Response to comments of Reviewer 1:

1. C1 is not essential to our analysis. All our technical statements hold even without **C1**; this assumption simply makes the VaR and CVaR easier to express. On the other hand, **C2** (finiteness $(1+\epsilon)$th moment for some $\epsilon > 0$) is essential to our analysis, though it is only slightly more restrictive than assuming that the MAB problem is well- posed (which corresponds to $\epsilon = 0$). Note that our algorithms do not know this ϵ.

2. It turns out that the regret minimization framework is much harder to work within a distribution oblivious setting. Indeed, to the best of our knowledge, none of the current algorithms for regret analysis are distribution oblivious.

3. We agree that it makes sense to include a brief review of various risk-aware statistical approaches in the literature – we will do so if the paper is accepted. However, we think the Bayesian risk framework may not be directly suited to our distribution oblivious setting, since it involves assigning a prior distribution over the unknown parameter space.

4. Linear combinations of the mean and a risk metric are analytically tractable, and therefore ubiquitous in the literature (for example, in modern portfolio theory). Of course, linear combinations also come up naturally in Lagrange relaxations of constrained optimization, such as minimizing expected loss subject to a given risk constraint.

5. Note that we only provide *upper bounds* on the misidentification probability in oblivious and non-oblivious settings. We observe that the actual performance is often better than the bounds, as depicted in the figure.

Response to comments of Reviewer 2:

The reviewer correctly notes that SR-type algorithms are distribution oblivious with standard empirical estimators for mean/CVaR, though these perform very poorly if reward distributions are heavy-tailed. This can be formalized via *lower bounds* for these algorithms. If accepted, we will elaborate on this aspect in the final version of this paper. We apologize for omitting the line numbers. Thank you for pointing out the typos; these will be fixed.

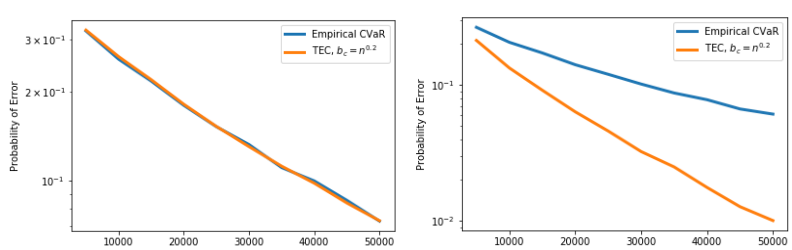
Response to comments of Reviewer 3:

1. Note that any SR-type algorithm requires apriori information about the time horizon. However, our uniform exploration (UE) algorithm can be made *any-time* using the doubling trick.

2. We submit that the proposed algorithms do not require all rewards to be stored. For SR, since the truncation parameters in each phase are known ahead of time, summary statistics of prior pulls can be stored for future rounds, leading to an $O(K^2)$ memory requirement (rather than $O(T)$). For UE, $O(K)$ memory suffices, even for the any-time modification.

3. On applicability: Heavy tailed distributions are ubiquitous, and are used to model financial rewards, losses due to natural calamities, wealth, Internet file sizes, social media followers, etc. Our formulations can, therefore, be applied to online learning problems in networking, economics, and finance. Of course, our CVaR concentration inequalities are of independent value, even beyond the MAB framework.

4. On experimental comparisons between the conventional CVaR estimator and the oblivious truncation-based CVaR estimator: Consider two pure exploration CVaR minimization scenarios. (a) **Light-tailed setting:** We have two exponentially distributed arms with the means such that the arms have CVaRs 1.0 and 0.98 at confidence 0.95. (b) **Heavy-tailed setting:** We have two Lomax distributed arms with shape parameter 2 and scale parameter set so that the arms have CVaRs 1.0 and 0.9 at confidence 0.95. For each value of T, we run the SR algorithm 50000 times. In the light-tailed setting (see Fig. 1a), the conventional empirical estimator works well, and truncation-aided bias-variance tradeoff does not help. But in the heavy-tailed setting (see Fig. 1b), the conventional empirical estimator performs poorly because of its considerable variance.



(a) Light Tailed Losses (b) Heavy Tailed Losses

Figure 1: Comparison of conventional and truncation-based CVaR estimators