Overcoming The Sample Complexity Barrier in Risk Analytics with Debiased Learning

Abstract. We introduce a new approach for robustifying objectives involving risk measures such as Conditional Value at Risk (CVaR) against the error introduced by "plugging-in" a model estimated from data. The proposed estimator seeks to automatically cancel the bias introduced despite not knowing the nature of the error committed in the estimation step. Interestingly, this bias correction can be achieved without committing to a specific distribution family and relies only on a sufficiently regular tail behaviour exhibited by a wide-variety of common probability distributions. Under mild conditions, this debiasing exercise is shown to lead to consistent decisions with exponentially fewer data samples than required by sample-average approximation. These findings add to the growing body of evidence on the effectiveness of debiasing towards tackling major challenges in "estimate, then-optimize" paradigm while optimizing under uncertainty.

Keywords: Tail risks; Conditional Value at Risk; Statistical Debiased Learning; Path Derivative; Rare Events

1. Problem Introduction

This work considers a novel approach to tackling the problem of insufficient data in risk analytics. Typically, the effect of distribution tails is introduced into optimisation paradigms through the incorporation of a tail risk measure such as Conditional value at risk (CVaR) into the problem formulation. This approach has served as a primary vehicle for introducing risk aversion in a variety of planning and resource allocation problems in operations research and quantitative finance (see for instance Rockafellar & Uryasev (2000), Chow *et al.* (2015), McNeil *et al.* (2015)). For a loss $\ell(\boldsymbol{X}, \boldsymbol{\theta})$ associated with a decision choice $\boldsymbol{\theta}$ under a random realization $\boldsymbol{X}$, let $v_\beta(\boldsymbol{\theta})$ denote the $(1-\beta)$-th quantile of $\ell(\boldsymbol{X}, \boldsymbol{\theta})$. Then its CVaR at the tail-level $\beta \in (0,1)$ measures the average loss over the worst β-fraction of the realizations. Minimizing CVaR $C_\beta(\boldsymbol{\theta})$ over a compact set Θ enjoys the following variational representation

$$c_\beta = \inf_{u\in\mathbb{R},\boldsymbol{\theta}\in\Theta} \left[u + \beta^{-1} E_P \left(\ell(\boldsymbol{X}, \boldsymbol{\theta}) - u\right)^+\right] = \inf_{u\in\mathbb{R},\boldsymbol{\theta}\in\Theta} f(u, \boldsymbol{\theta}), \tag{1}$$

which is a convenient starting point for solving optimization problems.

Since (1) is rarely solvable in closed form, one often constructs a sample average approximation (SAA): given data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, one instead solves

$$\hat{c}_{\beta,n} = \inf_{u,\boldsymbol{\theta}} \left[u + \frac{1}{n\beta}\sum_{i=1}^{n}(\ell(\boldsymbol{X}_i, \boldsymbol{\theta}) - u)^+\right].$$

Large sample properties for the above SAA formulation have been extensively studied, and show that the limiting variances scale inversely with the tail level β under consideration. Consequently at even moderate tail levels, for practical data sizes, errors in CVaR estimation may affect the optimal decision (see Lim *et al.* (2011)) and data requirements to solve (1) accurately become unreasonable (see Caccioli *et al.* (2018)).

One approach to overcome the above difficulty is to commit a-priori to a model for the distribution of $\boldsymbol{X}$. In this case, (1) may be solved through using simulation and variance reduction techniques as in Deo *et al.* (2022), Barrera *et al.* (2015), He *et al.* (2021). Unfortunately, the error incurred in optimal decisions by committing to the wrong model is often unquantifiable. Alternately, one can estimate a model for the distribution of $\boldsymbol{X}$ from data, and plug-in the estimated model into the optimisation formulation (1) (that is solve (1) with the estimated model $\hat{P}$ rather than P). However, in many instances, these plug-in models are estimated non-parametrically (say using kernel density estimation), leading to poor rates of convergence of the optimisation problem with the plugged in distribution; see Hines *et al.* (2022) for more details.

Our technique attempts to combat this slow rate of convergence. We systematically characterise the bias incurred by plug-in estimation, and devise a statistical methodology to eliminate it. Our de-biased estimator enjoys a variance that is orders of magnitude lower than SAA, as well as a bias that smaller (quantified precisely in Theorem 3.2) than plug-in estimators. We show that under mild conditions, our estimator achieves the accuracy guarantees obtained when the underlying distribution of $\boldsymbol{X}$ is known (that is, one can commit to the correct model), and state of the art variance reduction techniques can be deployed. While such statistical debiaisng has been adopted increasingly in settings where non-parametric estimation is required (see Chernozhukov *et al.* (2018), Ichimura & Newey (2022), Foster & Syrgkanis (2019) and the references therein), to the extent of our knowledge, this is the first work which incorporates these ideas into tail risk analytics.

