decision processes with average-value-at-risk criteria." *Mathematical Methods of Operations Research*, 74, 3, 361–379.

B. Behzadian, R. H. Russel, C. P. Ho, and M. Petrik. 2021. "Optimizing percentile criterion using robust MDPs." In: *International Conference on Artificial Intelligence and Statistics (AISTATS)*.

T. Brázdil, K. Chatterjee, V. Forejt, and A. Kučera. 2017. "Trading performance for stability in Markov decision processes." *Journal of Computer and System Sciences*, 84, 144–170. doi:10.1016/j.jcss.2016.09.009.

T. Brihaye, K. Chatterjee, S. Mohr, and M. Weininger. 2025. "Risk-aware Markov decision processes using cumulative prospect theory." In: *ACM/IEEE Symp. on Logic in Computer Science (LICS)*, 458–471. doi:10.1109/LICS65433.2025.00041.

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

C. E. Budde, A. Hartmanns, T. Meggendorfer, M. Weininger, and P. Wienhöft. 2025. "Statistical model checking beyond means: quantiles, CVaR, and the DKW inequality." In: *Intl. Conf. on Quantitative Evaluation of Systems (QEST+FORMATS)*.

Y. Chow, M. Ghavamzadeh, L. Janson, and M. Pavone. 2018. "Risk-constrained reinforcement learning with percentile risk criteria." *Journal of Machine Learning Research*, 18, 1–51.

M. Cubuktepe, N. Jansen, S. Junges, A. Marandi, M. Suilen, and U. Topcu. 2021. "Robust finite-state controllers for uncertain POMDPs." In: *Proceedings of the AAAI Conference on Artificial Intelligence* 13. Vol. 35, 11792–11800.

W. Dabney, M. Rowland, M. Bellemare, and R. Munos. 2018. "Distributional reinforcement learning with quantile regression." In: *Proceedings of the AAAI conference on artificial intelligence* 1. Vol. 32.

E. Delage and S. Mannor. 2010. "Percentile optimization for Markov decision processes with parameter uncertainty." *Operations Research*, 58, 1, 203–213.

A. Dixit, M. Ahmadi, and J. W. Burdick. 2023. "Risk-averse receding horizon motion planning for obstacle avoidance using coherent risk measures." *Artificial Intelligence*, 325, 104018.

M. D. Donsker. 1952. "Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems." *Annals of Mathematical Statistics*, 23, 2, 277–281.

A. Dvoretzky, J. Kiefer, and J. Wolfowitz. 1956. "Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator." *The Annals of Mathematical Statistics*, 27, 3, 642–669. doi:10.1214/aoms/1177728174.

H. Föllmer and A. Schied. 2002. "Convex measures of risk and trading constraints." *Finance and Stochastics*, 6, 4, 429–447. doi:10.1007/s007800200072.

A. Ghosh, L. Prashanth, and K. Jagannathan. 2024. "Concentration Bounds for Optimized Certainty Equivalent Risk Estimation." *arXiv preprint arXiv:2405.20933*.

P. Hou, W. Yeoh, and P. Varakantham. 2016. "Solving risk-sensitive POMDPs with and without cost observations." In: *Proceedings of the AAAI Conference on Artificial Intelligence* 1. Vol. 30.

D. R. Jiang and W. B. Powell. 2018. "Risk-averse approximate dynamic programming with quantile-based risk measures." *Mathematics of Operations Research*, 43, 2, 554–579. doi:10.1287/moor.2017.0872.

S. Koenig and R. G. Simmons. 1994. "Risk-sensitive planning with probabilistic decision graphs." In: *Principles of Knowledge Representation and Reasoning*. Elsevier, 363–373.

J. Křetínský and T. Meggendorfer. 2018. "Conditional value-at-risk for reachability and mean payoff in Markov decision processes." In: *ACM/IEEE Symp. on Logic in Computer Science (LICS)*, 609–618. doi:10.1145/3209108.3209176.

I. Lev-Yehudi, M. Barenboim, and V. Indelman. Mar. 2024. "Simplifying Complex Observation Models in Continuous POMDP Planning with Probabilistic Guarantees and Practice." In: 18. Vol. 38. (Mar. 2024), 20176–20184. doi:10.1609/aaai.v38i18.29997.

M. H. Lim, T. J. Becker, M. J. Kochenderfer, C. J. Tomlin, and Z. N. Sunberg. 2023. "Optimality guarantees for particle belief approximation of POMDPs." *Journal of Artificial Intelligence Research*, 77, 1591–1636.

E. Lobo, C. Cousins, M. Petrik, and Y. Zick. 2023. "Percentile criterion optimization in offline reinforcement learning." *Advances in Neural Information Processing Systems*, 36, 9322–9352.

O.-A. Maillard. 2021. "Local Dvoretzky–Kiefer–Wolfowitz Confidence Bands." *Mathematical Methods of Statistics*, 30, 1, 16–46.

A. Majumdar and M. Pavone. 2020. "How should a robot assess risk? towards an axiomatic theory of risk in robotics." In: *Robotics Research: The 18th International Symposium ISRR*. Springer, 75–84.

S. Mannor and J. N. Tsitsiklis. 2011. *Mean-variance optimization in Markov decision processes*. (2011). arXiv: 1104.5601.

J. Marecki and P. Varakantham. 2010. "Risk-sensitive planning in partially observable environments." In: *AAMAS*, 1357–1368.

P. Massart. 1990. "The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality." *The annals of Probability*, 1269–1283.

T. Meggendorfer. 2022. "Risk-aware stochastic shortest path." In: *AAAI Conf. on Artificial Intelligence* 9. Vol. 36, 9858–9867. doi:10.1609/aaai.v36i9.21222.

R. J. Moss, A. Corso, J. Caers, and M. J. Kochenderfer. 2024. "BetaZero: Belief-State Planning for Long-Horizon POMDPs using Learned Approximations." In: *Reinforcement Learning Conference (RLC)*.

R. J. Moss, A. Jamgochian, J. Fischer, A. Corso, and M. J. Kochenderfer. 2024. "ConstrainedZero: Chance-Constrained POMDP Planning Using Learned Probabilistic Failure Surrogates and Adaptive Safety Constraints." In: *International Joint Conference on Artificial Intelligence (IJCAI)*.

M. Ono, M. Pavone, Y. Kuwata, and J. Balaram. 2015. "Chance-constrained dynamic programming with application to risk-aware robotic space exploration." *Autonomous Robots*, 39, 555–571.

T. Osogami. 2015. "Robust partially observable Markov decision process." In: *International Conference on Machine Learning*. PMLR, 106–115.

C. H. Papadimitriou and J. N. Tsitsiklis. Aug. 1987. "The Complexity of Markov Decision Processes." *Math. Oper. Res.*, 12, 3, (Aug. 1987), 441–450.

Y. Pariente and V. Indelman. 2026. *POMDPPlanners: Open-Source Package for POMDP Planning*. (2026). https://arxiv.org/abs/2602.20810 arXiv: 2602.20810 (cs.AI).

G. C. Pflug. 2000. "Some remarks on the value-at-risk and the conditional value-at-risk." *Probabilistic constrained optimization: Methodology and applications*, 272–281.

M. Randour, J.-F. Raskin, and O. Sankur. 2017. "Percentile queries in multi-dimensional Markov decision processes." *Formal Methods in System Design*, 50, 2-3, 207–248. doi:10.1007/s10703-016-0262-7.

R. T. Rockafellar, S. Uryasev, et al.. 2000. "Optimization of conditional value-at-risk." *Journal of risk*, 2, 21–42.

P. Rodrigues Quemel e Assis Santana, S. Thiébaux, and B. Williams. Mar. 2016. "RAO*: An Algorithm for Chance-Constrained POMDP's." *Proceedings of the AAAI Conference on Artificial Intelligence*, 30, 1, (Mar. 2016). doi:10.1609/aaai.v30i1.10423.

A. Ruszczyński. 2010. "Risk-averse dynamic programming for Markov decision processes." *Mathematical Programming*, 125, 2, 235–261.

G. R. Shorack and J. A. Wellner. 1982. "Limit theorems and inequalities for the uniform empirical process indexed by intervals." *Annals of Probability*, 10, 3, 639–652. doi:10.1214/aop/1176993773.

D. Silver and J. Veness. 2010. "Monte-Carlo planning in large POMDPs." *Advances in neural information processing systems*, 23.

Z. Sunberg and M. Kochenderfer. 2018. "Online algorithms for POMDPs with continuous state, action, and observation spaces." In: *Proceedings of the International Conference on Automated Planning and Scheduling*. Vol. 28, 259–263.

P. Thomas and E. Learned-Miller. June 2019. "Concentration Inequalities for Conditional Value at Risk." In: *Proceedings of the 36th International Conference on Machine Learning* (Proceedings of Machine Learning Research). Ed. by K. Chaudhuri and R. Salakhutdinov. Vol. 97. PMLR, (June 2019), 6225–6233. https://proceedings.mlr.press/v97/thomas19a.html.

M. Ummels and C. Baier. 2013. "Computing quantiles in Markov reward models." In: *Foundations of Software Science and Computation Structures (FoSSaCS)* (LNCS). Vol. 7794, 353–368. doi:10.1007/978-3-642-37075-5_23.
H. Xu and S. Mannor. 2010. "Distributionally robust Markov decision processes." *Advances in Neural Information Processing Systems*, 23.
Z. Zhang, A. Ghosh, and T. Lan. 2025. "Tail-Risk-Safe Monte Carlo Tree Search under PAC-Level Guarantees." *arXiv preprint arXiv:2508.05441*.
A. Zhitnikov and V. Indelman. 2022. "Simplified Risk-aware Decision Making with Belief-dependent Rewards in Partially Observable Domains." *Artificial Intelligence*, 312, 103775. doi:https://doi.org/10.1016/j.artint.2022.103775.
A. Zhitnikov, O. Sztyglic, and V. Indelman. 2025. "No compromise in solution quality: Speeding up belief-dependent continuous partially observable Markov decision processes via adaptive multilevel simplification." *The International Journal of Robotics Research*, 44, 2, 157–195. eprint: https://doi.org/10.1177/02783649241261398. doi:10.1177/02783649241261398.

A Intermediate CVaR Concentration Bounds

The following intermediate results are obtained by combining the deterministic bounds of Theorem 5.1 with the DKW inequality. They require exact knowledge of the support bounds, which enter at $O(1/\sqrt{n})$. The support-relaxed bounds in Theorem 5.6 strictly improve upon these results by confining the support dependence to an $O(1/n)$ tail correction.

Theorem A.1 (Single-Variable CVaR Concentration). *Let X be a random variable, $\alpha \in (0,1], \delta \in (0, 0.5), \epsilon \triangleq \sqrt{\ln(1/\delta)/(2n)}$. Let $X_1, \dots, X_n \overset{iid}{\sim} F_X$ be random variables that define the ECDF $\hat{F}_X$.*

(1) ***Upper Bound:*** *If $b \geq \operatorname{ess\,sup} X$, then*
   (a) *If $\alpha > \epsilon$ then $\mathbb{P}[\mathrm{CVaR}_\alpha(X) \leq (1 - \frac{\epsilon}{\alpha})C^{\hat{F}_X}_{\alpha-\epsilon} + \frac{\epsilon}{\alpha} b] > 1 - \delta$.*
   (b) *If $\alpha \leq \epsilon$ then $\mathrm{CVaR}_\alpha(X) \leq b$*

(2) ***Lower Bound:*** *If $a \leq \operatorname{ess\,inf} X$, then*
   (a) *If $\alpha + \epsilon < 1$, then $\mathbb{P}[\mathrm{CVaR}_\alpha(X) \geq (1 + \frac{\epsilon}{\alpha})C^{\hat{F}_X}_{\alpha+\epsilon} - \frac{\epsilon}{\alpha}C^{\hat{F}_X}_{\epsilon}] > 1 - \delta$.*
   (b) *If $\alpha + \epsilon \geq 1$, then $\mathbb{P}[\mathrm{CVaR}_\alpha(X) \geq \frac{1}{\alpha}[(\alpha + \epsilon - 1)a + E_{\hat{F}_X}[X] - \epsilon C^{\hat{F}_X}_{\epsilon}]] > 1 - \delta$.*

Proof. From DKW inequality the following inequalities can be derived (Thomas and Learned-Miller 2019)

$$\mathbb{P}\left[\sup_{x\in\mathbb{R}} \left(\hat{F}(x) - F(x)\right) \leq \sqrt{\frac{\ln(1/\delta)}{2n}}\right] \geq 1 - \delta, \quad \mathbb{P}\left[\sup_{x\in\mathbb{R}} \left(F(x) - \hat{F}(x)\right) \leq \sqrt{\frac{\ln(1/\delta)}{2n}}\right] \geq 1 - \delta. \tag{57}$$

Let $\epsilon \triangleq \sqrt{\frac{\ln(1/\delta)}{2n}}$. By using Theorem 5.1 we get the following.

If $\alpha > \epsilon$ then

$$\begin{aligned}\mathbb{P}[\mathrm{CVaR}_\alpha(X) \leq \frac{\alpha-\epsilon}{\alpha}C^{\hat{F}_X}_{\alpha-\epsilon} + \frac{\epsilon}{\alpha}b] = &\underbrace{\mathbb{P}[C_\alpha(X) \leq \frac{\alpha-\epsilon}{\alpha}C^{\hat{F}_X}_{\alpha-\epsilon} + \frac{\epsilon}{\alpha}b \Big| \sup_{z\in\mathbb{R}}(\hat{F}_X - F_X) \leq \epsilon]}_{=1}\underbrace{\mathbb{P}[\sup_{z\in\mathbb{R}}(\hat{F}_X - F_X) \leq \epsilon]}_{>1-\delta} \\ &+ \underbrace{\mathbb{P}[C_\alpha(X) \leq \frac{\alpha-\epsilon}{\alpha}C^{\hat{F}_X}_{\alpha-\epsilon} + \frac{\epsilon}{\alpha}b \Big| \sup_{z\in\mathbb{R}}(\hat{F}_X - F_X) > \epsilon]}_{\geq 0}\underbrace{\mathbb{P}[\sup_{z\in\mathbb{R}}(\hat{F}_X - F_X) > \epsilon]}_{\geq 0} \\ &> 1 - \delta.\end{aligned} \tag{58}$$

Observe that, conditional on the event

$$\sup_{z\in\mathbb{R}} \left(\hat{F}_X(z) - F_X(z)\right) \leq \epsilon,$$

the bound in Theorem 5.1 holds deterministically. Consequently, the probability that

$$C_\alpha(X) \leq \frac{\alpha-\epsilon}{\alpha}C^{\hat{F}_X}_{\alpha-\epsilon} + \frac{\epsilon}{\alpha}b$$

holds, given

$$\sup_{z\in\mathbb{R}}\left(\hat{F}_X(z)-F_X(z)\right)\leq\epsilon,$$

is equal to one. The same observation is necessary through the rest of the proof in a similar manner. If $\alpha+\epsilon<1$, then

$$\begin{aligned}\mathbb{P}[C_\alpha(X)\geq(1+\frac{\epsilon}{\alpha})C^{\hat{F}_X}_{\alpha+\epsilon}-\frac{\epsilon}{\alpha}C^{\hat{F}_X}_{\epsilon}] &= \underbrace{\mathbb{P}[C_\alpha(X)\geq(1+\frac{\epsilon}{\alpha})C^{\hat{F}_X}_{\alpha+\epsilon}-\frac{\epsilon}{\alpha}C^{\hat{F}_X}_{\epsilon}\Big|\sup_{z\in\mathbb{R}}(F_X-\hat{F}_X)\leq\epsilon]}_{=1}\underbrace{\mathbb{P}[\sup_{z\in\mathbb{R}}(F_X-\hat{F}_X)\leq\epsilon]}_{>1-\delta}\\ &\quad+\underbrace{\mathbb{P}[C_\alpha(X)\geq(1+\frac{\epsilon}{\alpha})C^{\hat{F}_X}_{\alpha+\epsilon}-\frac{\epsilon}{\alpha}C^{\hat{F}_X}_{\epsilon}\Big|\sup_{z\in\mathbb{R}}(F_X-\hat{F}_X)>\epsilon]}_{\geq0}\underbrace{\mathbb{P}[\sup_{z\in\mathbb{R}}(F_X-\hat{F}_X)>\epsilon]}_{\geq0}\\ &>1-\delta\end{aligned}\tag{59}$$

If $\alpha+\epsilon\geq1$, then

$$\begin{aligned}&\mathbb{P}[C_\alpha(X)\geq\frac{1}{\alpha}[(\alpha+\epsilon-1)a+E_{\hat{F}_X}[X]-\epsilon C^{\hat{F}_X}_{\epsilon}]]\\ &=\underbrace{\mathbb{P}[C_\alpha(X)\geq\frac{1}{\alpha}[(\alpha+\epsilon-1)a+E_{\hat{F}_X}[X]-\epsilon C^{\hat{F}_X}_{\epsilon}]\Big|\sup_{z\in\mathbb{R}}(F_X-\hat{F}_X)\leq\epsilon]}_{=1}\underbrace{\mathbb{P}[\sup_{z\in\mathbb{R}}(F_X-\hat{F}_X)\leq\epsilon]}_{>1-\delta}\\ &+\underbrace{\mathbb{P}[C_\alpha(X)\geq\frac{1}{\alpha}[(\alpha+\epsilon-1)a+E_{\hat{F}_X}[X]-\epsilon C^{\hat{F}_X}_{\epsilon}]\Big|\sup_{z\in\mathbb{R}}(F_X-\hat{F}_X)>\epsilon]}_{\geq0}\underbrace{\mathbb{P}[\sup_{z\in\mathbb{R}}(F_X-\hat{F}_X)>\epsilon]}_{\geq0}\\ &>1-\delta\end{aligned}\tag{60}$$

□

Theorem A.1 is obtained by applying Theorem 5.1 with Y instantiated as the ECDF of X, using the DKW inequality to control ϵ. The bounds coincide with those of Thomas and Learned-Miller (2019) (Theorems 3.2 and 3.3), which express the same result through reweighted order statistics; our formulation in terms of CVaR is more interpretable.

Corollary A.2. *Let X be a random variable, $\alpha\in(0,1], \delta\in(0,1), \epsilon\triangleq\sqrt{\ln(1/\delta)/(2n)}, a\in\mathbb{R}, b\in\mathbb{R}$. Let $X_1,\ldots,X_n\overset{iid}{\sim}F_X$ be random variables that define the ECDF $\hat{F}_X$. Denote by $U(n)$ and $L(n)$ the upper and lower bounds respectively from Theorem A.1, where n is the number of samples, then*

(1) *If $b\geq\operatorname{ess\,sup}X$, then $\lim_{n\to\infty}U(n)\overset{a.s.}{=}\mathrm{CVaR}_\alpha(X)$.*
(2) *If $a\leq\operatorname{ess\,inf}X$, then $\lim_{n\to\infty}L(n)\overset{a.s.}{=}\mathrm{CVaR}_\alpha(X)$.*

where a.s. denotes almost sure convergence.

