Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

ON THE IMPACTS OF TAIL MODEL UNCERTAINTY IN RARE-EVENT ESTIMATION

Zhiyuan Huang

Department of IOE
University of Michigan
1205 Beal Avenue
Ann Arbor, MI 48109, USA

Henry Lam

Department of IEOR
Columbia University
500 W. 120th Street
New York, NY 10027, USA

ABSTRACT

Rare-event probabilities and risk measures that quantify the likelihood of catastrophic or failure events can be sensitive to the accuracy of the underlying input models, especially regarding their tail behaviors. We investigate how the lack of tail information of the input can affect the output extremal measures, in relation to the level of data that are needed to inform the input tail. Using the basic setting of estimating the probability of the overshoot of an aggregation of i.i.d. input variables, we argue that heavy-tailed problems are much more vulnerable to input uncertainty than light-tailed problems. We explain this phenomenon via their large deviations behaviors, and substantiate with some numerical experiments.

1 INTRODUCTION

Assessing rare-event probabilities and extremal measures for the likelihood of catastrophic events is ubiquitous in risk analysis and management. Examples include the prediction of large asset losses in finance, imbalance of cash flows and ruin in insurance, and system overloads in service operations. In many cases, these extremal quantities are outputs that rely on underlying, granular stochastic components. For example, a financial portfolio may consist of a weighted combination of assets each having its own (correlated) return pattern, and an insurance portfolio consists of the cash flows of many different policyholders. Estimating these extremal quantities hinges on the provision of accurate probabilistic descriptions of these input components, with any deviations away from the reality leading to potential errors or even meaningless estimates.

The latter issue has been studied and has gathered growing literature in recent years, generally known as the problem of model uncertainty or input uncertainty. Its main focus is to develop methodologies that can quantify the impact of model misspecifications or errors that propagate to output estimation or decision-making. See, e.g., Barton et al. (2002), Henderson (2003), Chick (2006), Barton (2012), Song et al. (2014), Lam (2016) in the stochastic simulation literature, and Petersen et al. (2000), Hansen and Sargent (2008), Glasserman and Xu (2014), Lim and Shanthikumar (2007) in finance, economics, control and operations management applications. In the extremal estimation setting, this problem is intimately related to extreme value theory, in which one attempts to extrapolate the tail beyond the scope of data in a statistically justified fashion, along with uncertainty quantification (Embrechts et al. 2013; Embrechts et al. 2005). Recently, the framework of so-called distributionally robust optimization (Delage and Ye 2010; Wiesemann et al. 2014; Ben-Tal et al. 2013; Ghaoui et al. 2003) has been studied to construct bounds on extremal measures with additional robustness properties beyond statistical asymptotics. This approach utilizes postulations such as the acknowledgement of the true distribution within a neighborhood of a baseline model measured by a suitable statistical distance (Atar et al. 2015; Blanchet and Murthy 2016), marginal information and extremal coefficients (Embrechts and Puccetti 2006; Puccetti and Rüschendorf 2013; Wang and Wang 2011; Dhara et al. 2017; Puccetti and Rüschendorf 2012; Embrechts et al. 2013;

Yuen et al. 2019), moments and shape assumptions on the tail such as monotonicity or convexity (Lam and Mottet 2017; Van Parys et al. 2019; Li et al. 2019). In simulation-based rare-event analysis, Nakayama (1995), Nakayama (1998) studied methods to efficiently compute sensitivities of rare-event probabilities with respect to model parameters, and Nelson et al. (2019) proposed an averaging of distributions to fit input models to enhance tail performances.

In contrast to most past literature that focused on the technique in quantifying model uncertainty impacts, here we address several validity questions that arise when, given input data, a modeler chooses to use "standard" approaches to obtain estimates and quantify uncertainty, namely:

1. By simply using the empirical distribution as my input model fit, would the rare-event estimate be reasonably close to the truth? (assuming computational or Monte Carlo noise is negligible)
2. Following the point estimate in Question 1, would it work if one runs a bootstrap to obtain a confidence interval that accounts for the input data noise?
3. If the bootstrap does not work, would incorporating extreme value theory in fitting the input tail helps with more reliable uncertainty quantification?