2. Model Assumptions and Methodology

We make the following very mild assumption on the density of $\boldsymbol{X}$

Assumption 1.

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \exp(-\varphi(\boldsymbol{x})) \text{ where } n^{-1}\varphi(n^p\boldsymbol{x}) \to \varphi^*(\boldsymbol{x}) \tag{2}$$

Assumption 1 simply means that the log of the density of $\boldsymbol{X}$ is asymptotically homogeneous and imposes a non-restrictive tail regularity constraint. A number of practically occurring *light tailed* distributions satisfy this; for a detailed list, we refer the reader to Deo & Murthy (2021). In what follows, assume for simplicity that $p = 1$.

Observe that in order to accurately solve (1), one needs to estimate $\Psi(P) = E_P[\ell(\boldsymbol{X}, \boldsymbol{\theta}) - u]^+$. A Taylor series expansion of Ψ shows that (see (Hines *et al.* 2022, Section 4))

$$\Psi(P) = \Psi(\hat{P}) + E_P[\mathrm{IF}(\hat{P})]. \tag{3}$$

The quantity $\mathrm{IF}(\hat{P})$ gives the effect of an infinitesimal perturbation of P in the direction of $\hat{P}$ on the value of the functional Ψ, and is referred to as the path derivative or influence function of Ψ at $\hat{P}$. Unfortunately, using (3) directly to compute $\Psi(P)$ simply results in SAA (see (Hines *et al.* 2022, Example 1)) and does not provide added benefits. To this end, model information given by Assumption 1 must be leveraged in constructing a more targeted debiased estimator.

Observe that since the density of $\boldsymbol{X}$ satisfies Assumption 1, the plug-in model $\hat{P}$ may be restricted to the class of distributions with log-homogeneous densities. Specifically, the density of $\hat{P}$ may be restricted to functions of the form $f_{\hat{P}}(\boldsymbol{x}) = \exp(-\hat{\varphi}(\boldsymbol{x}))$ where $\hat{\varphi}$ is a shape parameter estimated from data (under Assumption 1, one can construct $\hat{\varphi}$ to be a consistent estimator of the true shape parameter φ^*). One can utilise this information to replace $\mathrm{IF}(\hat{P})$ by its restriction to homogeneous functions. The appropriate notion of restriction here is given by the conditional expectation of $\mathrm{IF}(\hat{P})$ given the relevant model information. Let $\delta(\hat{P}) = E[\mathrm{IF}(\hat{P}) \mid \mathcal{G}_{\mathrm{Hom}}]$, where $\mathcal{G}_{\mathrm{Hom}}$ is the σ-algebra generated by all homogeneous functions that satisfy $E_{\hat{P}}[g] = 0$.

Lemma 2.1. *Suppose Assumption 1 holds, and let $\hat{\varphi}$ be an estimate of the tail shape. Then,*

$$\Psi(P) = \Psi(\hat{P}) + E_P[\delta(\hat{P})] + O_P(\|\hat{\varphi} - \varphi^*\|^2). \tag{4}$$

Loosely speaking, (4) is a consequence of projection and the Taylor approximation: $dP/d\hat{P} = (1 + [\hat{\varphi} - \varphi^*] + O(\|\hat{\varphi} - \varphi^*\|^2))$. An estimator for $\Psi(P)$ may now be constructed as follows (conditional on the computation of δ, which we demonstrate shortly):

$$\Psi_{db,n}(\hat{P}) = \Psi(\hat{P}) + \frac{1}{n}\sum_{i=1}^{n} \delta(\boldsymbol{X}_i, \hat{P}), \text{ where } \boldsymbol{X}_i \text{ are i.i.d. data samples.} \tag{5}$$

The CVaR optimisation problem may now be solved by plugging in the debiased estimator (5) into (1).

3. Key Results and Implementation

For any tail model P_0, with a homogeneous shape parameter φ (i.e., $f_{P_0}(\boldsymbol{x}) = \exp(-\varphi_0(\boldsymbol{x}))$), the projected influence function $\delta(\cdot, P_0)$ may obtained as the solution to the projection problem:

$$\min_{E_{P_0}[g]=0, g(r\boldsymbol{\phi})=rg(\boldsymbol{\phi})} E_{P_0}[\mathrm{IF}(P_0) - g]^2.$$

For $\boldsymbol{x} \in \mathbb{R}^d$, denote $\|\boldsymbol{x}\|_\infty = \max_{i \le d} |x_i|$ and $\hat{\boldsymbol{x}} = \boldsymbol{x}/\|\boldsymbol{x}\|_\infty$.