Proof.

$$\lim_{n\to\infty}U(n)\overset{a.s}{=}\lim_{n\to\infty}(1-\frac{\epsilon}{\alpha})C^{\hat{F}_X}_{\alpha-\epsilon}+\frac{\epsilon}{\alpha}b\overset{a.s}{=}\lim_{n\to\infty}C^{\hat{F}_X}_{\alpha-\epsilon}=\lim_{n\to\infty}C^{\hat{F}_X}_{\alpha}=C^{F_X}_{\alpha}\tag{61}$$

where the first and second equalities follow from the fact that $\epsilon\to0$ as $n\to\infty$; the third equality holds by the continuity of CVaR with respect to α; and the fourth equality holds by the almost sure convergence of the

empirical CVaR of i.i.d. samples to the true CVaR. For the same reason,

$$\lim_{n\to\infty} L(n) \overset{a.s}{=} \lim_{n\to\infty} (1+\frac{\epsilon}{\alpha})C^{\hat{F}_X}_{\alpha+\epsilon} - \frac{\epsilon}{\alpha}C^{\hat{F}_X}_{\epsilon} \overset{a.s}{=} \lim_{n\to\infty} C^{\hat{F}_X}_{\alpha+\epsilon} \overset{a.s}{=} \lim_{n\to\infty} C^{\hat{F}_X}_{\alpha} \overset{a.s}{=} \lim_{n\to\infty} C^{F_X}_{\alpha}. \tag{62}$$

□

Theorem A.3 (Auxiliary-Variable CVaR Concentration). *Let X and Y be random variables, $\epsilon \in [0,1], \delta \in (0, 0.5)$. Let $Y_1, \ldots, Y_n$ be independent and identically distributed samples from F_Y, and denote by $\hat{F}_Y$ the associated empirical cumulative distribution function. Let $\eta \geq 0$ satisfy $\mathbb{P}[\sup_z |\hat{F}_Y(z) - F_Y(z)| \leq \eta] > 1-\delta$, and define $\epsilon' = \min(\epsilon+\eta, 1)$.*

(1) ***Upper Bound:*** *If $\forall z \in \mathbb{R}, F_Y(z) - F_X(z) \leq \epsilon$ and $b_X \geq \operatorname{ess\,sup} X$ and $b_Y \geq \operatorname{ess\,sup} Y$, then*
   (a) *If $\alpha > \epsilon'$ then $\mathbb{P}\Big[\ \mathrm{CVaR}_\alpha(X) \leq \frac{\epsilon'}{\alpha}\max(b_X, b_Y) + (1-\frac{\epsilon'}{\alpha})\,\mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\epsilon'}\ \Big] > 1-\delta$.*
   (b) *If $\alpha \leq \epsilon'$, then $\mathrm{CVaR}_\alpha(X) \leq \max(b_X, b_Y)$*

(2) ***Lower Bound:*** *If $\forall z \in \mathbb{R}, F_X(z) - F_Y(z) \leq \epsilon$ and $a_X \leq \operatorname{ess\,inf} X$ and $a_Y \leq \operatorname{ess\,inf} Y$, then*
   (a) *If $\alpha + \epsilon' \leq 1$, then*

$$\mathbb{P}\Big[\ \mathrm{CVaR}_\alpha(X) \geq (1+\frac{\epsilon'}{\alpha})\,\mathrm{CVaR}^{\hat{F}_Y}_{\alpha+\epsilon'} - \frac{\epsilon'}{\alpha}\,\mathrm{CVaR}^{\hat{F}_Y}_{\epsilon'}\ \Big] > 1-\delta. \tag{63}$$

   (b) *If $\alpha + \epsilon' > 1$, then*

$$\mathbb{P}\Big[\ \mathrm{CVaR}_\alpha(X) \geq \frac{1}{\alpha}\big(\mathbb{E}_{\hat{F}_Y}[Y] - \epsilon'\,\mathrm{CVaR}^{\hat{F}_Y}_{\epsilon'} + (\alpha+\epsilon'-1)a_{min}\big)\Big] > 1-\delta. \tag{64}$$

Proof. We begin by establishing that ϵ' bounds the distributional discrepancy between $\hat{F}_Y$ and F_X, under the assumption that $\sup_x \big(\hat{F}_Y(x) - F_Y(x)\big) \leq \eta$. Let $x \in \mathbb{R}$,

$$\begin{aligned} \hat{F}_Y(x) - F_X(x) &= \hat{F}_Y(x) - F_Y(x) + F_Y(x) - F_X(x) \\ &\leq |\hat{F}_Y(x) - F_Y(x)| + |F_Y(x) - F_X(x)| \leq \eta + \epsilon = \epsilon'. \end{aligned} \tag{65}$$

Assume that $\alpha > \epsilon'$. Note that, conditional on the event $\sup_{z\in\mathbb{R}} \big(\hat{F}_Y(z) - F_X(z)\big) \leq \epsilon'$, the upper bound in Theorem 5.1 holds deterministically. Consequently, the probability that

$$\mathrm{CVaR}_\alpha(X) \leq \frac{\epsilon'}{\alpha}\max(b_X, b_Y) + (1-\frac{\epsilon'}{\alpha})\,\mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\epsilon'} \tag{66}$$

holds is equal to one. From the law of total probability,

$$\begin{aligned} &\mathbb{P}\Big[\ \mathrm{CVaR}_\alpha(X) \leq \frac{\epsilon'}{\alpha}\max(b_X, b_Y) + (1-\frac{\epsilon'}{\alpha})\,\mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\epsilon'}\ \Big] \\ &= \mathbb{P}\Big[\ \mathrm{CVaR}_\alpha(X) \leq \frac{\epsilon'}{\alpha}\max(b_X, b_Y) + (1-\frac{\epsilon'}{\alpha})\,\mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\epsilon'}\ \Big|\ \sup_{z\in\mathbb{R}} \big(\hat{F}_Y(z) - F_X(z)\big) \leq \epsilon'\Big]\mathbb{P}[\sup_{z\in\mathbb{R}} \big(\hat{F}_Y(z) - F_X(z)\big) \leq \epsilon'] \\ &+ \mathbb{P}\Big[\ \mathrm{CVaR}_\alpha(X) \leq \frac{\epsilon'}{\alpha}\max(b_X, b_Y) + (1-\frac{\epsilon'}{\alpha})\,\mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\epsilon'}\ \Big|\ \sup_{z\in\mathbb{R}} \big(\hat{F}_Y(z) - F_X(z)\big) > \epsilon'\Big]\mathbb{P}[\sup_{z\in\mathbb{R}} \big(\hat{F}_Y(z) - F_X(z)\big) > \epsilon'] \\ &= \mathbb{P}[\sup_{z\in\mathbb{R}} \big(\hat{F}_Y(z) - F_X(z)\big) \leq \epsilon'] \\ &+ \mathbb{P}\Big[\ \mathrm{CVaR}_\alpha(X) \leq \frac{\epsilon'}{\alpha}\max(b_X, b_Y) + (1-\frac{\epsilon'}{\alpha})\,\mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\epsilon'}\ \Big|\ \sup_{z\in\mathbb{R}} \big(\hat{F}_Y(z) - F_X(z)\big) > \epsilon'\Big]\mathbb{P}[\sup_{z\in\mathbb{R}} \big(\hat{F}_Y(z) - F_X(z)\big) > \epsilon'] \\ &\geq \mathbb{P}[\sup_{z\in\mathbb{R}} \big(\hat{F}_Y(z) - F_X(z)\big) \leq \epsilon'] \end{aligned} \tag{67}$$

From DKW (Dvoretzky et al. 1956) inequality the following inequalities can be derived (Thomas and Learned-Miller 2019)

$$\mathbb{P}\left[\sup_{x\in\mathbb{R}}\left(\hat{F}(x)-F(x)\right)\leq\sqrt{\frac{\ln(1/\delta)}{2n}}\right]\geq 1-\delta,\qquad \mathbb{P}\left[\sup_{x\in\mathbb{R}}\left(\hat{F}(x)-F(x)\right)\geq\sqrt{\frac{\ln(1/\delta)}{2n}}\right]\geq 1-\delta. \tag{68}$$

$$\begin{aligned}\mathbb{P}[\sup_{z\in\mathbb{R}}\left(\hat{F}_Y(z)-F_X(z)\right)\leq\epsilon'] &\geq \mathbb{P}[\sup_{z\in\mathbb{R}}\left(\hat{F}_Y(z)-F_Y(z)\right)+\sup_{z\in\mathbb{R}}\left(F_Y(z)-F_X(z)\right)\leq\epsilon']\\ &\geq \mathbb{P}[\sup_{z\in\mathbb{R}}\left(\hat{F}_Y(z)-F_Y(z)\right)+\epsilon\leq\epsilon+\eta]\\ &= \mathbb{P}[\sup_{z\in\mathbb{R}}\left(\hat{F}_Y(z)-F_Y(z)\right)\leq\eta]>1-\delta.\end{aligned} \tag{69}$$

The first inequality follows from the triangle inequality; the second holds since ϵ bounds the distributional discrepancy between X and Y; and the third follows from (68).

As with the preceding equations, all bounds in Theorem 5.1 hold deterministically for a given distributional discrepancy, where the discrepancy between $\hat{F}_Y$ and F_X is ϵ'. Consequently, the remaining probabilistic guarantees hold with probability at least $1-\delta$. □

B CVaR Bounds Proofs

Theorem 5.1 (restated). *Let X and Y be random variables and $\epsilon\in[0,1]$.*

(1) ***Upper Bound:*** *assume that finite $b_X, b_Y\in\mathbb{R}$ satisfy $b_X\geq\operatorname{ess\,sup}X$ and $b_Y\geq\operatorname{ess\,sup}Y$, and $\forall z\in\mathbb{R}, F_Y(z)-F_X(z)\leq\epsilon$, then*
   (a) *If $\alpha>\epsilon$,*
$$\mathrm{CVaR}_\alpha(X)\leq\frac{\epsilon}{\alpha}\max(b_X,b_Y)+(1-\frac{\epsilon}{\alpha})\,\mathrm{CVaR}_{\alpha-\epsilon}(Y). \tag{70}$$
   (b) *If $\alpha\leq\epsilon$, then $\mathrm{CVaR}_\alpha(X)\leq\max(b_X,b_Y)$.*

(2) ***Lower Bound:*** *If finite $a_X, a_Y\in\mathbb{R}$ satisfy $a_X\leq\operatorname{ess\,inf}X$ and $a_Y\leq\operatorname{ess\,inf}Y$, and $\forall z\in\mathbb{R}, F_X(z)-F_Y(z)\leq\epsilon$, then*
   (a) *If $\alpha+\epsilon<1$,*
$$\mathrm{CVaR}_\alpha(X)\geq(1+\frac{\epsilon}{\alpha})\,\mathrm{CVaR}_{\alpha+\epsilon}(Y)-\frac{\epsilon}{\alpha}\,\mathrm{CVaR}_\epsilon(Y) \tag{71}$$
   (b) *If $\alpha+\epsilon\geq1$ and $a_{min}=\min(a_X,a_Y)$,*
$$\mathrm{CVaR}_\alpha(X)\geq\frac{1}{\alpha}\big(\mathbb{E}[Y]-\epsilon\,\mathrm{CVaR}_\epsilon(Y)+(\alpha+\epsilon-1)a_{min}\big) \tag{72}$$

Proof. The strategy of the proof is to construct two distributions derived from F_Y, denoted F_Y^L and F_Y^U, such that F_X is stochastically bounded between them; that is, $F_Y^L\leq F_X\leq F_Y^U$. Consequently, since CVaR is a coherent risk measure, it follows that

$$\mathrm{CVaR}_\alpha^{F_Y^L}\leq\mathrm{CVaR}_\alpha^{F_X}\leq\mathrm{CVaR}_\alpha^{F_Y^U}, \tag{73}$$

where $\mathrm{CVaR}_\alpha^{F_Y^L}$, $\mathrm{CVaR}_\alpha^{F_X}$, and $\mathrm{CVaR}_\alpha^{F_Y^U}$ denote the CVaR at level α corresponding to the distributions F_Y^L, F_X, and F_Y^U, respectively.

Let $a_{\min}\triangleq\min(a_X,a_Y)$ and $b_{\min}\triangleq\min(b_X,b_Y)$. Define the interval $[b_{\min},b_X)$ to be $\emptyset$ if $b_{\min}=b_X$, and equal to $[b_Y,b_X)$ otherwise. Analogously, define $[a_{\min},a_X)$ in the same manner. We then define upper and lower bounds

for F_Y as follows (see Figure 16):

$$F_Y^U(y) \triangleq \begin{cases} 0 & y < \max(a_X, q^Y_{1-\epsilon}) \\ \min(F_Y(y) - \epsilon, 1-\epsilon) & y \in [\max(a_X, q^Y_{1-\epsilon}), b_{min}) \\ 1-\epsilon & y \in [b_{min}, b_X) \\ 1 & y \geq b_X, \end{cases} \tag{74}$$

$$F_Y^L(y) \triangleq \begin{cases} 0 & y < a_{min} \\ \epsilon & y \in [a_{min}, a_Y) \\ \min(F_Y(y) + \epsilon, 1) & y \in [a_Y, \min(q^Y_{\epsilon}, b_X)) \\ 1 & y \geq \min(q^Y_{\epsilon}, b_X). \end{cases} \tag{75}$$

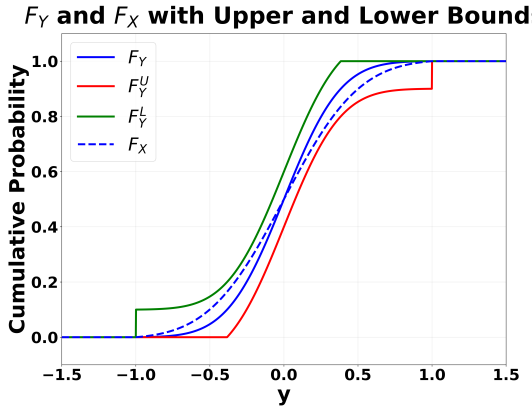


Fig. 16. Illustration of the upper and lower bounds F_Y^U and F_Y^L on the cumulative distribution function F_X.

As a first step, we verify that F_Y^L and F_Y^U are valid CDFs and that they satisfy $F_Y^L \leq F_X \leq F_Y^U$. To establish that F is a CDF, it suffices to verify the following properties:

(1) F is non-decreasing;
(2) $F : \mathbb{R} \to [0, 1]$ with $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$;
(3) F is right-continuous.

Proof that F_Y^L is a CDF:

(1) On the interval $[a_Y, \min(q^Y_{1-\epsilon}, b_X))$, the function F_Y^L is monotone increasing, since F_Y is monotone increasing by virtue of being a CDF. Outside this interval, F_Y^L is constant and consequently preserves monotonicity.
(2) By definition, $\lim_{y\to\infty} F_Y^L(y) = 1$ and $\lim_{y\to-\infty} F_Y^L(y) = 0$. We need to show that F_Y^L is bounded between 0 and 1. By its definition, F_Y^L is bounded between 0 and 1.
(3) Within the interval $[a_Y, \min(q^Y_{\epsilon}, b_X))$, F_Y^L is right-continuous since F_Y is right-continuous as a CDF. Outside this interval, F_Y^L is constant and hence also right-continuous.

Proof that F_Y^U is a CDF:

(1) On the interval $[\max(a_X, q^Y_{1-\epsilon}), b_{min})$, the function F_Y^U is monotone increasing, since F_Y is monotonically increasing by virtue of being a CDF. Outside this interval, F_Y^U is constant and consequently preserves monotonicity.

(2) By definition, $\lim_{y\to\infty} F_Y^U(y) = 1$ and $\lim_{y\to-\infty} F_Y^U(y) = 0$. By its definition, F_Y^U is bounded between 0 and 1.
(3) Within the interval $[\max(a_X, q^Y_{1-\epsilon}), b_{min})$, F_Y^U is right-continuous since F_Y is right-continuous as a CDF. Outside this interval, F_Y^U is constant and hence also right-continuous.

Proof that $F_Y^L \le F_X \le F_Y^U$: To establish that $F_Y^L \le F_X \le F_Y^U$, we need to show that for all $z \in \mathbb{R}$, $F_Y^L(z) \ge F_X(z) \ge F_Y^U(z)$. Let $z \in \mathbb{R}$.

- If $z < a_{min}$ then $F_Y^L(z) = 0 = F_X(z)$.
- If $z \in [a_{min}, a_Y)$ then $F_X(y) = F_X(y) - F_Y(y) + F_Y(y) \le |F_X(y) - F_Y(y)| + F_Y(y) \le \epsilon + F_Y(y) = \epsilon = F_Y^L(y)$.
- If $z \in [a_Y, \min(q^Y_\epsilon, b_X))$ we assume that $F_Y(z) + \epsilon \le 1$ because otherwise $F_Y^L(z) = 1$ and the inequality holds. $F_Y^L(z) = F_Y(z) + \epsilon = F_Y(z) - F_X(z) + F_X(z) + \epsilon \ge F_X(z) - |F_Y(z) - F_X(z)| + \epsilon \ge F_X(z)$.
- If $z \ge \min(q^Y_\epsilon, b_X)$ then $F_Y^L(z) = 1 \ge F_X(z)$.

and therefore $F_Y^L \le F_X$. Note that the last proof holds when $b_X = \infty$, making F_Y^L a valid CDF when the support of X is not bounded from above.

- If $z < \max(a_X, q^Y_{1-\epsilon})$ then $F_Y^U(z) = 0 \le F_X(z)$.
- If $z \in [\max(a_X, q^Y_{1-\epsilon}), b_{min})$ then $F_Y^U(z) \le F_Y(z) - \epsilon = F_Y(z) - F_X(z) + F_X(z) - \epsilon \le |F_Y(z) - F_X(z)| + F_X(z) - \epsilon \le \epsilon + F_X(z) - \epsilon = F_X(z)$.
- If $z \in [b_{min}, b_X)$ then $F_Y^U(z) = 1 - \epsilon = F_Y(z) - \epsilon = F_Y(z) - \epsilon + F_X(z) - F_X(z) \le F_X(z) - \epsilon + |F_Y(z) - F_X(z)| \le F_X(z)$

and therefore $F_X \le F_Y^U$. Note that the last proof holds when $a_X = -\infty$, making F_Y^U a valid CDF when the support of X is not bounded from below.

Next, we derive bounds for the CVaR associated with F_Y^U and F_Y^L.

Upper bound for $\text{CVaR}_\alpha^{F_Y^U}$: From Acerbi and Tasche (2002), CVaR is equal to an integral of the VaR

$$\text{CVaR}_\alpha^{F_Y^U} = \frac{1}{\alpha}\int_{1-\alpha}^{1} \inf\{y \in \mathbb{R} : F_Y^U(y) \ge \tau\} d\tau = \frac{1}{\alpha}\int_{1-\alpha+\epsilon}^{1+\epsilon} \inf\{y \in \mathbb{R} : F_Y^U(y) \ge \tau - \epsilon\} d\tau \tag{76}$$

If $\alpha \le \epsilon$, then $1 - \alpha + \epsilon \ge 1$, rendering the bound in the preceding equation trivial, as it attains the maximum value of the support of both X and Y.

$$\frac{1}{\alpha}\int_{1-\alpha+\epsilon}^{1+\epsilon} \inf\{y \in \mathbb{R} : F_Y^U(y) \ge \tau - \epsilon\} \le \frac{1}{\alpha}\int_{1-\alpha+\epsilon}^{1+\epsilon} \max(b_X, b_Y) d\tau = \max(b_X, b_Y). \tag{77}$$

That is, $\text{CVaR}_\alpha(X) \le \max(b_X, b_Y)$.