Our viewpoint is that the main source of uncertainty in determining the accuracy of rare-event estimation comes from the lack of knowledge of the tail of the input models. The main body (i.e., non-tail) part of the input distribution can be fit by both parametric and nonparametric techniques, where there are typically adequate data to perform such fit (and in Question 1 above, we simply use the empirical distribution as the fit). However, it is the portion beyond the scope of data that determines the distributional tail and in turn the rare-event behaviors. Thus, before we go to the above questions, we first focus on:

0. How does truncating the tail of the input model affect the rare-event estimate?

Our main contention is that heavy-tailed problems could be much more challenging than light-tailed counterparts regarding estimation and uncertainty quantification using the standard approaches in Questions 1-3. This challenge roots from Question 0 in that truncating the input tail in a heavy-tailed system exerts a huge effect on the rare-event estimate, when the truncation level represents the typical level of knowledge that the data informs (e.g., the top 1% or 0.1% of the data). As a consequence, using empirical distribution, or bootstrap on the empirical distribution, which significantly ignores the tail content, would fail to estimate the rare-event quantity and vastly under-estimate the uncertainty. Using extreme value theory in Question 3 to extrapolate tail (such as the peak-over-threshold method, e.g. Leadbetter 1991) helps to an extent, but could introduce extra bias, at least using our fitting methods (though we should point out that better techniques are available). On the other hand, the effect of missing tails on light-tailed estimation is relatively milder.

The larger effect from truncating the input tail on heavy-tailed problems can be explained from their large deviations behaviors that pertain to the one or several "big jumps" (Embrechts et al. 2013; Denisov et al. 2008; Rhee et al. 2016), i.e., to invoke a rare event, one or several input components exhibit huge values. On the other hand, light-tailed systems invoke rare events by having each component contributing a small shift and adding these contributions. Thus, to accurately estimate a heavy-tail rare event, one needs to accurately estimate the far tail of each input component, whereas this is not necessary in light-tailed systems. In fact, ignoring the tail of heavy-tail inputs would lead to estimates as if the system is light-tail, and the ultimate effect could be that the estimation error is as large as the rare-event probability of interest, deeming the estimation meaningless.

We point out that, regarding Question 0, our study is related to Dupuis et al. (2018), Blanchet et al. (2014), Lam and Mottet (2015) and especially Olvera-Cravioto (2006) and Jelenkovic (1999). Dupuis et al. (2018) investigated the sensitivities on the large deviations rate when the input model deviates within a Rényi divergence ball. Blanchet et al. (2014) showed in a similar context that imposing a single ball over all inputs, thus allowing the distortion of dependency structure among the inputs, can lead to a

substantially heavier tail than the original model when the Kullback-Leibler divergence is used. Lam and Mottet (2015) studied robust rare-event simulation when the input tail is unknown but subject to geometric assumptions. Olvera-Cravioto (2006) studied the impacts on the waiting times when the tail of service times is misspecified or truncated. Relating to Jelenkovic (1999), they investigated the truncation threshold needed to retain the heavy-tail characteristic of a system. They also contrasted it with the light-tail case and observed that the required threshold is higher for heavy tail. Our observation in this regard is thus similar to Olvera-Cravioto (2006), but with a different setting (aggregation of i.i.d. variables) and focus on the statistical implications asked in Questions 1-3. Moreover, we investigate extensively on the numerical evidence and identify situations where the theoretical findings hold or deviate.

In the remainder of this paper, we will focus on a basic setup on the overshoot of an aggregation of i.i.d. variables. Section 2 describes the estimation target and explains the impacts of tail truncation in light- versus heavy-tailed cases. Section 3 shows the numerical results and comparisons in input tail truncation, and the use of empirical distributions and bootstrapping. We leave the full derivations and generalizations to the journal version of this work.