Proposition 3.1. *The efficient influence function $\delta(\cdot, P_0)$ of model with shape parameter φ_0 is given by,*

$$\delta(\boldsymbol{x}, P_0) = \|\boldsymbol{x}\|_\infty \frac{\zeta_3(\hat{\boldsymbol{x}}, P_0) - \mu(P_0)\zeta_1(\hat{\boldsymbol{x}}, P_0)}{\zeta_2(\hat{\boldsymbol{x}}, P_0)} \quad \text{where} \tag{6}$$

$$\zeta_1(\boldsymbol{\phi}, P_0) := E_{P_0}\left[R \mid \Phi = \boldsymbol{\phi}\right], \; \zeta_2(\boldsymbol{\phi}, P_0) := E_{P_0}\left[R^2 \mid \Phi = \boldsymbol{\phi}\right], \; \zeta_3(\boldsymbol{\phi}, P_0) := E_{P_0}\left[R \times IF(R\boldsymbol{\Phi}, P_0) \mid \Phi = \boldsymbol{\phi}\right],$$

and the constant $\mu(P_0) := \mu_2(P_0)/\mu_1(P_0)$, with $\mu_1(P_0) := E_{P_0}[\|\boldsymbol{X}\|_\infty \zeta_1(\hat{\boldsymbol{X}}, P_0)/\zeta_2(\hat{\boldsymbol{X}}, P_0)]$ and $\mu_2(P_0) := E_{P_0}[\|\boldsymbol{X}\|_\infty \zeta_3(\hat{\boldsymbol{X}}, P_0)/\zeta_2(\hat{\boldsymbol{X}}, P_0)]$, and where the density of P_0 is $f_{P_0}(r, \phi) \propto \exp(-r\varphi_0(\phi))$.

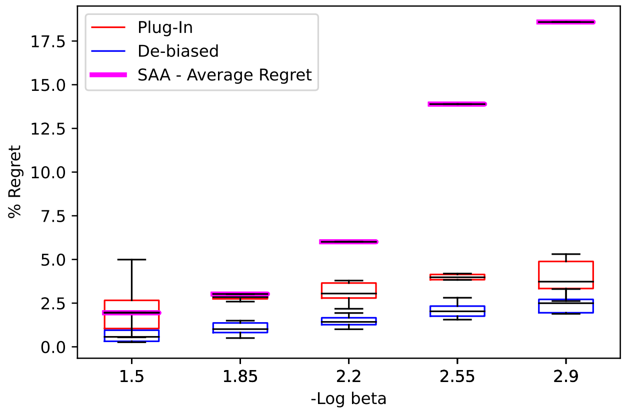

FIGURE 1. Box-plots for the errors for % regret for plug-in and debiased estimators are shown. To compare to sample averages, we also plot the average regret incurred by sample average approximation. In this experiment, we have used $n = 50$ samples to estimate the shape parameter φ for the former two methods, and $n = 500$ samples for SAA.

The expression for the efficient influence function $\delta(\cdot, \hat{P})$ given by Proposition 3.1 allows us to formulate an asymptotic expansion for $\Psi_{db}(P) - \Psi(P)$, in terms of the error in estimation of the shape parameter φ^* and the variance of the efficient influence function δ. This may be further utilised to give an error bound on the quality of solution obtained by debiasing:

Theorem 3.2. *Suppose Assumption 1 holds, and assume that the error in estimation of the shape parameter* $\|\hat{\varphi}-\varphi^*\| = O(n^{-c(d)})$. *Let* $\hat{c}_{\beta,n,db} = \inf_{u,\boldsymbol{\theta}} \left[u + \beta^{-1}\Psi_{db}(\hat{P})\right]$. *Then we have* $[\hat{c}_{\beta,n,db} - c_\beta] = O_P(n^{-1/2}\sigma_\beta + n^{-2c(d)})$. *In particular, whenever* $c(d) > 1/4$, *we obtain a central limit theorem:*

$$\sqrt{n}[\hat{c}_{\beta,n,db} - c_\beta] \Rightarrow N(0, \sigma_\beta^2), \text{ where } \sigma_\beta^2 = o(\beta^{-\varepsilon}) \text{ for any } \varepsilon > 0.$$

An important observation from Theorem 3.2 is that when the tail parameter φ^* can estimated sufficiently well ($c(d) > 1/4$), the debiased estimator converges at the canonical rate, with an error smaller than SAA. When the convergence of $\hat{\varphi}$ is slow, that is the $n^{-2c(d)}$ term dominates, the debiased estimator has a significantly smaller error than its plug-in counterpart (it can be shown that in this case, the plug in would incur a bias of $n^{-c(d)}$). Table 1 offers a more detailed comparison. A key point is that when the estimated shape $\hat{\varphi}$ converges to φ^* sufficiently fast, the debiased estimator performs as well as the one where the distribution of $\boldsymbol{X}$ is known, and state of the art variance reduction methods (for e.g. He *et al.* (2021), Deo *et al.* (2022)) can be deployed.