If $\alpha > \epsilon$, then $1 - \alpha + \epsilon < 1$, and the integral may be decomposed into one term that is trivially bounded by the maximum of the support and another term that can be computed explicitly

$$\begin{aligned}\text{CVaR}_\alpha^{F_Y^U} = \frac{1}{\alpha}\int_{1-\alpha+\epsilon}^{1+\epsilon} \inf\{y \in \mathbb{R} : F_Y^U(y) \ge \tau - \epsilon\} d\tau = \frac{1}{\alpha}\Big[\underbrace{\int_{1}^{1+\epsilon} \inf\{y \in \mathbb{R} : F_Y^U(y) \ge \tau - \epsilon\} d\tau}_{\triangleq A_1} \\ + \underbrace{\int_{1-\alpha+\epsilon}^{1} \inf\{y \in \mathbb{R} : F_Y^U(y) \ge \tau - \epsilon\} d\tau}_{\triangleq A_2}\Big]\end{aligned} \tag{78}$$

The term A_1 is bounded, in a manner analogous to (77), by $\epsilon \max(b_X, b_Y)$, which is essentially the tightest bound attainable given the definition of F_Y^U. The term A_2 can be expressed in terms of the CVaR of Y, evaluated at a shifted confidence level. Specifically, the variable $\tau - \epsilon$ in A_2 lies within the interval $[1 - \alpha, 1 - \epsilon]$. Over this range,

for all $y \in [\max(a_X, q^Y_{1-\epsilon}), b_{\min}]$, the inequality $F^U_Y(y) \leq F_Y(y) - \epsilon$ holds. This follows because if $q^Y_{1-\epsilon} \geq a_X$, then by definition $F^U_Y(y) = \min(F_Y(y) - \epsilon, 1 - \epsilon)$, and if $a_X > q^Y_{1-\epsilon}$, then $F^U_Y(y) = 0 \leq F_Y(y) - \epsilon$ for $y \in [q^Y_{1-\epsilon}, a_X]$ and again $F^U_Y(y) = \min(F_Y(y) - \epsilon, 1 - \epsilon)$ for $y \in [a_X, b_{min}]$.

$$A_2 \leq \int_{1-\alpha+\epsilon}^{1} \inf\{y \in \mathbb{R} : F_Y(y) - \epsilon \geq \tau - \epsilon\} d\tau = \int_{1-(\alpha-\epsilon)}^{1} \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\} d\tau = (\alpha - \epsilon)\,\mathrm{CVaR}_{\alpha-\epsilon}(Y). \tag{79}$$

Note that the confidence level is valid, as $\alpha, \epsilon \in (0, 1)$ and $\alpha > \epsilon$ imply $\alpha - \epsilon \in (0, 1)$. By combining the previous two expressions, we obtain a single bound:

$$\mathrm{CVaR}^{F^U_Y}_\alpha \leq \epsilon \frac{\max(b_X, b_Y)}{\alpha} + (1 - \frac{\epsilon}{\alpha})\,\mathrm{CVaR}_{\alpha-\epsilon}(Y). \tag{80}$$

In the case where $a_X = -\infty$, we have $\max(a_X, q^Y_{1-\epsilon}) = q^Y_{1-\epsilon}$, and the definition of F^U_Y no longer involves a_X. Therefore, the inequality remains valid in the case of $a_X = -\infty$.

Lower bound for $\mathrm{CVaR}^{F^L_Y}_\alpha$**:** From Acerbi and Tasche (2002), CVaR is equal to an integral of the VaR

$$\mathrm{CVaR}^{F^L_Y}_\alpha = \frac{1}{\alpha} \int_{1-\alpha}^{1} \inf\{y \in \mathbb{R} : F^L_Y(y) \geq \tau\} d\tau = \frac{1}{\alpha} \int_{1-(\epsilon+\alpha)}^{1-\epsilon} \inf\{y \in \mathbb{R} : F^L_Y(y) \geq \tau + \epsilon\} d\tau. \tag{81}$$

It holds that $F^L_Y(y) \leq F_Y(y) + \epsilon$ and therefore

$$\begin{aligned}
\mathrm{CVaR}^{F^L_Y}_\alpha &\geq \frac{1}{\alpha} \int_{1-(\epsilon+\alpha)}^{1-\epsilon} \inf\{y \in \mathbb{R} : F_Y(y) + \epsilon \geq \tau + \epsilon\} d\tau = \frac{1}{\alpha} \int_{1-(\epsilon+\alpha)}^{1-\epsilon} \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\} d\tau \\
&= \frac{1}{\alpha} [\int_{1-(\epsilon+\alpha)}^{1} \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\} d\tau - \int_{1-\epsilon}^{1} \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\} d\tau] \\
&= \frac{1}{\alpha} [(\alpha + \epsilon)\,\mathrm{CVaR}_{\alpha+\epsilon}(Y) - \epsilon\,\mathrm{CVaR}_\epsilon(Y)].
\end{aligned} \tag{82}$$

In the case where $\epsilon + \alpha > 1$,

$$\begin{aligned}
\mathrm{CVaR}^{F^L_Y}_\alpha &= \frac{1}{\alpha} \int_{1-(\epsilon+\alpha)}^{1-\epsilon} \inf\{y \in \mathbb{R} : F^L_Y(y) \geq \tau + \epsilon\} d\tau \\
&= \frac{1}{\alpha} [\underbrace{\int_{0}^{1-\epsilon} \inf\{y \in \mathbb{R} : F^L_Y(y) \geq \tau + \epsilon\} d\tau}_{\triangleq A_3} + \underbrace{\int_{1-(\epsilon+\alpha)}^{0} \inf\{y \in \mathbb{R} : F^L_Y(y) \geq \tau + \epsilon\} d\tau}_{\triangleq A_4}].
\end{aligned} \tag{83}$$

$$\begin{aligned}
A_3 &\geq \int_{0}^{1-\epsilon} \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\} d\tau = \int_{0}^{1} \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\} d\tau - \int_{1-\epsilon}^{1} \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\} d\tau \\
&= \mathbb{E}[Y] - \epsilon\,\mathrm{CVaR}_\epsilon(Y)
\end{aligned} \tag{84}$$

In the case of A_4, after the change of variable, τ ranges from $1 - \alpha$ to ϵ, and therefore

$$A_4 = \int_{1-\alpha}^{\epsilon} \inf\{y \in \mathbb{R} : F^L_Y(y) \geq \tau\} d\tau \geq (\alpha + \epsilon - 1) a_{min} \tag{85}$$

By combining the last equations to one bound we get

$$\mathrm{CVaR}_\alpha(X) \geq \mathrm{CVaR}^{F^L_Y}_\alpha \geq \frac{1}{\alpha}\big(\mathbb{E}[Y] - \epsilon\,\mathrm{CVaR}_\epsilon(Y) + (\alpha + \epsilon - 1) a_{min}\big). \tag{86}$$

□

Theorem 5.3 (restated). *Given the definition of X, Y, ϵ and α as in Theorem 5.1, the lower and upper bounds of Theorem 5.1 converge to* $\mathrm{CVaR}_\alpha(X)$ *as* $\epsilon \to 0$.

Proof. Let f and g be continuous functions such that $f(0)$ and $g(0)$ are finite. Then $\lim_{x\to 0} f(x)g(x) = \big(\lim_{x\to 0} f(x)\big)\big(\lim_{x\to 0} g(x)\big)$. This property will be used throughout the proof.

Assume that $F_Y(z) - F_X(z) \leq \epsilon$ for all $z \in \mathbb{R}$. Since we consider the limit as $\epsilon \to 0$, we further assume that $\alpha > \epsilon$ when computing the bound. Noting that CVaR is continuous with respect to the confidence level, it follows that $\lim_{\epsilon\to 0} \mathrm{CVaR}_{\alpha-\epsilon}(Y) = \mathrm{CVaR}_\alpha(Y)$.

$$\begin{aligned}
&\lim_{\epsilon\to 0} \frac{\epsilon}{\alpha} \max(b_X, b_Y) + (1 - \frac{\epsilon}{\alpha})\,\mathrm{CVaR}_{\alpha-\epsilon}(Y) \\
&= \lim_{\epsilon\to 0} (1 - \frac{\epsilon}{\alpha})\,\mathrm{CVaR}_{\alpha-\epsilon}(Y) \\
&= \lim_{\epsilon\to 0} 1 - \frac{\epsilon}{\alpha} \lim_{\epsilon\to 0} \mathrm{CVaR}_{\alpha-\epsilon}(Y) = \mathrm{CVaR}_\alpha(Y)
\end{aligned} \tag{87}$$

If $F_X(z) - F_Y(z) \leq \epsilon$ for all $z \in \mathbb{R}$, then, since we consider the limit as $\epsilon \to 0$, we assume in the derivation of the bound that $\alpha + \epsilon \leq 1$.

$$\lim_{\epsilon\to 0} (1 + \frac{\epsilon}{\alpha})\,\mathrm{CVaR}_{\alpha+\epsilon}(Y) - \frac{\epsilon}{\alpha}\,\mathrm{CVaR}_\epsilon(Y) = \lim_{\epsilon\to 0} 1 + \frac{\epsilon}{\alpha} \lim_{\epsilon\to 0} \mathrm{CVaR}_{\alpha+\epsilon}(Y) - \lim_{\epsilon\to 0} \frac{\epsilon}{\alpha} \lim_{\epsilon\to 0} \mathrm{CVaR}_\epsilon(Y) = \mathrm{CVaR}_\alpha(Y). \tag{88}$$

□

C Pointwise Discrepancy Refinements

This appendix provides the detailed discussion and proofs for the pointwise discrepancy refinements introduced in Section 5.1.

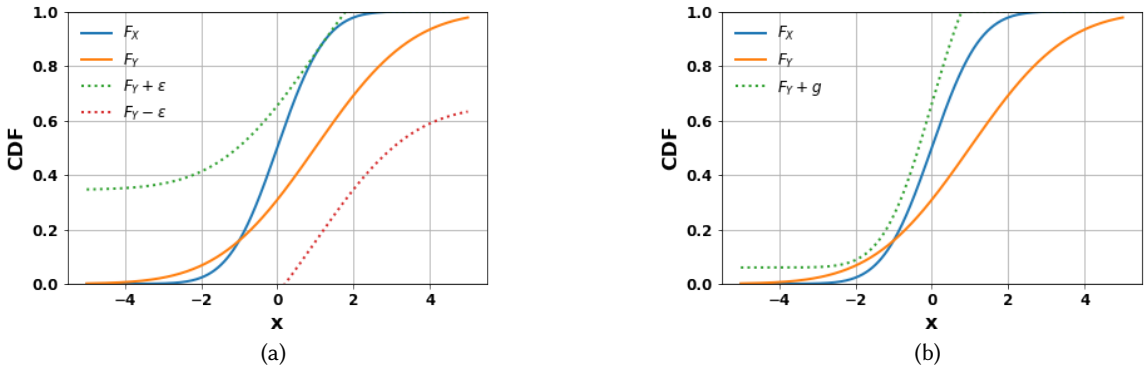


Fig. 17. Illustrations of bounds on $F_X(x)$. (a) Bounds from Theorem 5.1, where ϵ is a uniform bound. (b) Pointwise discrepancy bound $g(x)$ (Theorem C.1), yielding a tighter result than $F_Y(x) + \epsilon$.

Pointwise CDF discrepancy. Theorem 5.1 assumes a uniform bound ϵ on the CDF discrepancy. When the actual discrepancy varies with x (Figure 17b), a pointwise bound $|F_X(x) - F_Y(x)| \leq g(x)$ can yield tighter results. Theorem C.1 constructs the largest valid CDF $F_{Y^L}(y) = \min(1, F_Y(y) + g(y))$ within the discrepancy band; since $F_X \leq F_{Y^L}$ by construction, monotonicity of CVaR yields the lower bound $\text{CVaR}_\alpha(Y^L) \leq \text{CVaR}_\alpha(X)$. This extends Theorem 5.1, which is recovered when $g(x) = \epsilon$ for all x.

Theorem C.1 (Tighter CVaR Lower Bound). *Let $\alpha \in (0,1)$, X and Y be random variables. Define the random variable Y^L such that $F_{Y^L}(y) \triangleq \min(1, F_Y(y) + g(y))$ for $g : \mathbb{R} \to [0,\infty)$. Assume $\lim_{x\to-\infty} g(x) = 0$, g is continuous from the right and monotonic increasing. If $\forall x \in \mathbb{R}, F_X(x) \leq F_Y(x) + g(x)$, then F_{Y^L} is a CDF and* $\text{CVaR}_\alpha(Y^L) \leq \text{CVaR}_\alpha(X)$.

Proof. In order to prove that F_{Y^L} is a CDF we need to prove that F_{Y^L} is:

(1) Monotonic increasing
(2) $F : \mathbb{R} \to [0,1]$, $\lim_{x\to\infty} F_{Y^L}(x) = 1$, $\lim_{x\to-\infty} F_{Y^L}(x) = 0$
(3) Continuous from the right

Monotonic increasing: Note that for every $f_i : \mathbb{R} \to \mathbb{R}, i = 1, 2$ that are monotonic increasing, $f_1(f_2(x)))$ is also monotonic increasing in x. Denote $f(x) \triangleq \min(x, 1)$ and $f_2(x) \triangleq F_Y(x) + g(x)$. $F_Y(x)$ is a CDF and therefore monotonic increasing, so $f_2(x)$ is monotonic increasing as a sum of monotonic increasing functions. f_1 is also monotonic increasing, and $F_{Y^L}(x) = f_1(f_2(x))$. Therefore $F_{Y^L}(x)$ is monotonic increasing.

Limits:

$$1 \geq \lim_{x\to\infty} F_{Y^L}(x) = \lim_{x\to\infty} \min(1, F_Y(x) + g(x)) \geq \lim_{x\to\infty} \min(1, F_Y(x)) = \lim_{x\to\infty} F_Y(x) = 1 \tag{89}$$

and therefore $\lim_{x\to\infty} F_{Y^L}(x) = 1$.

$$0 \leq \lim_{x\to-\infty} F_{Y^L}(x) = \lim_{x\to-\infty} \min(1, F_Y(x) + g(x)) \leq \lim_{x\to-\infty} F_Y(x) + g(x) = \lim_{x\to-\infty} F_Y(x) + \lim_{x\to-\infty} g(x) = 0$$

and therefore $\lim_{x\to-\infty} F_{Y^L}(x) = 0$. By definition $\forall x \in \mathbb{R}, F_{Y^L}(x) \leq 1$, and $\forall x \in \mathbb{R}, F_{Y^L}(x) \geq 0$ because both g and F_{Y^L} are non negative functions.

Continuity from the right: F_Y is continuous from the right because it is a CDF, and g is continuous from the right by assumption. Their sum $F_Y + g$ is therefore continuous from the right, and since the constant function 1 is continuous, $F_{Y^L} = \min(1, F_Y + g)$ is continuous from the right as the minimum of two right-continuous functions. Thus, F_{Y^L} is a CDF.

Bound proof: If $Y^L \leq X$, then $\text{CVaR}_\alpha(Y^L) \leq \text{CVaR}_\alpha(X)$ because CVaR is a coherent risk measure. Note that if $F_Y(x) + g(x) < 1$ then

$$F_X(x) \leq F_Y(x) + g(x) = F_{Y^L}(x),$$

and if $F_Y(x) + g(x) \geq 1$, $1 = F_{Y^L}(x) \geq F_X(x)$. Therefore $Y^L \leq X$. □

Theorem C.2. *Let $\alpha \in (0,1)$, X and Y be random variables. Define $h : \mathbb{R} \to [0,\infty)$ to be a continuous function, $g(z) \triangleq \int_{-\infty}^{z} h(x)dx$ and Y^L to be a random variable such that $F_{Y^L}(y) \triangleq \min(1, F_Y(y) + g(y))$. If $\lim_{z\to-\infty} g(z) = 0$ and $\forall x \in \mathbb{R}, f_x(x) \leq f_y(x) + h(x)$, then F_{Y^L} is a CDF and* $\text{CVaR}_\alpha(Y^L) \leq \text{CVaR}_\alpha(X)$.

Proof. We will show the g satisfies the properties of Theorem C.1, and therefore this theorem holds. We need to prove that

(1) $\lim_{z\to-\infty} g(z) = 0$
(2) g is continuous from the right.

(3) g is monotonic increasing.
(4) $F_X(y) \leq F_Y(y) + g(y)$

It is given in the theorem's assumptions that $\lim_{z\to-\infty} g(z) = 0$, so (1) holds. h is non negative and therefore g is monotonic increasing, so (3) holds. g is continuous if its derivative exists for all $z \in \mathbb{R}$. Let $z \in \mathbb{R}$ and $a < z$.

$$\frac{d}{dz}g(z) = \frac{d}{dz}\int_{-\infty}^{z} h(x)dx = \frac{d}{dz}[\int_{-\infty}^{a} h(x)dx + \int_{a}^{z} h(x)dx] = \frac{d}{dz}\int_{a}^{z} h(x)dx = h(z) \tag{90}$$

where the third equality holds because $\int_{-\infty}^{a} h(x)dx = g(a)$ is a constant that does not depend on z. The last equality holds from the fundamental theorem of calculus because h is continuous. Finally, (4) holds because

$$F_X(y) \triangleq \int_{-\infty}^{y} f_x(x)dx \leq \int_{-\infty}^{y} f_y(x) + h(x)dx \triangleq F_Y(y) + g(y). \tag{91}$$

□

Without the monotonicity and right-continuity assumptions on g, the most general form of the CVaR bounds is:

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{\alpha}\int_{1-\alpha}^{1} \inf\{z \in \mathbb{R} : F_X(z) \geq \tau\}d\tau \geq \frac{1}{\alpha}\int_{1-\alpha}^{1} \inf\{z \in \mathbb{R} : F_Y(z) + g(z) \geq \tau\}d\tau, \tag{92}$$

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{\alpha}\int_{1-\alpha}^{1} \inf\{z \in \mathbb{R} : F_X(z) \geq \tau\}d\tau \leq \frac{1}{\alpha}\int_{1-\alpha}^{1} \inf\{z \in \mathbb{R} : F_Y(z) - g(z) \geq \tau\}d\tau. \tag{93}$$

Density-level discrepancy. Theorem C.2 demonstrates that the pointwise CDF discrepancy function g can be generated from a density-level discrepancy h via $g(z) = \int_{-\infty}^{z} h(x)\,dx$. This construction automatically satisfies the monotonicity and right-continuity requirements of Theorem C.1, showing that the pointwise bounds can be derived from a broad class of density discrepancy functions.