2 SETTING AND THEORY

We consider estimating the overshoot of an aggregation of n i.i.d. variables, i.e., consider $p = P(S_n > \gamma)$ where $S_n = X_1 + \cdots + X_n$ and $X_i \in \mathbb{R}$ are i.i.d. variables drawn from the distribution F. We denote X as a generic copy of X_i for convenience. We assume the density of X exists and denote as f. Correspondingly, we let $\bar{F}(x) = 1 - F(x)$ be the tail distribution function. We let $\mu = E[X] < \infty$. Suppose $\gamma = \gamma(n)$ is a high level that grows to ∞ as $n \to \infty$. Throughout this paper, for any sequences $a_n, b_n \in \mathbb{R}$ we write $a_n = o(b_n)$ if $a_n/b_n \to 0$ as $n \to \infty$, $a_n = \omega(b_n)$ if $a_n/b_n \to \infty$ as $n \to \infty$, and $a_n = \Theta(b_n)$ if there exists an integer n_0 such that $\underline{M} \le |a_n/b_n| \le \overline{M}$ for $n \ge n_0$ and $0 < \underline{M} \le \overline{M} < \infty$.

It is well known that, if $\gamma = bn$ for some constant $b > \mu$ and X possesses exponential moments, then under mild additional assumptions p decays exponentially in n (Dembo and Zeitouni 1998). On the other hand, if X is Pareto-tailed, then, as $\gamma = \omega(\sqrt{n})$, p approximately equals $P(\max_i X_i > \gamma - n\mu)$ or $n\bar{F}(\gamma - n\mu)$, which corresponds to the one-big-jump behavior.

Our investigation pertaining to Question 0 is the following. Suppose we truncate the distribution $F(x)$ at the point u so that the density becomes 0 for $x > u$, i.e., consider the truncated distribution function given by

$$\tilde{F}_u(x) = \begin{cases} F(x)/F(u) & \text{for } x \le u \\ 1 & \text{for } x > u \end{cases}$$

and correspondingly the truncated density $\tilde{f}_u(x) = (f(x)/F(u))I(x \le u)$, where $I(\cdot)$ denotes the indicator function. For convenience, denote $p(G)$ as the probability $P_G(S_n > \gamma)$ where X_i's are governed by an arbitrary distribution G, and we simply denote $P(S_n > \gamma)$ if X_i's are governed by F. We consider the approximation error $p(\tilde{F}_u) - p(F)$.

Note that, roughly speaking, this situation captures the case where we use the empirical distribution to plug into our input model, so that the probability mass is zero for regions outside the scope of data or close to zero at the very tail of the data. The proportional constant $F(u)$ is introduced to ensure a proper truncated distribution and has little effect on the mass below u when u is reasonably big.

By definition, the approximation error is

$$p(\tilde{F}_u) - p(F) = \frac{P(S_n > \gamma, X_i \le u \ \forall i = 1, \ldots, n)}{F(u)^n} - P(S_n > \gamma). \tag{1}$$

2.1 Heavy-Tail Case

We first consider the Pareto-tail case. Suppose that $\bar{F}$ has a regularly varying tail in the form

$$\bar{F}(x) = L(x)x^{-\alpha}(1 + o(1)) \tag{2}$$

for some slowly varying function $L(\cdot)$ and $\alpha > 2$, and $E|X_i|^{2+\delta} < \infty$.

Suppose $n \to \infty$ and $\gamma = \Theta(n)$ (or more generally $\gamma = \omega(\sqrt{n \log n})$). In this case, it is known that $P(S_n > \gamma)$ is approximately $P(\text{at least one } X_i > \gamma - n\mu)$, or probabilistically, that the rare event $S_n > \gamma$

happens most likely due to a big jump from one of the X_i's (e.g. Embrechts et al. 2013). Thus, if the truncation level u is too small compared to $\gamma - n\mu$, then the big jump that contributes to the dominating mass of the rare event is barred, making $P(S_n > \gamma, X_i \le u\ \forall i = 1,\ldots,n)$ substantially smaller than $P(S_n > \gamma)$. In this situation, $p(\tilde{F}_u)$ becomes negligible compared to $P(S_n > \gamma)$, and the approximation error (1) is effectively $-P(S_n > \gamma)$. In other words, using a truncated input distribution leads to a substantial under-estimate with a bias almost equal to the magnitude of the rare-event probability itself.