TABLE 1. Comparing various methods

Method	Asymptotic Error	Samples required for vanishing error
SAA	$n^{-1/2}\beta^{-1/2}$	β^{-1}
Debiased SAA, $c(d) > 1/4$	$n^{-1/2}\sigma_\beta$	$o(\beta^{-\varepsilon})$, for ε small
Model known, Importance Sampling	$n^{-1/2}\sigma_\beta$	$o(\beta^{-\varepsilon})$, for ε small

Numerical Exploration: We consider a simple setup where $\ell(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{\theta}^\intercal \boldsymbol{x}$ and $\boldsymbol{X}$ is exponentially distributed with independent marginals. Figure 1 compares the relative errors in CVaR optimisation using the respective methods. Notice that the use of debiasing helps significantly reduce out of sample errors: for instance, with $\beta = 3/1000$, debiasing gives a relative error in the range $1 - 2.4\%$, while the plug-in estimator gives of $4 - 5\%$.

Conclusion: In this work, we present a novel approach to estimate tail risk measures. We leverage the theory of semi parametric estimation to propose a debiased Sample Average Approximation based estimator for Conditional Value at Risk Optimisation. We show the our estimator enjoys a low variance, and also incurs a bias which is orders of magnitude smaller than the plug-in estimators widely used in practice. A simple numerical example validates our theoretical findings.

References

Barrera, Javiera, de Mello, Tito Homem, Moreno, Eduardo, Pagnoncelli, Bernardo K., & Canessa, Gianpiero. 2015. Chance-constrained problems and rare events: an importance sampling approach. *Mathematical Programming*, **157**(1), 153–189.

Caccioli, Fabio, Kondor, Imre, & Papp, Gábor. 2018. Portfolio optimization under expected shortfall: contour maps of estimation error. *Quantitative Finance*, **18**(8), 1295–1313.

Chernozhukov, Victor, Chetverikov, Denis, Demirer, Mert, Duflo, Esther, Hansen, Christian, Newey, Whitney, & Robins, James. 2018. *Double/debiased machine learning for treatment and structural parameters.*

Chow, Yinlam, Tamar, Aviv, Mannor, Shie, & Pavone, Marco. 2015. Risk-sensitive and robust decision-making: a cvar optimization approach. *Pages 1522–1530 of: Advances in Neural Information Processing Systems.*

Deo, Anand, & Murthy, Karthyek. 2021. Achieving Efficiency in Black Box Simulation of Distribution Tails with Self-structuring Importance Samplers. *arXiv preprint arXiv:2102.07060.*

Deo, Anand, Murthy, Karthyek, & Sarker, Tirtho. 2022. Combining Retrospective Approximation with Importance Sampling for Optimising Conditional Value at Risk. *arXiv preprint arXiv:2102.07060.*

Foster, Dylan J, & Syrgkanis, Vasilis. 2019. Orthogonal statistical learning. *arXiv preprint arXiv:1901.09036.*

He, Shengyi, Jiang, Guangxin, Lam, Henry, & Fu, Michael C. 2021. Adaptive Importance Sampling for Efficient Stochastic Root Finding and Quantile Estimation. *arXiv preprint arXiv:2102.10631.*

Hines, Oliver, Dukes, Oliver, Diaz-Ordaz, Karla, & Vansteelandt, Stijn. 2022. Demystifying statistical learning based on efficient influence functions. *The American Statistician*, 1–13.

Ichimura, Hidehiko, & Newey, Whitney K. 2022. The influence function of semiparametric estimators. *Quantitative Economics*, **13**(1), 29–61.

Lim, Andrew EB, Shanthikumar, J George, & Vahn, Gah-Yi. 2011. Conditional value-at-risk in portfolio optimization: Coherent but fragile. *Operations Research Letters*, **39**(3), 163–171.

McNeil, Alexander J, Frey, Rüdiger, & Embrechts, Paul. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools.* Princeton university press.

Rockafellar, R Tyrrell, & Uryasev, Stanislav. 2000. Optimization of Conditional Value-at-Risk. *Journal of Risk*, **2**, 21–42.