D Truncation Decomposition Proofs

Proposition 5.4 (restated)**.** *Let X be a random variable and $b \in \mathbb{R}$ with $b \geq \operatorname{ess\,sup} X$, and let $\kappa \in \mathbb{R}$. Define $X_\kappa \triangleq \min(X, \kappa)$ and $p \triangleq \mathbb{P}[X > \kappa]$. Then:*

$$\mathrm{CVaR}_\alpha(X) = \mathrm{CVaR}_\alpha(X_\kappa) + \frac{1}{\alpha}\int_{1-\alpha}^{1} \left[F_X^{-1}(v) - \kappa\right]^+ dv. \tag{94}$$

Consequently,

$$\mathrm{CVaR}_\alpha(X_\kappa) \leq \mathrm{CVaR}_\alpha(X) \leq \mathrm{CVaR}_\alpha(X_\kappa) + \frac{\min(p,\,\alpha)}{\alpha}\,(b - \kappa). \tag{95}$$

Proof of Proposition 5.4. Under the generalized-inverse convention above, since $X_\kappa = \min(X, \kappa)$, the quantile function of X_κ is $F_{X_\kappa}^{-1}(v) = \min(F_X^{-1}(v), \kappa)$. Using the identity $x = \min(x, \kappa) + [x - \kappa]^+$ with $x = F_X^{-1}(v)$:

$$\begin{aligned}
\mathrm{CVaR}_\alpha(X) &= \frac{1}{\alpha}\int_{1-\alpha}^{1} F_X^{-1}(v)\,dv \\
&= \frac{1}{\alpha}\int_{1-\alpha}^{1} \left[\min\left(F_X^{-1}(v),\,\kappa\right) + \left[F_X^{-1}(v) - \kappa\right]^+\right] dv \\
&= \underbrace{\frac{1}{\alpha}\int_{1-\alpha}^{1} F_{X_\kappa}^{-1}(v)\,dv}_{=\mathrm{CVaR}_\alpha(X_\kappa)} + \frac{1}{\alpha}\int_{1-\alpha}^{1} \left[F_X^{-1}(v) - \kappa\right]^+ dv,
\end{aligned} \tag{96}$$

which gives (26). The correction term is non-negative, so $\mathrm{CVaR}_\alpha(X_\kappa) \le \mathrm{CVaR}_\alpha(X)$. For the upper bound, the integrand $[F_X^{-1}(v)-\kappa]^+ > 0$ only when $v > F_X(\kappa) = 1-p$, so the effective integration region is $[1-\alpha, 1] \cap (1-p, 1]$, which has measure $\min(p, \alpha)$. On this interval $F_X^{-1}(v) \le b$, so:

$$\frac{1}{\alpha}\int_{1-\alpha}^{1} \left[F_X^{-1}(v) - \kappa\right]^+ dv \le \frac{\min(p,\ \alpha)}{\alpha}\,(b-\kappa). \tag{97}$$

□

Proposition 5.5 (restated)**.** *Let X be a random variable and $a \in \mathbb{R}$ with $a \le \operatorname{ess\,inf} X$, and let $\kappa \in \mathbb{R}$. Define $X^\kappa \triangleq \max(X, \kappa)$ and $q \triangleq \mathbb{P}[X < \kappa]$. Then:*

$$\mathrm{CVaR}_\alpha(X^\kappa) - \frac{\min(q,\ \alpha)}{\alpha}\,(\kappa - a) \le \mathrm{CVaR}_\alpha(X) \le \mathrm{CVaR}_\alpha(X^\kappa). \tag{98}$$

Proof of Proposition 5.5. Under the generalized-inverse convention above, since $X^\kappa = \max(X, \kappa)$, the quantile function of X^κ is $F_{X^\kappa}^{-1}(v) = \max(F_X^{-1}(v), \kappa)$. Using the identity $x = \max(x, \kappa) - [\kappa - x]^+$ with $x = F_X^{-1}(v)$:

$$\begin{aligned}
\mathrm{CVaR}_\alpha(X) &= \frac{1}{\alpha}\int_{1-\alpha}^{1} F_X^{-1}(v)\,dv \\
&= \frac{1}{\alpha}\int_{1-\alpha}^{1} \left[\max\left(F_X^{-1}(v),\ \kappa\right) - \left[\kappa - F_X^{-1}(v)\right]^+\right] dv \\
&= \underbrace{\frac{1}{\alpha}\int_{1-\alpha}^{1} F_{X^\kappa}^{-1}(v)\,dv}_{=\mathrm{CVaR}_\alpha(X^\kappa)} - \frac{1}{\alpha}\int_{1-\alpha}^{1} \left[\kappa - F_X^{-1}(v)\right]^+ dv,
\end{aligned} \tag{99}$$

which gives (28). The correction term is non-negative, so $\mathrm{CVaR}_\alpha(X) \le \mathrm{CVaR}_\alpha(X^\kappa)$. For the lower bound, the integrand $[\kappa - F_X^{-1}(v)]^+ > 0$ only when $F_X^{-1}(v) < \kappa$, equivalently (up to an endpoint of measure zero), when $v \le F_X(\kappa^-) = \mathbb{P}[X < \kappa] = q$. Thus the effective integration region has measure at most $\min(q, \alpha)$. On this interval $F_X^{-1}(v) \ge a$, so:

$$\frac{1}{\alpha}\int_{1-\alpha}^{1} \left[\kappa - F_X^{-1}(v)\right]^+ dv \le \frac{\min(q,\ \alpha)}{\alpha}\,(\kappa - a). \tag{100}$$

□

Lemma D.1 (Discrepancy Preservation). *Let X and Y be random variables and $\epsilon \ge 0$.*

(1) *If $F_Y(z) - F_X(z) \le \epsilon$ for all z, then for any κ and $X_\kappa \triangleq \min(X, \kappa)$: $F_Y(z) - F_{X_\kappa}(z) \le \epsilon$ for all z.*
(2) *If $F_X(z) - F_Y(z) \le \epsilon$ for all z, then for any κ and $X^\kappa \triangleq \max(X, \kappa)$: $F_{X^\kappa}(z) - F_Y(z) \le \epsilon$ for all z.*

Proof of Lemma D.1. *Part 1.* For $z < \kappa$: $F_{X_\kappa}(z) = F_X(z)$, so $F_Y(z) - F_{X_\kappa}(z) = F_Y(z) - F_X(z) \le \epsilon$. For $z \ge \kappa$: $F_{X_\kappa}(z) = 1 \ge F_Y(z)$, so $F_Y(z) - F_{X_\kappa}(z) \le 0 \le \epsilon$.

Part 2. For $z < \kappa$: $F_{X^\kappa}(z) = 0 \le F_Y(z)$, so $F_{X^\kappa}(z) - F_Y(z) \le 0 \le \epsilon$. For $z \ge \kappa$: $F_{X^\kappa}(z) = F_X(z)$, so $F_{X^\kappa}(z) - F_Y(z) = F_X(z) - F_Y(z) \le \epsilon$. □

Lemma D.2 (Tail Probability Bound). *Let $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} F_Y$ and $\hat{b}_Y \triangleq \max_i Y_i$. If $\sup_z (F_Y(z) - F_X(z)) \le \epsilon$, then for any $\delta \in (0, 1)$:*

$$\mathbb{P}\left[1 - F_X(\hat{b}_Y) \le 1 - \delta^{1/n} + \epsilon\right] \ge 1 - \delta. \tag{101}$$

For $Y = X$ (i.e., $\epsilon = 0$): $\mathbb{P}[1 - F_X(X^{(n)}) \le 1 - \delta^{1/n}] \ge 1 - \delta$.

Proof of Lemma D.2. When F_Y is continuous, the probability integral transform gives $F_Y(Y_i) \overset{\text{iid}}{\sim} \text{Uniform}(0,1)$, so $F_Y(Y^{(n)}) \sim \text{Beta}(n,1)$ and $\mathbb{P}[F_Y(Y^{(n)}) \geq \delta^{1/n}] = 1-\delta$. For general (possibly discrete) F_Y, the stochastic domination $\mathbb{P}[F_Y(Y) \leq u] \leq u$ for all $u \in [0,1]$ yields, by independence, $\mathbb{P}[F_Y(Y^{(n)}) < \delta^{1/n}] = \mathbb{P}[\max_i F_Y(Y_i) < \delta^{1/n}] \leq (\delta^{1/n})^n = \delta$, so the same tail bound holds. Since $F_X(z) \geq F_Y(z) - \epsilon$ for all z, we have $1 - F_X(\hat{b}_Y) \leq 1 - F_Y(\hat{b}_Y) + \epsilon$ almost surely. Therefore,

$$\begin{aligned}\mathbb{P}\Big[1 - F_X(\hat{b}_Y) \leq 1 - \delta^{1/n} + \epsilon\Big] &\geq \mathbb{P}\Big[1 - F_Y(\hat{b}_Y) + \epsilon \leq 1 - \delta^{1/n} + \epsilon\Big] \\ &= \mathbb{P}\Big[F_Y(\hat{b}_Y) \geq \delta^{1/n}\Big] \geq 1 - \delta. \end{aligned} \tag{102}$$

□

Lemma D.3 (Lower Tail Probability Bound). *Let $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} F_Y$ with F_Y continuous, and let $\hat{a}_Y \triangleq \min_i Y_i$. If $\sup_z(F_X(z) - F_Y(z)) \leq \epsilon$, then for any $\delta \in (0,1)$:*

$$\mathbb{P}\Big[F_X(\hat{a}_Y) \leq 1 - \delta^{1/n} + \epsilon\Big] \geq 1 - \delta. \tag{103}$$

Proof of Lemma D.3. Since F_Y is continuous, the probability integral transform gives $F_Y(Y_i) \overset{\text{iid}}{\sim} \text{Uniform}(0,1)$, so $F_Y(Y^{(1)}) = \min_i F_Y(Y_i) \sim \text{Beta}(1,n)$ and $\mathbb{P}[F_Y(Y^{(1)}) \leq 1-\delta^{1/n}] = 1-(\delta^{1/n})^n = 1-\delta$. Since $F_X(z) \leq F_Y(z) + \epsilon$ for all z, we have $F_X(\hat{a}_Y) \leq F_Y(\hat{a}_Y) + \epsilon$ almost surely. Therefore,

$$\begin{aligned}\mathbb{P}\Big[F_X(\hat{a}_Y) \leq 1 - \delta^{1/n} + \epsilon\Big] &\geq \mathbb{P}\Big[F_Y(\hat{a}_Y) + \epsilon \leq 1 - \delta^{1/n} + \epsilon\Big] \\ &= \mathbb{P}\Big[F_Y(\hat{a}_Y) \leq 1 - \delta^{1/n}\Big] \geq 1 - \delta. \end{aligned} \tag{104}$$

□

Since $1 - e^{-x} \leq x$ for $x \geq 0$, for any $\delta \in (0,1)$ and every $n \geq 1$, $1 - \delta^{1/n} \leq \ln(1/\delta)/n = O(1/n)$, confirming the faster decay rate.

Theorem 5.6 (restated). *Let X and Y be random variables. Let $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} F_Y$ and let $\delta_1, \delta_2 \in (0, 0.5)$, $\alpha \in (0,1]$, $\epsilon \geq 0$. Define $\hat{b}_Y \triangleq \max_i Y_i$, $\hat{a}_Y \triangleq \min_i Y_i$. Let $\eta \geq 0$. Define the effective discrepancy $\epsilon' \triangleq \min(\epsilon + \eta, 1)$ and the tail factors $\hat{p} \triangleq \hat{q} \triangleq 1 - \delta_2^{1/n} + \epsilon$.*

Upper Bound. *Assume $\sup_z(F_Y(z) - F_X(z)) \leq \epsilon$, $\mathbb{P}[\sup_z(\hat{F}_Y(z) - F_Y(z)) \leq \eta] > 1 - \delta_1$, suppose $b_X \geq \operatorname{ess\,sup} X$, and let $\bar{b} \geq b_X$. If $\epsilon' < \alpha$:*

$$U \triangleq \underbrace{\left(1 - \frac{\epsilon'}{\alpha}\right) \text{CVaR}_{\alpha-\epsilon'}^{\hat{F}_Y} + \frac{\epsilon'}{\alpha}\, \hat{b}_Y}_{U_{\text{trunc}}} + \underbrace{\frac{\hat{p}}{\alpha}\,(\bar{b} - \hat{b}_Y)^+}_{\Delta_U}. \tag{105}$$

Then $\mathbb{P}[\text{CVaR}_\alpha(X) \leq U] > 1 - \delta_1 - \delta_2$.

Lower Bound. *Assume $\sup_z(F_X(z) - F_Y(z)) \leq \epsilon$, $\mathbb{P}[\sup_z(F_Y(z) - \hat{F}_Y(z)) \leq \eta] > 1 - \delta_1$, F_Y continuous, $a_X \leq \operatorname{ess\,inf} X$, and $\underline{a} \leq a_X$. If $\epsilon' + \alpha < 1$:*

$$L \triangleq \underbrace{\left(1 + \frac{\epsilon'}{\alpha}\right) \text{CVaR}_{\alpha+\epsilon'}^{\hat{F}_Y} - \frac{\epsilon'}{\alpha}\, \text{CVaR}_{\epsilon'}^{\hat{F}_Y}}_{L_{\text{trunc}}} - \underbrace{\frac{\hat{q}}{\alpha}\,(\hat{a}_Y - \underline{a})^+}_{\Delta_L}. \tag{106}$$

Then $\mathbb{P}[\text{CVaR}_\alpha(X) \geq L] > 1 - \delta_1 - \delta_2$.

Proof of Theorem 5.6. Both bounds follow the same structure: (i) a truncation decomposition splits $\mathrm{CVaR}_\alpha(X)$ into a bulk term and a tail correction; (ii) the bulk is bounded via a quantile chain linking X, Y, and the empirical CDF $\hat{F}_Y$; (iii) the tail correction is controlled by an order-statistic bound; (iv) a union bound over the two probabilistic events yields the confidence guarantee. We present each bound in full. We first state a quantile inversion fact used in both directions.

Lemma D.4 (Quantile inversion under CDF discrepancy). *Let G and H be CDFs satisfying $\sup_z(G(z) - H(z)) \le \epsilon$. Then $H^{-1}(v) \le G^{-1}(v+\epsilon)$ for all v with $v + \epsilon \le 1$.*

Proof. Fix v with $v + \epsilon \le 1$. For any z with $G(z) \ge v + \epsilon$, the discrepancy assumption gives

$$H(z) \ge G(z) - \epsilon \ge v, \tag{107}$$

so the two level sets satisfy

$$\{z \in \mathbb{R} : G(z) \ge v + \epsilon\} \subseteq \{z \in \mathbb{R} : H(z) \ge v\}. \tag{108}$$

The infimum over a subset is at least the infimum over its superset, so by the definition of the generalized inverse (Section 3.3)

$$G^{-1}(v+\epsilon) = \inf\{z : G(z) \ge v + \epsilon\} \ge \inf\{z : H(z) \ge v\} = H^{-1}(v). \tag{109}$$

If the first level set is empty, then $G^{-1}(v+\epsilon) = \inf \emptyset = +\infty$ and the bound holds trivially. □

Upper bound.

Truncation decomposition. Define $X_{\hat{b}_Y} \triangleq \min(X, \hat{b}_Y)$ and $p \triangleq \mathbb{P}[X > \hat{b}_Y] = 1 - F_X(\hat{b}_Y)$. By Proposition 5.4,

$$\mathrm{CVaR}_\alpha(X) \le \mathrm{CVaR}_\alpha(X_{\hat{b}_Y}) + \frac{p}{\alpha}(\bar{b} - \hat{b}_Y)^+. \tag{110}$$

Bounding the bulk $\mathrm{CVaR}_\alpha(X_{\hat{b}_Y})$. By Lemma D.1(1), $\sup_z(F_Y(z) - F_{X_{\hat{b}_Y}}(z)) \le \epsilon$. Applying Lemma D.4 with $G = F_Y$ and $H = F_{X_{\hat{b}_Y}}$:

$$F^{-1}_{X_{\hat{b}_Y}}(v) \le F_Y^{-1}(v+\epsilon) \quad \text{for } v + \epsilon \le 1. \tag{111}$$

By the CDF concentration assumption (i.e., $\mathbb{P}[\sup_z(\hat{F}_Y(z) - F_Y(z)) \le \eta] > 1 - \delta_1$), on this event, Lemma D.4 with $G = \hat{F}_Y$ and $H = F_Y$ gives:

$$F_Y^{-1}(v) \le \hat{F}_Y^{-1}(v+\eta) \quad \text{for } v + \eta \le 1. \tag{112}$$

Composing (111) and (112):

$$F^{-1}_{X_{\hat{b}_Y}}(v) \le \hat{F}_Y^{-1}(v+\epsilon') \quad \text{for } v + \epsilon' \le 1, \tag{113}$$

and $F^{-1}_{X_{\hat{b}_Y}}(v) \le \hat{b}_Y$ for all v (since $X_{\hat{b}_Y} \le \hat{b}_Y$ by construction). Using (113) on $[1-\alpha, 1-\epsilon']$ and the deterministic bound on $[1-\epsilon', 1]$:

$$\begin{aligned} \alpha\ \mathrm{CVaR}_\alpha(X_{\hat{b}_Y}) &= \int_{1-\alpha}^{1} F^{-1}_{X_{\hat{b}_Y}}(v)\, dv \le \int_{1-\alpha}^{1-\epsilon'} \hat{F}_Y^{-1}(v+\epsilon')\, dv + \int_{1-\epsilon'}^{1} \hat{b}_Y\, dv \\ &= (\alpha - \epsilon')\ \mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\epsilon'} + \epsilon'\, \hat{b}_Y, \end{aligned} \tag{114}$$

where the first integral follows from the substitution $u = v + \epsilon'$, giving $\int_{1-\alpha+\epsilon'}^{1} \hat{F}_Y^{-1}(u)\, du = (\alpha - \epsilon')\ \mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\epsilon'}$. Dividing by α yields $\mathrm{CVaR}_\alpha(X_{\hat{b}_Y}) \le U_{\mathrm{trunc}}$ with probability $> 1 - \delta_1$.

Bounding the tail probability p. By Lemma D.2, $\mathbb{P}[p \le \hat{p}] \ge 1 - \delta_2$, where $\hat{p} \triangleq 1 - \delta_2^{1/n} + \epsilon$.

Combining via union bound. Define the events $E_1 \triangleq \{\sup_z(\hat{F}_Y(z) - F_Y(z)) \leq \eta\}$ and $E_2 \triangleq \{p \leq \hat{p}\}$, with $\mathbb{P}[E_1^c] \leq \delta_1$ and $\mathbb{P}[E_2^c] \leq \delta_2$. On $E_1 \cap E_2$, the truncation decomposition (110) gives

$$\mathrm{CVaR}_\alpha(X) \leq \mathrm{CVaR}_\alpha(X_{\hat{b}_Y}) + \frac{p}{\alpha}(\bar{b} - \hat{b}_Y)^+ \leq U_{\text{trunc}} + \frac{\hat{p}}{\alpha}(\bar{b} - \hat{b}_Y)^+ = U_{\text{trunc}} + \Delta_U, \tag{115}$$

where the first term uses E_1 and the second uses E_2. Since $\{\mathrm{CVaR}_\alpha(X) > U_{\text{trunc}} + \Delta_U\} \subseteq E_1^c \cup E_2^c$:

$$\mathbb{P}[\mathrm{CVaR}_\alpha(X) > U_{\text{trunc}} + \Delta_U] \leq \mathbb{P}[E_1^c \cup E_2^c] \leq \mathbb{P}[E_1^c] + \mathbb{P}[E_2^c] \leq \delta_1 + \delta_2. \tag{116}$$

Lower bound.