Alternately, we can write the approximation error (1) as

$$\frac{-P(S_n > \gamma,\ \text{at least one } X_i > u) + P(S_n > \gamma)(1 - F(u)^n)}{F(u)^n}. \tag{3}$$

Again, when u is relatively small compared to $\gamma - n\mu$, then the event $\{\text{at least one } X_i > u\}$ inside the probability $P(S_n > \gamma,\ \text{at least one } X_i > u)$ is redundant, making this probability asymptotically equivalent to $P(S_n > \gamma)$ and that (3) is asymptotically equivalent to $-P(S_n > \gamma)$.

We summarize the above as:

Theorem 1 Suppose X_i's are i.i.d. random variables with regularly varying tail distribution $\bar{F}$ in the form (2) with $\alpha > 2$ and $E|X|^{2+\delta} < \infty$. Let $n \to \infty$ and $\gamma = n\mu + \omega(\sqrt{n \log n})$. Assume $u \le (\gamma - n\mu)/\sqrt{\log n}$. The discrepancy between using a truncated distribution $\tilde{F}_u$ and the original distribution F in evaluating the probability $p(F) = P(S_n > \gamma)$ as $n \to \infty$ is given by

$$p(\tilde{F}_u) - p(F) = -p(F)(1 + o(1)).$$

Proof. Using equation (1.45) in (Nagaev et al. 1979), when $(\gamma - n\mu)/\sqrt{n \log n} \to \infty$ and $u \le (\gamma - n\mu)/\sqrt{\log n}$, we have $P(S_n > \gamma, X_i \le u\ \forall i = 1,\ldots,n) = o(n\bar{F}(\gamma - n\mu))$. Moreover, by Theorem 1.9 in (Nagaev et al. 1979) (or equation (1.25b) therein), we have $P(S_n > \gamma) = n\bar{F}(\gamma - n\mu)(1 + o(1))$ under the given conditions. Thus $P(S_n > \gamma, X_i \le u\ \forall i = 1,\ldots,n) = o(P(S_n > \gamma))$.

Moreover, note that if $u = (\gamma - n\mu)/\sqrt{\log n}$, we have $F(u)^n \to 1$. Thus

$$\frac{P(S_n > \gamma, X_i \le u\ \forall i = 1,\ldots,n)}{F(u)^n} = o(P(S_n > \gamma))$$

if $u = (\gamma - n\mu)/\sqrt{\log n}$. However, since the truncated distribution $\tilde{F}_u$ stochastically dominates $\tilde{F}_{u'}$, i.e., $\bar{\tilde{F}}_u(\cdot) \ge \bar{\tilde{F}}_{u'}(\cdot)$, for any u, u' such that $u > u'$, we must have, for given γ, $P(S_n > \gamma, X_i \le u\ \forall i = 1,\ldots,n)/F(u)^n$ non-decreasing in u. Therefore we have

$$\frac{P(S_n > \gamma, X_i \le u\ \forall i = 1,\ldots,n)}{F(u)^n} = o(P(S_n > \gamma))$$

for any $u \le (\gamma - n\mu)/\sqrt{\log n}$. This concludes the theorem. □

Consider the case $\gamma = bn$ for some $b > \mu$. Theorem 1 states that when u is below $(b - \mu)n/\sqrt{\log n}$, the rare-event estimation is essentially void, at least asymptotically. Note that this threshold is approximately linear in n. When the number of input components n is large, it could be difficult to sustain an accuracy level given a finite set of input data.

