We would like to thank the reviewers for their comments.

Reviewer 1

On suggested improvements to Sec 6: We agree that an empirical evaluation would be interesting. However, given that the thrust of this work has been to lay the foundations necessary for employing risk measures in a ML-type application (for instance, the CVaR-bandits application), and coupled with space limitations, we chose to postpone a detailed empirical investigation to future work. Nevertheless, we shall add numerical experiments in a longer version of this paper, and make it available on arxiv.

On CPT result being out of place: We believe that CPT is a general risk measure, and our result on CPT-concentration ties in well with the overall theme of the paper, which is to obtain concentration bounds for risk measures, and use Wasserstein distance along the way to achieve the result. The bounds we obtain for the case of CPT improve on the state-of-the-art, esp. for the sub-Gaussian case. As an aside, one can choose an appropriate weight function in the definition of CPT-value, and recover CVaR.

On the source of sharper results: We obtain sharper bounds by 1) relating the estimation error directly to one of the three characterizations given in eq. (2) of the Wasserstein distance between the empirical and true distributions, and 2) using concentration bounds for the Wasserstein distance between the true and empirical distribution given in Fournier and Guillin (2015).

On why Wasserstein metric was used: The estimation errors for all three risk measures can be directly related to the Wasserstein distance (this is clear from the proofs, but we will also clarify this in the main body of the final paper). Moreover, the bounds resulting from its use apply to unbounded random variables as well. If the DWK inequality were to be used to obtain concentration bounds for CVaR/CPT, the sup norm in the DKW inequality will make the resulting bounds applicable only to bounded r.v.s.

On how sharp are our results: We believe our bounds for the sub-Gaussian case are the best achievable in a minimax sense. Setting alpha=0, we recover the expected value, and it is apparent that the dependence on number of samples n and accuracy epsilon cannot be any better. On the other hand, for the case of sub-exponential or even the case of distributions with bounded higher moments, our bounds could be improved. As remarked on l. 153-159, the source of sub-optimality is not in our analysis. Instead, the Wasserstein concentration result from [Fournier and Guillin, 2015] for the latter case is far from optimal, as it involves a power law decay instead of an exponential one.

Reviewer 2

On the connection between CVaR and the Wasserstein distance: As observed by reviewer 3, the estimation error in each of the three risk measures that we consider is directly related to one of the three characterizations of Wasserstein distance given in eq. (2). For instance, as shown in the proof of Prop. 1, the estimation error in CVaR is related to the first equality in (2). Similarly, the estimation errors in CPT and spectral risk measures are related to the 2nd and 3rd equalities in (2).

On the bound being simple application of a sub-Gaussian concentration result We strongly disagree. In fact, sub-Gaussian concentration bounds (for e.g., Hoeffding's inequality) is never invoked in our proofs. Instead, we relate the CVaR estimation error to the distance between empirical and true distributions, and then invoke an inequality from [Fournier and Guillin, 2015]. The concentration bound that we derive is a convergence rate result, as it shows that the CVaR estimate converges at an exponential rate to the true CVaR.

On the strength of the technical contributions As outlined in l. 43-60, our bounds show a much better dependence on the number of samples n and accuracy $epsilon$, as compared to the state-of-the-art. Moreover, unlike our approach, an alternative proof that is based on quantiles (cf. [Kolla et al. 2019]) does not allow a bandit application. Finally, our approach is unified in the sense that one does not require separate proofs to handle the cases of sub-Gaussian, sub-exponential, or even distributions with bounded higher moments.

Reviewer 3

On the reliance on existing results While the estimation error for each risk measure is related to the Wasserstein distance between EDF and CDF, the proof takes a different route in arriving at this relation for CVaR, spectral risk measures, and CPT. In the first case, we use Lipschitz-ness and inf-norm for CVaR, while spectral risk measures are handled by a relation involving distribution inverses. Finally, for the case of CPT, when the distribution has bounded support, we can relate the estimation error to Wasserstein distance; the case of sub-Gaussian distribution involves more work, as we employ a truncated CPT-value estimator, and show that for such a scheme, one can handle the truncated and non-truncated part separately in the proof.