Truncation decomposition. Define $X^{\hat{a}_Y} \triangleq \max(X, \hat{a}_Y)$ and $q \triangleq \mathbb{P}[X < \hat{a}_Y]$. By Proposition 5.5,

$$\mathrm{CVaR}_\alpha(X) \geq \mathrm{CVaR}_\alpha(X^{\hat{a}_Y}) - \frac{q}{\alpha}(\hat{a}_Y - \underline{a})^+. \tag{117}$$

Bounding the bulk $\mathrm{CVaR}_\alpha(X^{\hat{a}_Y})$. By Lemma D.1(2), $\sup_z(F_{X^{\hat{a}_Y}}(z) - F_Y(z)) \leq \epsilon$. Applying Lemma D.4 with $G = F_{X^{\hat{a}_Y}}$ and $H = F_Y$:

$$F_Y^{-1}(v) \leq F_{X^{\hat{a}_Y}}^{-1}(v + \epsilon) \quad \text{for } v + \epsilon \leq 1, \tag{118}$$

equivalently, $F_{X^{\hat{a}_Y}}^{-1}(v) \geq F_Y^{-1}(v - \epsilon)$ for $v > \epsilon$. On the CDF concentration event (i.e., $\mathbb{P}[\sup_z(F_Y(z) - \hat{F}_Y(z)) \leq \eta] > 1 - \delta_1$), Lemma D.4 with $G = F_Y$ and $H = \hat{F}_Y$ gives:

$$\hat{F}_Y^{-1}(v) \leq F_Y^{-1}(v + \eta) \quad \text{for } v + \eta \leq 1, \tag{119}$$

equivalently, $F_Y^{-1}(v) \geq \hat{F}_Y^{-1}(v - \eta)$ for $v > \eta$. Composing:

$$F_{X^{\hat{a}_Y}}^{-1}(v) \geq \hat{F}_Y^{-1}(v - \epsilon') \quad \text{for } v > \epsilon'. \tag{120}$$

Since $\alpha + \epsilon' < 1$ implies $1 - \alpha > \epsilon'$, the chain (120) applies throughout $[1 - \alpha, 1]$. Substituting $u = v - \epsilon'$:

$$\begin{aligned} \alpha\ \mathrm{CVaR}_\alpha(X^{\hat{a}_Y}) &= \int_{1-\alpha}^{1} F_{X^{\hat{a}_Y}}^{-1}(v)\, dv \geq \int_{1-\alpha}^{1} \hat{F}_Y^{-1}(v - \epsilon')\, dv = \int_{1-\alpha-\epsilon'}^{1-\epsilon'} \hat{F}_Y^{-1}(u)\, du \\ &= \int_{1-\alpha-\epsilon'}^{1} \hat{F}_Y^{-1}(u)\, du - \int_{1-\epsilon'}^{1} \hat{F}_Y^{-1}(u)\, du \\ &= (\alpha + \epsilon')\ \mathrm{CVaR}_{\alpha+\epsilon'}^{\hat{F}_Y} - \epsilon'\ \mathrm{CVaR}_{\epsilon'}^{\hat{F}_Y}. \end{aligned} \tag{121}$$

Dividing by α yields $\mathrm{CVaR}_\alpha(X^{\hat{a}_Y}) \geq L_{\text{trunc}}$ with probability $> 1 - \delta_1$.

Bounding the tail probability q. By Lemma D.3, $\mathbb{P}[F_X(\hat{a}_Y) \leq \hat{q}] \geq 1 - \delta_2$, where $\hat{q} \triangleq 1 - \delta_2^{1/n} + \epsilon$. Since $q = \mathbb{P}[X < \hat{a}_Y] \leq F_X(\hat{a}_Y)$, this implies $\mathbb{P}[q \leq \hat{q}] \geq 1 - \delta_2$.

Combining via union bound. Define the events $E_1 \triangleq \{\sup_z(F_Y(z) - \hat{F}_Y(z)) \leq \eta\}$ and $E_2 \triangleq \{q \leq \hat{q}\}$, with $\mathbb{P}[E_1^c] \leq \delta_1$ and $\mathbb{P}[E_2^c] \leq \delta_2$. On $E_1 \cap E_2$, the truncation decomposition (117) gives

$$\mathrm{CVaR}_\alpha(X) \geq \mathrm{CVaR}_\alpha(X^{\hat{a}_Y}) - \frac{q}{\alpha}(\hat{a}_Y - \underline{a})^+ \geq L_{\text{trunc}} - \frac{\hat{q}}{\alpha}(\hat{a}_Y - \underline{a})^+ = L_{\text{trunc}} - \Delta_L, \tag{122}$$

where the first term uses E_1 and the second uses E_2. Since $\{\mathrm{CVaR}_\alpha(X) < L_{\text{trunc}} - \Delta_L\} \subseteq E_1^c \cup E_2^c$:

$$\mathbb{P}[\mathrm{CVaR}_\alpha(X) < L_{\text{trunc}} - \Delta_L] \leq \mathbb{P}[E_1^c \cup E_2^c] \leq \mathbb{P}[E_1^c] + \mathbb{P}[E_2^c] \leq \delta_1 + \delta_2. \tag{123}$$

Degenerate cases. When $\epsilon' \geq \alpha$ (upper bound), the entire integration region $[1 - \alpha, 1]$ falls within $[1 - \epsilon', 1]$, so $F_{X_{\hat{b}_Y}}^{-1}(v) \leq \hat{b}_Y$ for all $v \in [1 - \alpha, 1]$, giving $\mathrm{CVaR}_\alpha(X_{\hat{b}_Y}) \leq \hat{b}_Y$. Thus $U_{\text{trunc}} = \hat{b}_Y$ and the upper bound becomes $U = \hat{b}_Y + (\hat{p}/\alpha)(\bar{b} - \hat{b}_Y)^+$ with probability $> 1 - \delta_2$ (no CDF concentration event is needed).

When $\epsilon' + \alpha \geq 1$ (lower bound), the shifted integration range $[1-\alpha-\epsilon', 1-\epsilon']$ extends below 0. We split the integral at $v = \epsilon'$. For $v \in [1-\alpha, \epsilon']$: $v - \epsilon' \leq 0$ and $F^{-1}_{X^{\hat{a}_Y}}(v) \geq \hat{a}_Y$ (since $X^{\hat{a}_Y} \geq \hat{a}_Y$), contributing $(\alpha+\epsilon'-1)\,\hat{a}_Y$. For $v \in [\epsilon', 1]$: the quantile chain (120) applies, and the substitution $u = v-\epsilon'$ gives $\int_0^{1-\epsilon'} \hat{F}_Y^{-1}(u)\,du = \mathbb{E}_{\hat{F}_Y}[Y] - \epsilon'\ \mathrm{CVaR}^{\hat{F}_Y}_{\epsilon'}$. Combining: $L_{\mathrm{trunc}} = \frac{1}{\alpha}\big(\mathbb{E}_{\hat{F}_Y}[Y] - \epsilon'\ \mathrm{CVaR}^{\hat{F}_Y}_{\epsilon'} + (\alpha+\epsilon'-1)\,\hat{a}_Y\big)$. □

Theorem 5.8 (restated). *Let X and Y be random variables, $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} F_Y$, $\delta_1, \delta_2, \delta_3 \in (0, 0.5)$, $\alpha \in (0, 1]$, and $\eta \geq 0$. Define $\hat{b}_Y \triangleq \max_i Y_i$ and $\hat{a}_Y \triangleq \min_i Y_i$.*

Upper Bound. *Define $\epsilon \triangleq \sup_z(F_Y(z) - F_X(z))$, and let $\hat{\epsilon} \geq 0$ be an estimator of ϵ. Define $\hat{\epsilon}' \triangleq \min(\hat{\epsilon}+\eta, 1)$ and $\hat{p} \triangleq 1 - \delta_2^{1/n} + \hat{\epsilon}$. Assume*

$$\mathbb{P}[\hat{\epsilon} \geq \epsilon] \geq 1-\delta_3, \qquad \mathbb{P}\big[\sup_z(\hat{F}_Y(z) - F_Y(z)) \leq \eta\big] > 1-\delta_1, \qquad \bar{b} \geq b_X \geq \operatorname{ess\,sup} X. \tag{124}$$

If $\hat{\epsilon}' < \alpha$:

$$U \triangleq \underbrace{\left(1 - \frac{\hat{\epsilon}'}{\alpha}\right)\mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\hat{\epsilon}'} + \frac{\hat{\epsilon}'}{\alpha}\,\hat{b}_Y}_{U_{\mathrm{trunc}}} + \underbrace{\frac{\hat{p}}{\alpha}\,(\bar{b} - \hat{b}_Y)^+}_{\Delta_U}. \tag{125}$$

If $\hat{\epsilon}' \geq \alpha$, define $U \triangleq \hat{b}_Y + \frac{\hat{p}}{\alpha}(\bar{b} - \hat{b}_Y)^+$. Then $\mathbb{P}[\mathrm{CVaR}_\alpha(X) \leq U] > 1 - \delta_1 - \delta_2 - \delta_3$.

Lower Bound. *Define $\epsilon \triangleq \sup_z(F_X(z) - F_Y(z))$, and let $\hat{\epsilon} \geq 0$ be an estimator of ϵ. Define $\hat{\epsilon}' \triangleq \min(\hat{\epsilon}+\eta, 1)$ and $\hat{q} \triangleq 1 - \delta_2^{1/n} + \hat{\epsilon}$. Assume F_Y is continuous and*

$$\mathbb{P}[\hat{\epsilon} \geq \epsilon] \geq 1-\delta_3, \qquad \mathbb{P}\big[\sup_z(F_Y(z) - \hat{F}_Y(z)) \leq \eta\big] > 1-\delta_1, \qquad \underline{a} \leq a_X \leq \operatorname{ess\,inf} X. \tag{126}$$

If $\hat{\epsilon}' + \alpha < 1$:

$$L \triangleq \underbrace{\left(1 + \frac{\hat{\epsilon}'}{\alpha}\right)\mathrm{CVaR}^{\hat{F}_Y}_{\alpha+\hat{\epsilon}'} - \frac{\hat{\epsilon}'}{\alpha}\ \mathrm{CVaR}^{\hat{F}_Y}_{\hat{\epsilon}'}}_{L_{\mathrm{trunc}}} - \underbrace{\frac{\hat{q}}{\alpha}\,(\hat{a}_Y - \underline{a})^+}_{\Delta_L}. \tag{127}$$

If $\hat{\epsilon}' + \alpha \geq 1$, define $L \triangleq \underline{a}$. Then $\mathbb{P}[\mathrm{CVaR}_\alpha(X) \geq L] > 1 - \delta_1 - \delta_2 - \delta_3$.

Proof of Theorem 5.8. We prove the upper bound; the lower bound follows by an analogous argument. Define the three events:

$$E_1 \triangleq \Big\{\sup_z\big(\hat{F}_Y(z) - F_Y(z)\big) \leq \eta\Big\}, \qquad \mathbb{P}(E_1^c) < \delta_1, \tag{128}$$

$$E_2 \triangleq \big\{1 - F_X(\hat{b}_Y) \leq 1 - \delta_2^{1/n} + \epsilon\big\}, \qquad \mathbb{P}(E_2^c) \leq \delta_2, \tag{129}$$

$$E_3 \triangleq \{\hat{\epsilon} \geq \epsilon\}, \qquad \mathbb{P}(E_3^c) \leq \delta_3. \tag{130}$$

On the event E_3, the estimated discrepancy satisfies $\hat{\epsilon} \geq \epsilon = \sup_z(F_Y(z) - F_X(z))$, and therefore, if $\hat{\epsilon}' < \alpha \leq 1$, $\hat{\epsilon}' = \hat{\epsilon} + \eta \geq \epsilon + \eta \geq \epsilon'$ (where ϵ' is the true effective discrepancy). On $E_1 \cap E_3$, the quantile chain from the proof of Theorem 5.6 holds with ϵ' replaced by $\hat{\epsilon}'$:

$$F^{-1}_{X_{\hat{b}_Y}}(v) \leq \hat{F}_Y^{-1}(v + \hat{\epsilon}') \quad \text{for } v + \hat{\epsilon}' \leq 1. \tag{131}$$

This is because the quantile inversion lemma (Lemma D.4) gives $F^{-1}_{X_{\hat{b}_Y}}(v) \leq F_Y^{-1}(v+\epsilon) \leq F_Y^{-1}(v+\hat{\epsilon})$ (using $\hat{\epsilon} \geq \epsilon$ from E_3 and monotonicity of F_Y^{-1}), and $F_Y^{-1}(v+\hat{\epsilon}) \leq \hat{F}_Y^{-1}(v+\hat{\epsilon}+\eta) = \hat{F}_Y^{-1}(v+\hat{\epsilon}')$ (from E_1). Integrating over

$[1-\alpha, 1]$ and using $F^{-1}_{X_{\hat{b}_Y}}(v) \leq \hat{b}_Y$:

$$\mathrm{CVaR}_\alpha(X_{\hat{b}_Y}) \leq \left(1 - \frac{\hat{\epsilon}'}{\alpha}\right) \mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\hat{\epsilon}'} + \frac{\hat{\epsilon}'}{\alpha}\hat{b}_Y = U_{\text{trunc}}. \tag{132}$$

The key observation is that the upper bound is non-decreasing in the effective discrepancy: replacing ϵ' with a larger $\hat{\epsilon}' \geq \epsilon'$ shifts more weight to $\hat{b}_Y$ (the sample maximum) and evaluates CVaR at a more conservative confidence level $\alpha - \hat{\epsilon}'$, both of which can only increase the bound.

Since E_2 is defined with the deterministic ϵ, Lemma D.2 gives $\mathbb{P}(E_2^c) \leq \delta_2$. On $E_2 \cap E_3$, the tail probability satisfies $p = 1 - F_X(\hat{b}_Y) \leq 1 - \delta_2^{1/n} + \epsilon \leq 1 - \delta_2^{1/n} + \hat{\epsilon} = \hat{p}$.

If $\hat{\epsilon}' < \alpha$, on $E_1 \cap E_2 \cap E_3$, by the truncation decomposition (Proposition 5.4):

$$\mathrm{CVaR}_\alpha(X) \leq \mathrm{CVaR}_\alpha(X_{\hat{b}_Y}) + \frac{p}{\alpha}(\bar{b} - \hat{b}_Y)^+ \leq U_{\text{trunc}} + \frac{\hat{p}}{\alpha}(\bar{b} - \hat{b}_Y)^+ = U. \tag{133}$$

If $\hat{\epsilon}' \geq \alpha$, on $E_2 \cap E_3$, $X_{\hat{b}_Y} \leq \hat{b}_Y$ gives $\mathrm{CVaR}_\alpha(X_{\hat{b}_Y}) \leq \hat{b}_Y$, and the same decomposition gives $\mathrm{CVaR}_\alpha(X) \leq U$. By the union bound:

$$\mathbb{P}(\mathrm{CVaR}_\alpha(X) > U) \leq \mathbb{P}(E_1^c \cup E_2^c \cup E_3^c) < \delta_1 + \delta_2 + \delta_3. \tag{134}$$

For the lower bound, use the lower bound's ϵ and $\hat{\epsilon}$, $E_1 \triangleq \{\sup_z(F_Y(z) - \hat{F}_Y(z)) \leq \eta\}$, $E_2 \triangleq \{F_X(\hat{a}_Y) \leq 1 - \delta_2^{1/n} + \epsilon\}$ (Lemma D.3, which uses continuity of F_Y), and $q = \mathbb{P}[X < \hat{a}_Y] \leq F_X(\hat{a}_Y) \leq \hat{q}$ on $E_2 \cap E_3$. If $\hat{\epsilon}' + \alpha \geq 1$, $\mathrm{CVaR}_\alpha(X) \geq \operatorname{ess\,inf} X \geq a_X \geq \underline{a} = L$ surely. □

E Static CVaR Simplification Proofs

Theorem 6.1 (restated)**.**

$$\sup_{l \in \mathbb{R}} |\mathbb{P}^\pi_{b_k,a_k}[R_{k:T} \leq l] - \mathbb{P}^{\pi,s}_{b_k,a_k}[R_{k:T} \leq l]| \leq \sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_t, a_t)], \tag{135}$$

where Δ^s is the TV distance that is defined by

$$\Delta^s(b_{t-1}, a_{t-1}) \triangleq \tfrac{1}{2} \int_{b_t \in B} |P(b_t|b_{t-1}, a_{t-1}) - P_s(b_t|b_{t-1}, a_{t-1})| db_t. \tag{136}$$

Proof. By expanding the belief path from time $k+1$ to time T we get

$$\begin{aligned} \mathbb{P}^\pi_{b_k}[\sum_{t=k}^{T} c(b_t, a_t) \leq l] &= \int_{b_{k+1:T} \in B^{T-k}} P(R_{k:T} \leq l | b_{k:T}, \pi) \prod_{i=k+1}^{T} P(b_i|b_{i-1}, \pi, a_i) db_{k+1:T} \\ &= \int_{b_{k+1:T} \in B^{T-k}} 1_{R_{k:T} \leq l} \prod_{i=k+1}^{T} P(b_i|b_{i-1}, \pi, a_i) db_{k+1:T}, \end{aligned} \tag{137}$$

where the second equality holds because $R_{k:T}$ is constant given π and $b_{k:T}$. Hence, the subtraction between the simplified and original return CDFs can be exhibited using the difference between the multiplication of the belief transition models.

$$\mathbb{P}^\pi_{b_k}[R_{k:T} \leq l] - \mathbb{P}^{\pi,s}_{b_k}[R_{k:T} \leq l] = \int_{b_{k+1:T} \in B^{T-k}} 1_{R_{k:T} \leq l} [\prod_{i=k+1}^{T} P(b_i|b_{i-1}, \pi, a_i) - \prod_{i=k+1}^{T} P_s(b_i|b_{i-1}, \pi, a_i)] db_{k+1:T} \tag{138}$$

By applying the triangle inequality we get the following bound

$$|\mathbb{P}^{\pi}_{b_k}[R_{k:T} \le l] - \mathbb{P}^{\pi,s}_{b_k}[R_{k:T} \le l]| \le \int_{b_{k+1:T} \in B^{T-k}} 1_{R_{k:T} \le l} |\prod_{i=k+1}^{T} P(b_i|b_{i-1},\pi,a_i) - \prod_{i=k+1}^{T} P_s(b_i|b_{i-1},\pi,a_i)| db_{k+1:T}$$
$$\le \int_{b_{k+1:T} \in B^{T-k}} |\prod_{i=k+1}^{T} P(b_i|b_{i-1},\pi,a_i) - \prod_{i=k+1}^{T} P_s(b_i|b_{i-1},\pi,a_i)| db_{k+1:T} = g(T), \tag{139}$$

The indicator can be centred: both products are densities over $b_{k+1:T}$, so $\int_{b_{k+1:T} \in B^{T-k}} \prod_{i=k+1}^{T} P(b_i|b_{i-1},\pi,a_i) db_{k+1:T}$ and $\int_{b_{k+1:T} \in B^{T-k}} \prod_{i=k+1}^{T} P_s(b_i|b_{i-1},\pi,a_i) db_{k+1:T}$ both equal one, their difference integrates to zero, and subtracting $\frac{1}{2}$ from the indicator leaves the integral unchanged while $|1_{R_{k:T} \le l} - \frac{1}{2}| = \frac{1}{2}$ halves the bound,

$$\begin{aligned} &|\mathbb{P}^{\pi}_{b_k}[R_{k:T} \le l] - \mathbb{P}^{\pi,s}_{b_k}[R_{k:T} \le l]| \\ &= |\int_{b_{k+1:T} \in B^{T-k}} (1_{R_{k:T} \le l} - \tfrac{1}{2})[\prod_{i=k+1}^{T} P(b_i|b_{i-1},\pi,a_i) - \prod_{i=k+1}^{T} P_s(b_i|b_{i-1},\pi,a_i)] db_{k+1:T}| \\ &\le \tfrac{1}{2} \int_{b_{k+1:T} \in B^{T-k}} |\prod_{i=k+1}^{T} P(b_i|b_{i-1},\pi,a_i) - \prod_{i=k+1}^{T} P_s(b_i|b_{i-1},\pi,a_i)| db_{k+1:T} = \tfrac{1}{2} g(T), \end{aligned} \tag{140}$$

where

$$g(t) \triangleq \int_{b_{k+1:t} \in B^{t-k+1}} |\prod_{i=k+1}^{t} P(b_i|b_{i-1},\pi,a_i) - \prod_{i=k+1}^{t} P_s(b_i|b_{i-1},\pi,a_i)| db_{k+1:t}. \tag{141}$$

We will prove that

$$g(t+1) \le g(t) + 2\mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_t,a_t)] = \epsilon_t \tag{142}$$

for $t \in \{k+1,\dots,T\}$, where $\epsilon_k = 0$, and therefore, by (140),

$$|\mathbb{P}^{\pi}_{b_k}[R_{k:t} \le l] - \mathbb{P}^{\pi,s}_{b_k}[R_{k:t} \le l]| \le \tfrac{1}{2}\epsilon_t. \tag{143}$$

For the case where $t = k$, b_t is constant, and therefore the $c(b_t,a_t)$ does not depend on the belief transition model. Hence, $\mathbb{P}^{\pi,s}_{b_t,a_t}[c(b_t,a_t) \le l] = \mathbb{P}^{\pi}_{b_t,a_t}[c(b_t,a_t) \le l]$. Therefore we get

$$\mathbb{P}^{\pi}_{b_k,a_k}[c(b_k,a_k) \le l] - \mathbb{P}^{\pi,s}_{b_k,a_k}[c(b_k,a_k) \le l]| = 0 = \epsilon_k. \tag{144}$$

The rest of the proof will use induction over $t \in \{k+1,\dots,T\}$.