2.2 Light-Tail Case

We now consider X that possesses finite exponential moment, i.e., the logarithmic moment generating function $\psi(\theta) = \log E[e^{\theta X}] < \infty$ for θ in a neighborhood of 0. Consider $\gamma = bn$ for some constant $b > \mu$. Suppose that there exists a unique solution θ^* to the equation $b = \psi'(\theta)$. Then $p(F) = P(S_n > \gamma)$ exhibits exponential decay as $n \to \infty$, i.e., $-(1/n)\log p(F) \to I$ where I is the rate function given by the Legendre transform or the convex conjugate of $\psi(\theta)$

$$I = \sup_{\theta} \{b\theta - \psi(\theta)\}.$$

In fact, if X is further assumed non-lattice, we have the following more accurate asymptotic (Bucklew 2013)

$$P(S_n > \gamma) = \frac{1}{\theta^* \sqrt{2\pi\psi''(\theta^*)n}} e^{-nI}(1+o(1)).$$

We have:

Theorem 2 Consider $\gamma = bn$ for some constant $b > \mu$. Suppose that F is non-lattice, satisfies $\psi(\theta) < \infty$ for θ in a neighborhood of 0, and there exists a unique solution θ^* to the equation $b = \psi'(\theta)$. Then, as long as the truncation level u is chosen such that $ne^{\theta^* u}\bar{F}(u) \to 0$, the discrepancy between using a truncated distribution $\tilde{F}_u$ and the original distribution F in evaluating the probability $p(F) = P(S_n > \gamma)$ is asymptotically negligible, i.e.,

$$p(\tilde{F}_u) - p(F) = -o(p(F)).$$

Sketch of Proof. We consider the rate function corresponding to $p(\tilde{F}_u)$, given by

$$I_u = \sup_\theta \{b\theta - \psi_u(\theta)\}$$

where $\psi_u(\theta)$ denotes the logarithmic moment generating function of $\tilde{F}_u$, namely $\log(E[e^{\theta X}; X \le u]/F(u))$. Now, consider a change of variable $r = F(u)$, and abuse notation slightly to write $\psi_r(\theta) = \log(E[e^{\theta X}; X \le F^{-1}(r)]/r)$ and the corresponding rate function as I_r. By Taylor series expansion we have, as $u \to \infty$ or $r \to 1$,

$$I_u \approx I - \frac{d}{dr}\psi_r(\theta^*)(F(u)-1) \tag{4}$$

where $-\frac{d}{dr}\psi_r(\theta^*)$ is the derivative of I_r by the generalized Danskin's Theorem (Clarke 1975). Note that

$$\begin{aligned}\frac{d}{dr}\psi_r(\theta^*) &= \frac{e^{\theta^* F^{-1}(r)} f(F^{-1}(r))}{f(F^{-1}(r))E[e^{\theta^* X}; X \le F^{-1}(r)]} - \frac{1}{r} \\ &= \frac{e^{\theta^* u}}{E[e^{\theta^* X}; X \le u]} - \frac{1}{F(u)}.\end{aligned}$$

Hence from (4) we have

$$I_u \approx I + \left(\frac{e^{\theta^* u}}{E[e^{\theta^* X}; X \le u]} - \frac{1}{F(u)}\right)\bar{F}(u).$$

Now, one can show that, as $u \to \infty$, we have $\theta^* + \delta$ with $\delta \to 0$. Thus the approximation error $p(\tilde{F}_u) - p(F)$ is given by

$$\begin{aligned}&\frac{P(S_n > \gamma, X_i \le u \ \forall i = 1,\ldots,n)}{F(u)^n} - P(S_n > \gamma) \\ \approx\ & \frac{1}{(\theta^*+\delta)\sqrt{2\pi\psi''(\theta^*+\delta, u)n}} e^{-nI - n\left(\frac{e^{\theta^* u}}{E[e^{\theta^* X}; X \le u]} - \frac{1}{F(u)}\right)\bar{F}(u)} - \frac{1}{\theta^*\sqrt{2\pi\psi''(\theta^*)n}} e^{-nI}.\end{aligned} \tag{5}$$

Thus, when

$$n\left(\frac{e^{\theta^* u}}{E[e^{\theta^* X}; X \le u]} - \frac{1}{F(u)}\right)\bar{F}(u) \to 0 \tag{6}$$

we have (5) being asymptotically negligible compared to $(1/(\theta^*\sqrt{2\pi\psi''(\theta^*)n}))e^{-nI}$, which would conclude our claim. Finally, we only need to observe that (6) is equivalent to $ne^{\theta^* u}\bar{F}(u) \to 0$. □