Base case, $t = k+1$**:**

$$\begin{aligned} g(k+1) &= \int_{b_{k+1} \in B} |\prod_{i=k+1}^{k+1} P(b_i|b_{i-1},\pi,a_i) - \prod_{i=k+1}^{k+1} P_s(b_i|b_{i-1},\pi,a_i)| db_{k+1:k+1} \\ &= \int_{b_{k+1} \in B} |P(b_{k+1}|b_k,\pi,a_k) - P_s(b_{k+1}|b_k,\pi,a_k)| db_{k+1} = 2\Delta^s(b_k,a_k) = 2\mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_k,a_k)] \\ &= \epsilon_{k+1} \end{aligned} \tag{145}$$

Induction step: Assume that the claim is true for $t \in \{k+1,\dots,T-1\}$ and prove for $t+1$.

Our proof's strategy is to express $g(t+1)$ recursively, using $g(t)$, and separate the bound at time $t+1$ into 2 components - one at time t and one at time $t+1$. We then bound the term at time $t+1$, and use the induction step to bound the recursive part of g at time t.

For all $x_i, y_i \in \mathbb{R}$,

$$\begin{aligned}
&\prod_{i=k+1}^{t+1} x_i - \prod_{i=k+1}^{t+1} y_i \\
&= \prod_{i=k+1}^{t+1} x_i - x_{t+1}\prod_{i=k+1}^{t} y_i + x_{t+1}\prod_{i=k+1}^{t} y_i - \prod_{i=k+1}^{t+1} y_i \\
&= x_{t+1}\Big(\prod_{i=k+1}^{t} x_i - \prod_{i=k+1}^{t} y_i\Big) + (x_{t+1} - y_{t+1})\prod_{i=k+1}^{t} y_i.
\end{aligned} \tag{146}$$

By denoting $x_i = P(b_i|b_{i-1}, a_{i-1})$ and $y_i = P_s(b_i|b_{i-1}, a_{i-1})$ we get

$$\begin{aligned}
&\prod_{i=k+1}^{t+1} P(b_i|b_{i-1}, \pi, a_i) - \prod_{i=k+1}^{t+1} P_s(b_i|b_{i-1}, \pi, a_i) \\
&= [P(b_{t+1}|b_t, a_t) - P_s(b_{t+1}|b_t, a_t)] \prod_{i=k+1}^{t} P_s(b_i|b_{i-1}, a_{i-1}) \\
&\quad + P(b_{t+1}|b_t, a_t)\Big[\prod_{i=k+1}^{t} P(b_i|b_{i-1}, \pi, a_i) - \prod_{i=k+1}^{t} P_s(b_i|b_{i-1}, \pi, a_i)\Big].
\end{aligned} \tag{147}$$

By combining the equation above with the triangle inequality, we get

$$g(t+1) \le A_1 + A_2 \tag{148}$$

for

$$A_1 \triangleq \int_{b_{k+1:t+1} \in B^{t-k}} |P(b_{t+1}|b_t, a_t) - P_s(b_{t+1}|b_t, a_t)| \prod_{i=k+1}^{t} P_s(b_i|b_{i-1}, a_{i-1}) db_{k+1:t}, \tag{149}$$

$$A_2 \triangleq \int_{b_{k+1:t+1} \in B^{t-k}} P(b_{t+1}|b_t, a_t)| \prod_{i=k+1}^{t} P(b_i|b_{i-1}, \pi, a_i) - \prod_{i=k+1}^{t} P_s(b_i|b_{i-1}, \pi, a_i)| db_{k+1:t} \tag{150}$$

A_1 is twice the expectation over the TV-distance of the belief at time t, since Δ^s carries the factor $\frac{1}{2}$ of (41).

$$A_1 = 2\mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_t, a_t)]. \tag{151}$$

As we see below, A_2 does not depend on the belief integration at time $t+1$, and therefore equals to $g(t)$. From the induction assumption, we get that $g(t) \le \epsilon_t$.

$$\begin{aligned}
A_2 &= \int_{b_{k+1:t} \in B^{t-k}} P(b_{t+1}|b_t, a_t)| \prod_{i=k+1}^{t} P(b_i|b_{i-1}, \pi, a_i) - \prod_{i=k+1}^{t} P_s(b_i|b_{i-1}, \pi, a_i)| db_{k:t+1} \\
&= \int_{b_{k+1:t} \in B^{t-k}} |\prod_{i=k+1}^{t} P(b_i|b_{i-1}, \pi, a_i) - \prod_{i=k+1}^{t} P_s(b_i|b_{i-1}, \pi, a_i)| \underbrace{\int_{b_{t+1}} P(b_{t+1}|b_t, a_t) db_{t+1}}_{=1} db_{k+1:t} \\
&= g(t) \le \epsilon_t
\end{aligned} \tag{152}$$

By combining the computations of A_1 and A_2,

$$g(t+1) \le A_1 + A_2 = g(t) + 2\mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_t, a_t)]. \tag{153}$$

The last equation completes the induction proof. Utilizing the recursive relation of ϵ_t together with (140), whose factor $\frac{1}{2}$ cancels the factor 2 that g carries relative to Δ^s, we get the bound we want to prove.

$$\begin{aligned}
&|\mathbb{P}^{\pi}_{b_k}[R_{k:T} \le l] - \mathbb{P}^{\pi,s}_{b_k}[R_{k:T} \le l]| \le \tfrac{1}{2} g(T) \\
&\le \tfrac{1}{2} g(T-1) + \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_{T-1}, a_{T-1})] \\
&\le \sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_t, a_t)].
\end{aligned} \tag{154}$$

□

Theorem 6.2 (restated). *Let $d_{min}, d_{max} \in \mathbb{R}$ be deterministic bounds such that $d_{min} \le R_{k:T} \le d_{max}$ almost surely under both $\mathbb{P}^{\pi}_{b_k,a_k}$ and $\mathbb{P}^{\pi,s}_{b_k,a_k}$, and denote*

$$\epsilon \triangleq \epsilon(b_k, a_k) \triangleq \min\Big(\sum_{i=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_i, a_i)], 1\Big), \tag{155}$$

(1) ***Upper Bound:***
   (a) *If $\epsilon < \alpha$, then $U_s \triangleq (1 - \frac{\epsilon}{\alpha}) Q^{\pi}_{M_s}(b_k, a_k, \alpha - \epsilon) + \frac{\epsilon}{\alpha} d_{max}$*
   (b) *If $\epsilon \ge \alpha$ then $U_s \triangleq d_{max}$*

(2) ***Lower Bound:***
   (a) *If $\epsilon + \alpha < 1$ then $L_s \triangleq (1 + \frac{\epsilon}{\alpha}) Q^{\pi}_{M_s}(b_k, a_k, \alpha + \epsilon) - \frac{\epsilon}{\alpha} Q^{\pi}_{M_s}(b_k, a_k, \epsilon)$*
   (b) *If $\epsilon + \alpha \ge 1$ then $L_s \triangleq \frac{1}{\alpha}[(\alpha + \epsilon - 1) d_{min} + \mathbb{E}^{\pi,s}_{b_k,a_k}[R_{k:T}] - \epsilon Q^{\pi}_{M_s}(b_k, a_k, \epsilon)]$*

where $\mathbb{E}^{\pi,s}_{b_k,a_k}[R_{k:T}] = Q^{\pi}_{M_s}(b_k, a_k, 1)$ denotes the expected return under the simplified model. For $\epsilon = 0$, interpret $\epsilon Q^{\pi}_{M_s}(b_k, a_k, \epsilon)$ as zero. Then $L_s \le Q^{\pi}_{M}(b_k, a_k, \alpha) \le U_s$ deterministically.

Proof. From Theorem 6.1,

$$\sup_{l \in \mathbb{R}} |\mathbb{P}^{\pi}_{b_k,a_k}[R_{k:T} \le l] - \mathbb{P}^{\pi,s}_{b_k,a_k}[R_{k:T} \le l]| \le \min\left\{\sum_{i=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_i, a_i)], 1\right\} = \epsilon, \tag{156}$$

The action-value function is defined as the CVaR of the return. Let X and Y be random variables distributed according to the original and simplified models P and P_s, respectively. Theorem 5.1 establishes bounds on $\mathrm{CVaR}_{\alpha}(X)$ in terms of CVaR of Y, under a known discrepancy between the cumulative distribution functions of X and Y. For the fixed initial pair (b_k, a_k), this distributional discrepancy is bounded by ϵ, and the deterministic endpoints satisfy the support assumptions with $a_X = a_Y = d_{min}$ and $b_X = b_Y = d_{max}$, so the result applies directly. □

Theorem 6.3 (restated). *Both models update the belief with the same kernel given the new observation,*

$$P(b_{t+1}|b_t, a_t, z_{t+1}) = P_s(b_{t+1}|b_t, a_t, z_{t+1}), \tag{157}$$

and differ only in the observation model: p_z under M, q_z under M_s. In the case where the simplified and original observation models are denoted by q_z and p_z respectively, it holds that

$$\sup_{l \in \mathbb{R}} |\mathbb{P}^{\pi}_{b_k}[R_{k:T} \le l] - \mathbb{P}^{\pi,s}_{b_k}[R_{k:T} \le l]| \le \sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s_x(x_{t+1})], \tag{158}$$

where Δ_x^s is the state-level TV distance defined by

$$\Delta_x^s(x_t) \triangleq \tfrac{1}{2}\int_{z_t\in Z} |p_z(z_t|x_t) - q_z(z_t|x_t)|dz_t, \tag{159}$$

and x_{t+1} is drawn by $x_t \sim b_t$, $x_{t+1} \sim P(\cdot|x_t, a_t)$, with (b_t, a_t) from the simplified trajectory started at (b_k, a_k).

Proof.

$$\begin{aligned}
&\sup_{l\in\mathbb{R}} |\mathbb{P}^{\pi}_{b_k,a_k}[R_{k:T} \le l] - \mathbb{P}^{\pi,s}_{b_k,a_k}[R_{k:T} \le l]| \overset{(a)}{\le} \sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta^s(b_t, a_t)] \\
&\overset{(b)}{=} \sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}\Big[\tfrac{1}{2}\int_B |P(b_{t+1}|b_t, a_t) - P_s(b_{t+1}|b_t, a_t)|\, db_{t+1}\Big] \\
&\overset{(c)}{=} \sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}\Big[\tfrac{1}{2}\int_B \Big|\int_X\int_X\int_Z P(b_{t+1}|b_t, a_t, z_{t+1})\big(p_z(z_{t+1}|x_{t+1}) - q_z(z_{t+1}|x_{t+1})\big) \\
&\qquad\qquad \cdot P(x_{t+1}|x_t, a_t)\, b_t(x_t)\, dz_{t+1}dx_{t+1}dx_t\Big| db_{t+1}\Big] \\
&\overset{(d)}{\le} \sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}\Big[\int_X\int_X b_t(x_t)P(x_{t+1}|x_t, a_t)\, \tfrac{1}{2}\int_Z |p_z(z_{t+1}|x_{t+1}) - q_z(z_{t+1}|x_{t+1})| \\
&\qquad\qquad \cdot \underbrace{\int_B P(b_{t+1}|b_t, a_t, z_{t+1})\, db_{t+1}}_{=1}\; dz_{t+1}dx_{t+1}dx_t\Big] \\
&\overset{(e)}{=} \sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}\Big[\int_X\int_X b_t(x_t)P(x_{t+1}|x_t, a_t)\, \Delta_x^s(x_{t+1})\, dx_{t+1}dx_t\Big] \overset{(f)}{=} \sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta_x^s(x_{t+1})].
\end{aligned} \tag{160}$$

(a) Theorem 6.1, eq. (40). (b) Definition of Δ^s, eq. (41). (c) Expand each belief kernel over (x_t, x_{t+1}, z_{t+1}); by the assumption $P(b_{t+1}|b_t, a_t, z_{t+1}) = P_s(b_{t+1}|b_t, a_t, z_{t+1})$ of the theorem the two expansions share this kernel and differ only in p_z against q_z. (d) Triangle inequality, then exchange the order of integration. (e) The inner integral is one; definition (159). (f) Tower property, with x_{t+1} drawn as in the statement. □

Corollary E.1. *The bound from Theorem 6.2 holds for $\epsilon(b_k, a_k) = \min(\sum_{t=k}^{T-1} \mathbb{E}^{\pi,s}_{b_k,a_k}[\Delta_x^s(x_{t+1})], 1)$.*

Proof. Theorem 6.2 establishes bounds on the value and action-value functions under a quantified distributional discrepancy ϵ between the original and simplified return distributions. As long as this discrepancy between their CDFs remains bounded by ϵ, the result remains valid. Consequently, by bounding the belief-based discrepancy ϵ in Theorem 6.2 via the state-based discrepancy bound established in Theorem 6.3, we recover the same form of the bound for the newly defined ϵ. □

Theorem 7.1 (restated). *Let $S \triangleq \{(x_n^\Delta, \hat{\Delta}_x^s(x_n^\Delta))\}_{n=1}^{N^\Delta}$ denote the offline sample, with $\hat{\Delta}_x^s$ unbiased for Δ_x^s given x_n^Δ. Assume $\bar{b}_t$ and a are fixed independently of S, the online trajectories $\{\bar{b}_\tau^i\}_{i=1}^{N_b}$ are i.i.d. and independent of S, and the pairs $(x_n^\Delta, \hat{\Delta}_x^s(x_n^\Delta))_{n=1}^{N^\Delta}$ are i.i.d. If $T > t$, then for any $u, v > 0$,*

$$\mathbb{P}^{\pi,s}_{\bar{b}_t,a}[|\hat{\epsilon}_t(\bar{b}_t, a) - \epsilon_t(\bar{b}_t, a)| \ge u + v] \le 2\exp\Big(-\frac{2N_b u^2}{D^2(T-t)^2}\Big) + 2\exp\Big(-\frac{2N^\Delta v^2}{D^2(T-t)^2}\Big), \tag{161}$$

for $D = \max_{i\in\{t,\dots,T-1\}} D_i$ where $D_i = \sup_{x_{i+1}} P(x_{i+1}|x_i, a_i)/Q_0(x_{i+1})$. If $T = t$ the failure probability is zero.

Proof. Let $S \triangleq \{(x_n^\Delta, \hat{\Delta}_x^s(x_n^\Delta))\}_{n=1}^{N^\Delta}$ and

$$\tilde{\epsilon}_t(S) \triangleq \sum_{\tau=t}^{T-1} \mathbb{E}\big[\hat{m}(\bar{b}_\tau, a_\tau)\big], \tag{162}$$

with the expectation taken over the trajectory and S held fixed, the estimated distributional discrepancy given the sample S. By the triangle inequality, $|\hat{\epsilon}_t - \epsilon_t| \le |\hat{\epsilon}_t - \tilde{\epsilon}_t(S)| + |\tilde{\epsilon}_t(S) - \epsilon_t|$, so

$$\mathbb{P}^{\pi,s}_{\bar{b}_t,a}[|\hat{\epsilon}_t(\bar{b}_t,a) - \epsilon_t(\bar{b}_t,a)| \ge u+v] \le \underbrace{\mathbb{E}\,\mathbb{P}^{\pi,s}_{\bar{b}_t,a}[|\hat{\epsilon}_t - \tilde{\epsilon}_t(S)| \ge u \mid S]}_{\text{(I) online error}} + \underbrace{\mathbb{P}[|\tilde{\epsilon}_t(S) - \epsilon_t| \ge v]}_{\text{(II) offline error}}. \tag{163}$$

Bounding (I). Define

$$X_i \triangleq \frac{1}{N_b} \sum_{\tau=t}^{T-1} \hat{m}(\bar{b}_\tau^i, \pi(\bar{b}_\tau^i)), \tag{164}$$

and note that

$$\sum_{i=1}^{N_b} X_i = \hat{\epsilon}_t(\bar{b}_t, a). \tag{165}$$

Each X_i reuses the offline sample S through $\hat{m}$, so the X_i are independent only conditionally on S. Given S, the online trajectories $\{\bar{b}_\tau^i\}_{i=1}^{N_b}$ are i.i.d. by assumption, hence so are the X_i, with

$$0 \le X_i \le \frac{D(T-t)}{N_b}. \tag{166}$$

Write $a_\tau^i \triangleq \pi(\bar{b}_\tau^i)$. Given S, the trajectories are i.i.d. and independent of S, so

$$\begin{aligned}\mathbb{E}[\hat{\epsilon}_t(\bar{b}_t,a) \mid S] &\overset{(52)}{=} \sum_{\tau=t}^{T-1} \mathbb{E}\Big[\hat{\mathbb{E}}^{\pi,s}_{\bar{b}_t,a}[m(\bar{b}_\tau,\pi_\tau)]\,\Big|\,S\Big] \overset{(51)}{=} \sum_{\tau=t}^{T-1} \frac{1}{N_b} \sum_{i=1}^{N_b} \mathbb{E}\big[\hat{m}(\bar{b}_\tau^i, a_\tau^i)\,\big|\,S\big] \\ &= \sum_{\tau=t}^{T-1} \mathbb{E}\big[\hat{m}(\bar{b}_\tau, a_\tau)\big] \overset{(162)}{=} \tilde{\epsilon}_t(S),\end{aligned} \tag{167}$$

where the third equality holds because every i gives the same conditional expectation, taken over one trajectory with S held fixed. By (165), $\sum_i \mathbb{E}[X_i \mid S] = \mathbb{E}[\hat{\epsilon}_t(\bar{b}_t,a) \mid S] = \tilde{\epsilon}_t(S)$, so Hoeffding's inequality on $\sum_i (X_i - \mathbb{E}[X_i \mid S])$ gives

$$\mathbb{P}^{\pi,s}_{\bar{b}_t,a}[|\hat{\epsilon}_t(\bar{b}_t,a) - \tilde{\epsilon}_t(S)| \ge u \mid S] \le 2\exp\Big(-\frac{2N_b u^2}{D^2(T-t)^2}\Big). \tag{168}$$

Bounding (II). For the offline error, define

$$\rho(x) \triangleq \sum_{\tau=t}^{T-1} \mathbb{E}^{\pi,s}_{\bar{b}_t,a}\Big[\sum_j \tilde{w}_\tau^j \frac{P(x|x_\tau^j, a_\tau)}{Q_0(x)}\Big], \qquad V_n \triangleq \rho(x_n^\Delta)\,\hat{\Delta}_x^s(x_n^\Delta). \tag{169}$$

The expectation in ρ is over the online particles x_τ^j, weights $\tilde{w}_\tau^j$, and actions a_τ, so all online data is integrated out of ρ: it is a deterministic function of x, and V_n depends only on the offline pair $(x_n^\Delta, \hat{\Delta}_x^s(x_n^\Delta))$. Since the pairs

$(x_n^\Delta, \hat{\Delta}_x^s(x_n^\Delta))_{n=1}^{N^\Delta}$ are i.i.d., the V_n are independent, with $0 \le V_n \le D(T-t)$. Since $\mathbb{E}[\hat{\Delta}_x^s(x_n^\Delta) \mid x_n^\Delta] = \Delta_x^s(x_n^\Delta)$,

$$\begin{aligned}\mathbb{E}[V_n] &= \mathbb{E}_{x\sim Q_0}\big[\rho(x)\,\mathbb{E}[\hat{\Delta}_x^s(x) \mid x]\big] = \mathbb{E}_{x\sim Q_0}\big[\rho(x)\,\Delta_x^s(x)\big] \\ &= \sum_{\tau=t}^{T-1}\mathbb{E}\Big[\sum_j \tilde{w}_\tau^j\,\mathbb{E}_{x\sim Q_0}\Big[\frac{P(x|x_\tau^j,a_\tau)}{Q_0(x)}\Delta_x^s(x)\Big]\Big] \\ &\overset{(46)}{=} \sum_{\tau=t}^{T-1}\mathbb{E}\Big[\sum_j \tilde{w}_\tau^j\, m(x_\tau^j,a_\tau)\Big] \overset{(47)}{=} \sum_{\tau=t}^{T-1}\mathbb{E}[m(\bar{b}_\tau,a_\tau)] = \epsilon_t(\bar{b}_t,a),\end{aligned} \tag{170}$$

and, writing $\bar{b}_\tau = \{(x_\tau^j, \tilde{w}_\tau^j)\}_{j=1}^{N_p}$ and using that $\hat{\Delta}_x^s(x_n^\Delta)$ does not depend on the trajectory,

$$\begin{aligned}\tilde{\epsilon}_t(S) &\overset{(162)}{=} \sum_{\tau=t}^{T-1}\mathbb{E}\big[\hat{m}(\bar{b}_\tau,a_\tau)\big] \overset{(50)}{=} \sum_{\tau=t}^{T-1}\mathbb{E}\Big[\sum_j \tilde{w}_\tau^j\,\hat{m}(x_\tau^j,a_\tau)\Big] \\ &\overset{(49)}{=} \sum_{\tau=t}^{T-1}\mathbb{E}\Big[\sum_j \tilde{w}_\tau^j \frac{1}{N^\Delta}\sum_{n=1}^{N^\Delta}\frac{P(x_n^\Delta|x_\tau^j,a_\tau)}{Q_0(x_n^\Delta)}\hat{\Delta}_x^s(x_n^\Delta)\Big] \\ &= \frac{1}{N^\Delta}\sum_{n=1}^{N^\Delta}\rho(x_n^\Delta)\,\hat{\Delta}_x^s(x_n^\Delta) \overset{(169)}{=} \frac{1}{N^\Delta}\sum_{n=1}^{N^\Delta} V_n,\end{aligned} \tag{171}$$

so Hoeffding's inequality gives

$$\mathbb{P}[|\tilde{\epsilon}_t(S) - \epsilon_t(\bar{b}_t,a)| \ge v] \le 2\exp\Big(-\frac{2N^\Delta v^2}{D^2(T-t)^2}\Big). \tag{172}$$

Combining. Substituting the two bounds into (163),

$$\mathbb{P}^{\pi,s}_{\bar{b}_t,a}[|\hat{\epsilon}_t(\bar{b}_t,a) - \epsilon_t(\bar{b}_t,a)| \ge u+v] \le 2\exp\Big(-\frac{2N_b u^2}{D^2(T-t)^2}\Big) + 2\exp\Big(-\frac{2N^\Delta v^2}{D^2(T-t)^2}\Big). \tag{173}$$

If $T = t$ every sum above is empty and $\hat{\epsilon}_t = \epsilon_t = 0$. □

Theorem 7.2 (restated). *Let b_k be an initial belief and $R^1,\dots,R^{N_b} \sim \mathbb{P}^{\pi,s}_{b_k}$. Let $\hat{\epsilon} \triangleq \hat{\epsilon}(b_k,a_k)$ be the importance-sampling estimator from (52), and assume that $\mathbb{P}[\hat{\epsilon} \ge \epsilon] \ge 1-\delta_3$ where ϵ is the true distributional discrepancy as defined in (42). Let $\delta_1,\delta_2,\delta_3 \in (0,0.5)$, and let $\eta \ge 0$. Define $\hat{\epsilon}' \triangleq \min(\hat{\epsilon}+\eta,1)$, $\hat{p} \triangleq 1-\delta_2^{1/N_b}+\hat{\epsilon}$, $\hat{q} \triangleq 1-\delta_2^{1/N_b}+\hat{\epsilon}$, and let $\bar{d}_{\max} \triangleq R_{\max}\sum_{i=0}^{T-k}\gamma^i$, $\underline{d}_{\min} \triangleq R_{\min}\sum_{i=0}^{T-k}\gamma^i$.*

(1) ***Upper Bound:*** *Assume $\mathbb{P}[\sup_z(\hat{F}_{R_s}(z) - F_{R_s}(z)) \le \eta] > 1-\delta_1$. If $\hat{\epsilon}' < \alpha$,*

$$U_s^{\mathrm{tr}} \triangleq \Big(1-\frac{\hat{\epsilon}'}{\alpha}\Big)\hat{Q}^\pi_{M_s}(b_k,a_k,\alpha-\hat{\epsilon}') + \frac{\hat{\epsilon}'}{\alpha}\hat{d}^\pi_{\max}(b_k) + \frac{\hat{p}}{\alpha}\big(\bar{d}_{\max} - \hat{d}^\pi_{\max}(b_k)\big)^+. \tag{174}$$

If $\hat{\epsilon}' \ge \alpha$, define $U_s^{\mathrm{tr}} \triangleq \hat{d}^\pi_{\max}(b_k) + \frac{\hat{p}}{\alpha}(\bar{d}_{\max} - \hat{d}^\pi_{\max}(b_k))^+$.

(2) ***Lower Bound:*** *Assume $\mathbb{P}[\sup_z(F_{R_s}(z) - \hat{F}_{R_s}(z)) \le \eta] > 1-\delta_1$. If $\hat{\epsilon}' + \alpha < 1$ and the CDF of $R_{k:T}$ under $\mathbb{P}^{\pi,s}_{b_k}$ is continuous,*

$$L_s^{\mathrm{tr}} \triangleq \Big(1+\frac{\hat{\epsilon}'}{\alpha}\Big)\hat{Q}^\pi_{M_s}(b_k,a_k,\alpha+\hat{\epsilon}') - \frac{\hat{\epsilon}'}{\alpha}\hat{Q}^\pi_{M_s}(b_k,a_k,\hat{\epsilon}') - \frac{\hat{q}}{\alpha}\big(\hat{d}^\pi_{\min}(b_k) - \underline{d}_{\min}\big)^+. \tag{175}$$

Then $\mathbb{P}^\pi_{b_k,a_k}[Q^\pi_M(b_k,a_k,\alpha) \le U_s^{\mathrm{tr}}] > 1-\delta_1-\delta_2-\delta_3$ and $\mathbb{P}^\pi_{b_k,a_k}[L_s^{\mathrm{tr}} \le Q^\pi_M(b_k,a_k,\alpha)] > 1-\delta_1-\delta_2-\delta_3$.

Proof of Theorem 7.2. This follows from Theorem 5.8 applied to the POMDP simplification setting. Let X denote the return under the original model P and Y the return under the simplified model P_s. From Theorem 6.1, and because $\sup_l |F_X(l) - F_Y(l)| \le 1$, the one-sided discrepancies satisfy $\max\big(\sup_l(F_Y(l) - F_X(l)), \sup_l(F_X(l) - F_Y(l))\big) \le \sup_l |F_X(l) - F_Y(l)| \le \epsilon$ with ϵ from (42). The estimator $\hat{\epsilon}$ from (52) satisfies $\mathbb{P}[\hat{\epsilon} \ge \epsilon] \ge 1 - \delta_3$ by assumption.

The return samples $R^1, \dots, R^{N_b} \sim \mathbb{P}^{\pi,s}_{b_k}$ are i.i.d. samples from F_Y. Setting $n = N_b$, $\bar{b} = \bar{d}_{\max}$, $\underline{a} = \underline{d}_{\min}$, $\hat{b}_Y = \hat{d}^{\pi}_{\max}(b_k)$, and $\hat{a}_Y = \hat{d}^{\pi}_{\min}(b_k)$ in Theorem 5.8, and noting that $\bar{d}_{\max} \triangleq R_{\max} \sum_{i=0}^{T-k} \gamma^i$ is a deterministic upper bound on the support, and $\underline{d}_{\min}$ a deterministic lower bound, because each term $\gamma^{t-k} c(b_t, a_t)$ of $R_{k:T}$ lies in $[\gamma^{t-k} R_{\min}, \gamma^{t-k} R_{\max}]$ since $\gamma^{t-k} \ge 0$, the result follows directly. □

F Deterministic CVaR Bounds via Auxiliary Distributions

The proof of Theorem 5.6 (Appendix D) relies on a deterministic result that bounds $\mathrm{CVaR}_\alpha(X)$ using the CDF of an auxiliary variable Y with known one-sided discrepancy. This deterministic result is equivalent to Theorem 5.1 (proved in Appendix B): both establish that a one-sided CDF discrepancy $\sup_z(F_Y(z) - F_X(z)) \le \epsilon$ yields the upper bound $\mathrm{CVaR}_\alpha(X) \le \frac{\epsilon}{\alpha} \max(b_X, b_Y) + (1 - \frac{\epsilon}{\alpha})\,\mathrm{CVaR}_{\alpha-\epsilon}(Y)$ when $\alpha > \epsilon$, and the corresponding lower bound under the reversed discrepancy condition. The full proof of the deterministic result is therefore subsumed by the proof of Theorem 5.1 and is not repeated here; the proof of Theorem 5.6 applies this deterministic bound with ϵ replaced by $\epsilon' = \epsilon + \eta$ to account for CDF estimation error.

G Optimal Confidence Split

For a total confidence budget $\delta = \delta_1 + \delta_2$, the bound width depends on δ_1 through η (rate $O(1/\sqrt{n})$ for all instantiations in Remark 5.9) and on δ_2 through $\hat{p} \approx \ln(1/\delta_2)/n$ (rate $O(1/n)$). We analyze the split for the global DKW instantiation $\eta_{\mathrm{DKW}} \triangleq \sqrt{\ln(1/\delta_1)/(2n)}$; the qualitative conclusion (allocate most budget to δ_1) holds for all choices. Equating the marginal costs $\partial U/\partial\delta_1 = \partial U/\partial\delta_2$ yields

$$\delta_2^* = \frac{2\delta\,(\bar{b} - \hat{b}_Y)^+}{\hat{b}_Y - \mathrm{CVaR}^{\hat{F}_Y}_{\alpha-\epsilon'}} \cdot \frac{\sqrt{2\ln(1/\delta)}}{\sqrt{n}} + O(1/n), \qquad \delta_1^* = \delta - \delta_2^*. \tag{176}$$

Since $\delta_2^* = \Theta(\delta/\sqrt{n})$, almost all the confidence budget should be allocated to the CDF concentration event for large n. The practical rule $\delta_2 = \min(\delta/\sqrt{n},\ \delta/2)$ and $\delta_1 = \delta - \delta_2$ is near-optimal and avoids computing data-dependent quantities before the split.

H Instantiations of η

Any CDF concentration inequality yields a valid η for Theorem 5.6. We highlight three choices, ordered from loosest to tightest:

(1) **Global DKW** (exact finite-sample). The one-sided DKW inequality (Dvoretzky et al. 1956; Massart 1990) gives

$$\eta_{\mathrm{DKW}} \triangleq \sqrt{\frac{\ln(1/\delta_1)}{2n}}. \tag{177}$$

(2) **Restricted KS** (asymptotic). Since the upper bound accesses quantiles only in $[1-\alpha, 1]$, we restrict the one-sided KS statistic to $\{z : F_Y(z) \ge 1 - \alpha\}$. By Donsker's theorem (Donsker 1952), the empirical process $\sqrt{n}(\hat{F}_n - F)$ converges weakly to a Brownian bridge B on $[0, 1]$. On the sub-interval $[1-\alpha, 1]$, the bridge variance $\mathrm{Var}(B(t)) = t(1-t) \le \alpha$ scales with α, so the one-sided supremum tail satisfies $\mathbb{P}[\sup_{t\in[1-\alpha,1]} B(t) > a] \lesssim 2e^{-a^2/(2\alpha)}$ (Shorack and Wellner 1982). Substituting $a = \sqrt{n}\,\eta$ and solving for η

yields

$$\eta_{\text{rKS}} \triangleq \sqrt{\frac{2\alpha \ln(2/\delta_1)}{n}}, \tag{178}$$

which is *asymptotically* valid (empirically accurate for $n \geq 50$, but not a rigorous finite-sample guarantee). The ratio $\eta_{\text{rKS}}/\eta_{\text{DKW}} \approx 2\sqrt{\alpha}$ (≈ 0.45 for $\alpha = 0.05$), giving 40–55% tighter bounds.

(3) **Local DKW** (exact finite-sample). The exact local CDF concentration of Maillard (2021) yields a parameter $\eta_{\text{lDKW}}(n, \delta_1)$ with a rigorous finite-sample guarantee; see Appendix I for the full derivation.

Remark H.1 (Hierarchy). *$\eta_{\text{lDKW}} \leq \eta_{\text{rKS}} \leq \eta_{\text{DKW}}$, with equality of the first two when $\alpha = 1$.*

Remark H.2 (Lower-bound η). *For lower bounds, the proof accesses quantile levels down to $1 - \alpha - \epsilon'$, which extends below $1 - \alpha$. The restricted region must be expanded to $[1 - \alpha - \epsilon, 1]$, yielding $\eta^L_{\text{rKS}} \triangleq \sqrt{2(\alpha + \epsilon)\ln(2/\delta_1)/n}$ and η^L_{lDKW} defined analogously (Appendix J). For single-variable bounds ($\epsilon = 0$), upper and lower parameters coincide.*

Remark H.3 (Optimal confidence split). *The near-optimal practical rule $\delta_2 = \min(\delta/\sqrt{n},\ \delta/2)$, $\delta_1 = \delta - \delta_2$ allocates most budget to CDF concentration (Appendix G).*

I Local DKW Concentration Parameter: Full Derivation

By the probability integral transform, the restricted one-sided KS statistic satisfies the distribution-free identity $D_n^{[1-\alpha]} \overset{\mathcal{L}}{=} \sup_{u \in [1-\alpha, 1]}(\hat{G}_n(u) - u)$, where $\hat{G}_n$ is the empirical CDF of n i.i.d. Uniform$(0, 1)$ samples. Maillard (2021, Corollary 6) gives the exact tail probability: for any $t \leq \alpha$,

$$\delta_{[1-\alpha,1]}(n, t) \triangleq \mathbb{P}\left(\sup_{u \in [1-\alpha,\, 1]} \hat{G}_n(u) - u > t\right) = \sum_{\ell=0}^{\lceil n(\alpha - t)\rceil - 1} t \left(\frac{\ell}{n} + t\right)^{\ell - 1} \binom{n}{\ell} \left(1 - t - \frac{\ell}{n}\right)^{n-\ell}. \tag{179}$$

This is an *equality* for every finite n. Inverting yields the exact local concentration parameter:

$$\eta_{\text{lDKW}}(n, \delta_1) \triangleq \inf\left\{t > 0 : \delta_{[1-\alpha,1]}(n, t) \leq \delta_1\right\}. \tag{180}$$

Since $\delta_{[1-\alpha,1]}$ is decreasing in t, this inversion is computed via bisection. The parameter η_{lDKW} satisfies the one-sided CDF concentration condition of Theorem 5.6 and is the tightest finite-sample instantiation listed in Section H. The local CDF concentration (179) is due to Maillard (2021); the translation to CVaR confidence intervals via truncation decomposition and confidence splitting is our contribution.

J Local Concentration for Lower Bounds

For *upper bounds*, the proof of Theorem 5.6 evaluates empirical quantiles at shifted levels $v + \epsilon' \in [1 - \alpha + \epsilon', 1]$, which lie entirely within $[1 - \alpha, 1]$. The restricted KS and local DKW parameters from Appendix I, computed on $[1 - \alpha, 1]$, therefore apply directly.

For *lower bounds*, the proof evaluates empirical quantiles at levels $v - \epsilon' \in [1 - \alpha - \epsilon', 1 - \epsilon']$, which extend *below* $1 - \alpha$ by ϵ'. The restricted KS and local DKW parameters above do not control the CDF deviation in this extended range. To obtain valid local concentration for lower bounds, the restricted region must be expanded to $[1 - \alpha - \epsilon, 1]$, yielding:

$$\eta^L_{\text{rKS}} \triangleq \sqrt{\frac{2(\alpha + \epsilon)\ln(2/\delta_1)}{n}}, \qquad \eta^L_{\text{lDKW}}(n, \delta_1) \triangleq \inf\left\{t > 0 : \delta_{[1-\alpha-\epsilon,\,1]}(n, t) \leq \delta_1\right\}, \tag{181}$$

where $\delta_{[1-\alpha-\epsilon,\,1]}$ is defined analogously to (179) with α replaced by $\alpha + \epsilon$.

These parameters satisfy $\eta_{\text{rKS}} \leq \eta^L_{\text{rKS}} \leq \eta_{\text{DKW}}$ and $\eta_{\text{lDKW}} \leq \eta^L_{\text{lDKW}} \leq \eta_{\text{DKW}}$, with equality when $\epsilon = 0$ (single-variable case).

K Environment Configurations

The following subsections describe the three POMDP environments used throughout the experiments, including the action elimination experiments of Figure 8. Each domain provides an *original* environment (used for sampling rollouts) and a *simplified* surrogate (used for belief updates and CVaR bound computation). The two models differ only in their observation models, as summarised in Table 1. In every environment the original observation noise is a Gaussian scale mixture: K components with a common mean of zero, equal weights, and standard deviations $\sigma_k = e^{\xi_k}\sigma_{\text{obs}}$ with $\xi_k \sim \mathcal{N}(0, \tau^2)$ drawn once with seed 42. The simplified model is the single Gaussian $\mathcal{N}(0, \sigma^2_{\text{obs}})$. The dispersion τ sets the discrepancy between the two models; it was chosen per environment so that the estimated ϵ/α of Table 6 lies between 3% and 17%, the regime in which the bounds are informative but the simplification is not free.

Table 1. Architectural differences between the original and simplified models across all domains.

Aspect	**Original**	**Simplified**
Observation noise	Gaussian Mixture Model (GMM)	Isotropic Gaussian

K.1 Light-Dark

State space: $\mathbf{s} = (x, y) \in \mathbb{R}^2$ (2D continuous position). **Action space**: Discrete — {up, down, left, right}, mapped to unit vectors in $\mathbb{R}^2$. **Observation space**: $\mathbf{o} \in \mathbb{R}^2$ — noisy 2D position.

Table 2. Configuration for the 2D Light-Dark POMDP.

Parameter	**Value**
Grid size	7×7
Start state	$(1,\ 1)$
Goal state	$(6,\ 6)$; radius 1.5
State transition covariance	$0.06 \cdot I_2$
Obs. noise (far from beacon)	$\mathcal{N}(0,\ 0.06 \cdot I_2)$ / GMM(K=2500)
Obs. noise (near beacon)	$\mathcal{N}(0,\ 0.03 \cdot I_2)$ / GMM(K=2500, scaled)
Beacons	$(1,1),\ (1,6),\ (6,1),\ (6,6)$; radius 1.0
Obstacle	$(5,2)$; radius 3.0; hit prob 1.0
Goal reward	+10
Obstacle penalty	−10
Step cost (fuel)	−2
Discount factor γ	0.95
Planning horizon H	10 steps in Figure 8; 9 otherwise

Original: Two independent 1D scale mixtures (one per axis), $K = 2500$ components, $\tau = 0.0347$. **Simplified**: Standard Gaussian $\mathcal{N}(0,\ \sigma^2 I_2)$ with the same σ.

K.2 Continuous Push

State space: $\mathbf{s} = (r_x, r_y, o_x, o_y, t_x, t_y) \in \mathbb{R}^6$ — robot, object, and target positions. **Action space**: Discrete — {up, down, left, right}, mapped to unit vectors. **Observation space**: $\mathbf{o} \in \mathbb{R}^6$ — robot and target positions observed exactly; object position is noisy: $\hat{o} = o + \epsilon,\ \epsilon \sim \mathcal{N}(0, \sigma^2 I_2)$.

Original: One 1D scale mixture ($K = 7000$ components, $\tau = 0.0344$) applied independently to each object-position axis. **Simplified**: Gaussian noise $\mathcal{N}(0,\ 0.01 \cdot I_2)$ on object position.

Table 3. Configuration for the Continuous Push POMDP.

Parameter	Value
Grid size	6×6
Initial state	$(0.5,\ 0.5,\ 1.0,\ 0.5,\ 5.0,\ 5.0)$
Robot transition covariance	$\mathrm{diag}(0.015,\ 0.015)$
Observation noise σ	0.1 (on object position)
Push threshold	1.0
Friction coefficient	0.3
Max push force	2.0
Robot radius	0.3
Dangerous area	$(5.0,\ 2.0)$; radius 1.0; penalty -100
Discount factor γ	0.95
Planning horizon H	9 steps

K.3 Continuous Laser Tag

State space: $\mathbf{s} = (r_x, r_y, o_x, o_y, \text{term}) \in \mathbb{R}^4 \times \{0, 1\}$ — robot and opponent positions plus terminal flag. **Action space**: Discrete — {up, down, left, right, tag}. **Observation space**: $\mathbf{o} \in \mathbb{R}^8$ — 8-directional laser range measurements with additive noise; terminal states return $\mathbf{o} = -\mathbf{1}$.

Table 4. Configuration for the Continuous Laser Tag POMDP.

Parameter	Value
Arena	11×7 continuous grid
Walls	8 axis-aligned wall cells
Initial state	$(2.0,\ 1.0,\ 8.0,\ 5.0,\ 0)$
Robot transition covariance	$0.0125 \cdot I_2$
Opponent transition covariance	$0.00625 \cdot I_2$
Opponent pursuit speed	0.6
Laser noise σ	1.0 (per measurement)
Robot / opponent radius	0.3
Tag radius	0.5
Tag reward	+10
Failed-tag penalty	-10
Step cost	-1
Dangerous areas	$(5.0,\ 3.0)$, $(7.0,\ 1.0)$; radius 2.0; penalty -150
Discount factor γ	0.95
Planning horizon H	8 steps (7 moves + tag)

Original: One 1D scale mixture with $K = 500$ components ($\tau = 0.0597$), applied independently to all 8 laser measurements. **Simplified**: Gaussian noise $\mathcal{N}(0,\ 1.0)$ per laser measurement.

K.4 Shared Planner Parameters

Table 5 lists the planner parameters that are common to the three environments.

The configurations in this appendix are those of the open-loop experiments in Figures 8, 9, and 10. Figure 8 evaluates the full action sequences (10 steps in Light-Dark, 9 in Push, 8 in Laser Tag); Figure 10 their prefixes of length 2 to 8 and, in the bottom row, the full sequences at depth 9 (8 for Laser Tag); Figure 9 horizons 4 to

Table 5. Planner parameters shared across all three environments. All experiments use $\alpha = 0.2$.

Parameter	Value
CVaR level α	0.2
Belief particles N_p	10
Return samples N_b	600
Δ-estimation states	100
Samples per state (for Δ)	2000
k-NN neighbours	10
Discount factor γ	0.95
Confidence parameter δ	0.05
Random seed	42

20. The support endpoints of the bounds are the environment's per-step reward range times the horizon. The lower endpoint assumes that every step is paid, which the reward model does not guarantee once a trajectory terminates; every return sample in the paper lies far above it, so no reported bound is affected.

Figure 8 presents the estimated bounds and CVaR values for pairs of action sequences across all three environments, computed using 600 sampled belief trajectories and a confidence parameter of $\delta = 0.05$.

The simplified planner samples $N^\Delta = 100$ states offline from Q_0, the uniform distribution over the state region of each environment (Section 9.3). For each sampled state, the distributional discrepancy Δ between the simplified and original observation models is estimated using $N_z = 2000$ samples. CVaR is computed at confidence level $\alpha = 0.2$, and the probabilistic bounds are constructed with confidence level $\delta = 0.05$. The confidence budget is split as $\delta_2 = \min(\delta/\sqrt{N_b},\ \delta/2)$ and $\delta_1 = (\delta - 2\delta_2)/2$, so that the two-sided interval spends $2\delta_1 + 2\delta_2 = \delta$ (see Appendix G for the rationale). No share of δ is reserved for δ_3, the failure probability of the discrepancy estimate; the plotted intervals therefore hold with probability $1 - \delta - \delta_3$, with δ_3 governed by Theorem 7.1. We set the number of sampled belief trajectories to $N_b = 600$.

L Distributional Discrepancy Details

Table 6 reports the estimated distributional discrepancies ϵ and the corresponding ratios ϵ/α (at $\alpha = 0.2$) for each environment and action sequence. The ratio ϵ/α determines the fraction of the risk budget consumed by the model simplification and directly governs bound tightness: it appears as the weight on the truncation correction term in Theorem A.3. Values range from 3.5% to 16.6%, indicating that the observation model approximation imposes a non-trivial but manageable cost on the bounds.

Table 6. Estimated distributional discrepancies and their ratios to the risk level $\alpha = 0.2$ for each environment and action sequence. The ratio ϵ/α governs the fraction of the risk budget consumed by the observation model simplification.

Environment	Path	ϵ	ϵ/α
Light-Dark	Dangerous	0.0118	5.9%
Light-Dark	Safe	0.0113	5.7%
Laser Tag	Dangerous	0.0332	16.6%
Laser Tag	Safe	0.0331	16.6%
Push	Dangerous	0.0074	3.7%
Push	Safe	0.0070	3.5%

Figure 18 reports the sensitivity of the estimator in (56) to the number of observation samples N_z. For each environment and each of five seeds, $N^\Delta = 100$ states are drawn from Q_0 once and the estimate $\frac{1}{N^\Delta}\sum_n \hat{\Delta}^s_x(x^\Delta_n)$ is

recomputed for $N_z \in \{50, 100, 200, 500, 1000, 2000, 5000\}$ with $K = 10$. In all three environments the estimate is flat from $N_z \approx 200$ on, and the change from $N_z = 1000$ to 2000 and from 2000 to 5000 is smaller than the spread across seeds, so $N_z = 2000$ lies past convergence. The remaining spread at large N_z comes from which N^Δ states were drawn, not from N_z.

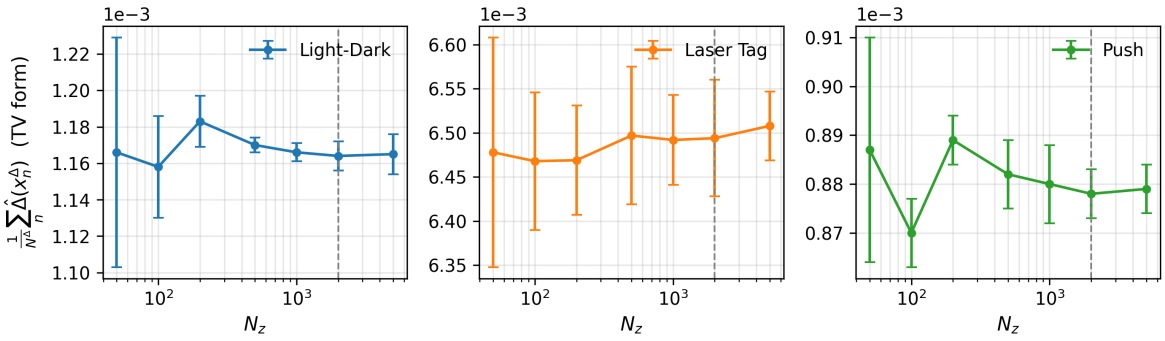


Fig. 18. Convergence of the distributional discrepancy estimate with the number of observation samples N_z (mean ± one standard deviation over five seeds; log scale). The dashed line marks the value $N_z = 2000$ used in the experiments.

M BetaZero Policy Configuration (Light-Dark)

The closed-loop policy evaluation uses BetaZero (Moss, Corso, et al. 2024), an AlphaZero-style algorithm for POMDPs. A neural network maps a belief summary vector $\phi(b)$ to a scalar value estimate $V(b)$ and a policy prior $\pi(a \mid b)$, which guide MCTS over particle beliefs.

M.1 Belief Representation

Particles are summarised into a fixed-size feature vector using the mean-and-standard-deviation representation:

$$\phi(b) \triangleq \left[\bar{\mu}_x,\ \bar{\mu}_y,\ \bar{\sigma}_x,\ \bar{\sigma}_y\right] \in \mathbb{R}^4, \tag{182}$$

where $\bar{\mu}$ and $\bar{\sigma}$ are the weighted mean and standard deviation of the particle set.

Table 7. Belief representation parameters for BetaZero on Light-Dark.

Parameter	Value
State dimension	2
Belief feature dimension	4 (= 2× state dim)
Particles per belief N_p	500

Input normalisation is applied using online statistics collected during training:
For the policy evaluation of Figure 11 the estimated discrepancy is $\hat{\epsilon} = 0.0109$, i.e. $\epsilon/\alpha = 5.4\%$ at $\alpha = 0.2$.

M.2 Network Architecture

Two-headed MLP with a shared trunk and separate value and policy heads:

Table 8. Input normalisation statistics for the BetaZero network.

Feature	Mean	Std
$\bar{\mu}_x$	3.39	2.00
$\bar{\mu}_y$	4.07	1.91
$\bar{\sigma}_x$	0.300	0.389
$\bar{\sigma}_y$	0.300	0.390
Value output	−19.65	30.69

Table 9. BetaZero network architecture.

Layer	Size
Input	4
Hidden 1	128
Hidden 2	128
Value head output	1 (scalar)
Policy head output	4 (one logit per action)

Table 10. MCTS parameters for BetaZero on Light-Dark.

Parameter	Value
Search depth	10
Simulations per step (training)	100
Simulations per step (inference)	1000
Exploration constant c	219.13 (tuned)
Temperature τ	0.319 (tuned)
Value weight z_q	1.0
Visit count weight z_n	1.0
DPW action widening k_a, α_a	1.0, 0.5
DPW observation widening k_o, α_o	1.0, 0.5
Discount factor γ	0.95

M.3 MCTS Parameters

M.4 Training Loop

Table 11. BetaZero training loop parameters.

Parameter	Value
Outer iterations	50
Episodes per iteration	50
Total training episodes	2500
Episode length	20 steps
Data collection	Full MCTS rollouts
Replay buffer	5 iterations

At each episode, a fresh initial belief is sampled: mean $\sim \mathcal{U}([-1, 8]^2)$, std $\sim \mathcal{U}(0.01,\ 2.0)$, then $N_p = 500$ particles are drawn from $\mathcal{N}(\text{mean},\ \text{std}^2 I)$. This diversity prevents the training distribution from collapsing to a uniform loop.

M.5 Optimizer and Loss

Table 12 lists the optimizer and loss settings used to train the BetaZero network.

Table 12. Optimizer and loss configuration for BetaZero training.

Parameter	Value
Optimizer	Adam
Learning rate	10^{-4}
Batch size	1024
Training epochs per iteration	50
Weight decay	10^{-5}
Loss	Cross-entropy (policy) + MSE (value)

M.6 Hyperparameter Search

The exploration constant c and temperature τ were selected via Optuna (TPE sampler, 20 trials) with $n_{\text{sim}} = 100$ during search:

Table 13. Hyperparameter search results for BetaZero on Light-Dark.

Hyperparameter	Search range	Best value
c (exploration constant)	$[1.0,\ 10 \times \Delta r] \approx [1,\ 220]$	219.13
τ (temperature)	$[0.05,\ 0.5]$	0.319
z_q, z_n	fixed	1.0, 1.0

Search used a MedianPruner (3 startup trials, 5 warmup steps) and a TPE sampler with seed 42. The objective minimised the final policy loss (cross-entropy vs. MCTS visit counts); the baseline uniform policy corresponds to $\ln 4 \approx 1.386$. The best trial was retrained at $n_{\text{sim}} = 1000$ for the final model used in all experiments. The search was run on an earlier version of the environment with $\tau = 0.1$; the final model was retrained with the same c and τ on the environment of Appendix K, reaching a final policy loss of 0.237.

N Hardware Specifications

All simulations were conducted on a machine equipped with an 11th Gen Intel® Core™ i9-11900K CPU running at 3.50 GHz and 64 GB of RAM.

O Experimental Validation of CVaR Bounds

This appendix provides full details for the CVaR concentration bounds experiments.

O.1 Synthetic Distributions: Full Results

Figure 7 shows the convergence comparison across eight distributions using exact local DKW, while Figure 19 shows sensitivity to the risk level α.

O.2 Truncation Decomposition Figure

Figure 20 illustrates the truncation decomposition of Section 5.2: the main CVaR term (blue) uses the observed sample extreme, and the support bound enters only through the tail correction (red) beyond it.

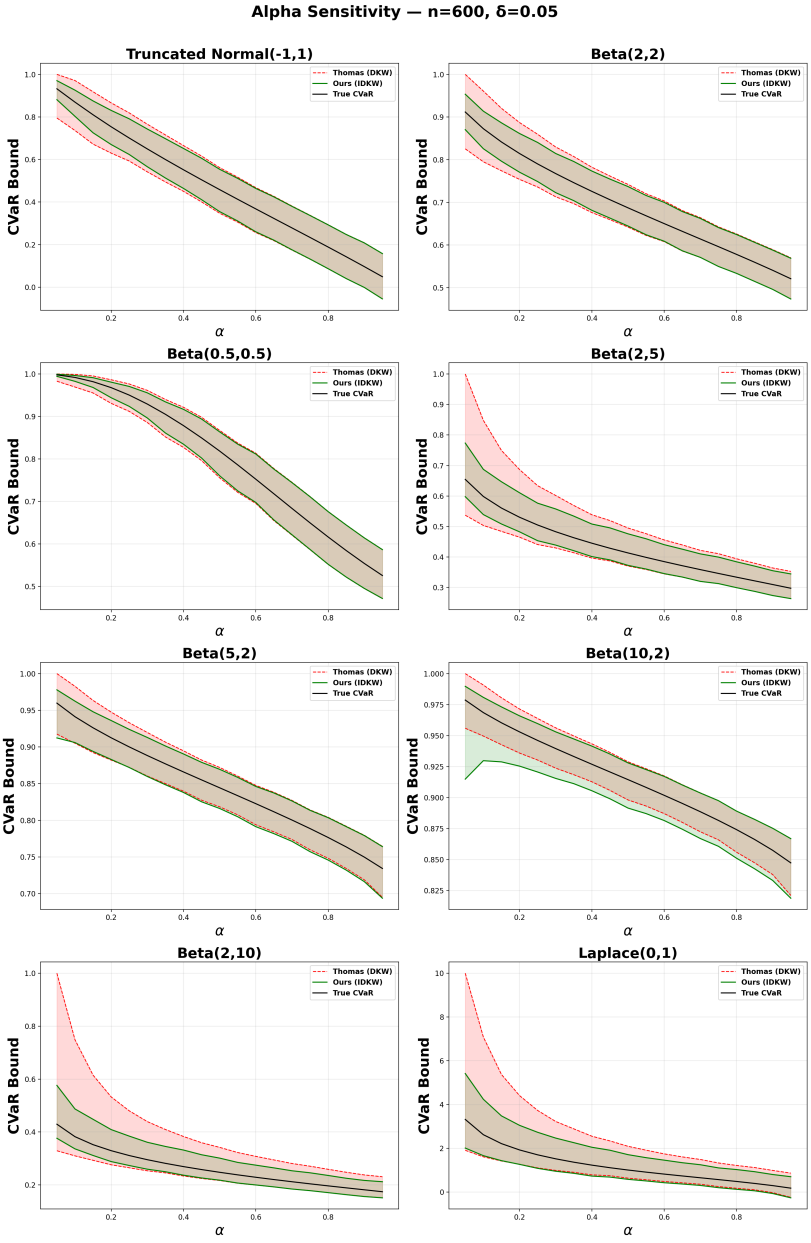
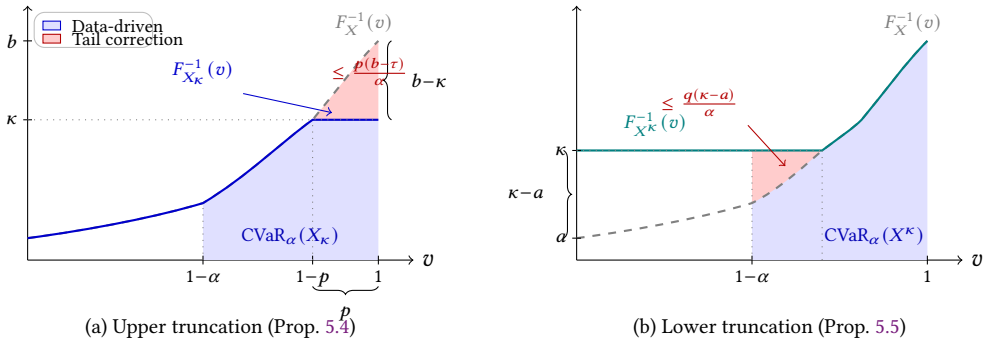


Fig. 19. Alpha sensitivity across eight distributions ($n = 600$, $\delta = 0.05$). Red dashed: Thomas bounds. Green solid: our bounds (η_{IDKW}). Black line: true CVaR. The gap between the two methods widens at small α, confirming that local concentration provides the largest benefit in the extreme-tail regime.



(a) Upper truncation (Prop. 5.4)

(b) Lower truncation (Prop. 5.5)

Fig. 20. Truncation decomposition of $\mathrm{CVaR}_\alpha(X)$. The dashed gray curve is $F_X^{-1}(v)$. The blue region is the data-driven term, requiring no support assumption. The red region is the tail correction, bounded by $O(1/n)$. **(a)** Upper truncation: correction is *added*, bounded by $(p/\alpha)(b-\kappa)$ with $p = O(1/n)$. **(b)** Lower truncation: correction is *subtracted*, bounded by $(q/\alpha)(\kappa - a)$ with $q = O(1/n)$.

Received 25 March 2026; accepted 09 October 2026