Theorem 2 postulates that as long as the truncation level u is chosen high enough relative to n such that $ne^{\theta^* u}\bar{F}(u) \to 0$, then the model error in using the truncated input is negligible. In contrast to the heavy-tail case, this condition on u dictates typically a logarithmic requirement on n. For instance, if F is an exponential distribution, say with rate λ, then we have $ne^{-(\lambda-\theta^*)u} \to 0$ which holds as long as $n = \omega(\log n)$. If F is a Gaussian distribution, say with mean μ and variance σ^2, then we have $ne^{\theta^* u - (u-\mu)^2/(2\sigma^2)}/\sqrt{2\pi u} \to 0$ which holds as long as $u = \omega(\sqrt{\log n})$.

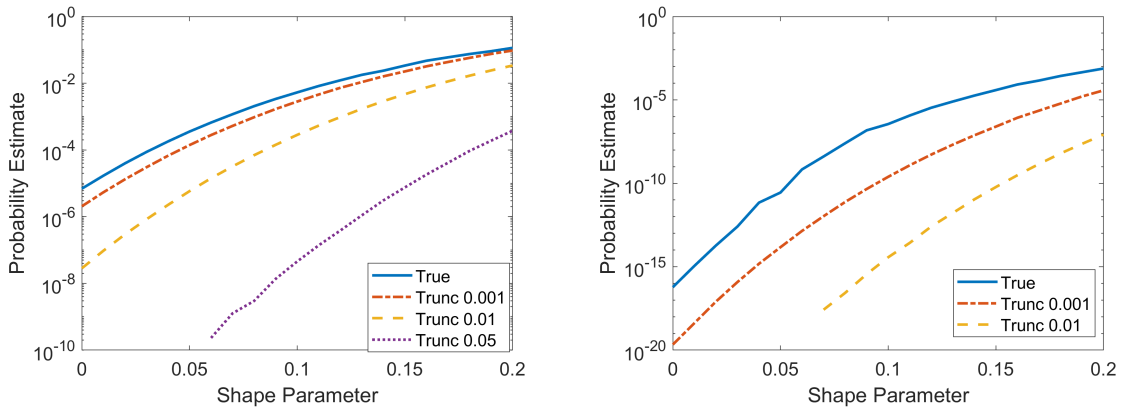


Figure 1: The probability estimation with untruncated and truncated distributions. "Trunc 0.001" denotes the probability estimate using distribution turncated at 0.001 tail quantile. (a) $\gamma = 60,\ n = 30$. (b) $\gamma = 100,\ n = 30$.

3 NUMERICAL EXPERIMENTS

We consider estimating the probability $p = P(\sum_{i=1}^{n} X_i > \gamma)$, with different number of variables n, rarity levels γ and distributions of X_i. In each of our experiments, we implement variance reduction techniques (including importance sampling and conditional Monte Carlo; for further details, see Asmussen and Glynn (2007)) to achieve better computation efficiency and use sufficient samples to ensure negligible simulation noise. We investigate the effect of input tail truncation (Question 0) in Section 3.1, using empirical distribution (Question 1) in Section 3.2, bootstrap (Question 2) in Section 3.3, and using peak-over-threshold (Question 3) in Section 3.4.

3.1 Truncating Input Tail

We test with truncation points on the input distribution corresponding to 0.05, 0.01 and 0.001 tail probability masses respectively, i.e., t such that $P(X > t) = \alpha$ where $\alpha = 0.05,\ 0.01,\ 0.001$ (We shall refer to as the α tail quantile). When we truncate the distribution at t, we only use $f(x|x < t)$ to generate X_i's. The estimate with untruncated distribution is set as a baseline for comparison.

We generate X_i's using the generalized Pareto distribution, varying the shape parameter ξ and the scale parameter σ and fixing the threshold parameter to be 0. Note that when $\xi = 0$, the distribution is light-tailed (equivalent to exponential distribution); when $\xi > 0$ the distribution is heavy-tailed and larger ξ gives heavier tail. We vary ξ from 0 to 0.2 to observe what would change if the tail part grows heavier. When ξ varies, we keep the mean of the distribution to be 1 by letting $\sigma = 1 - \xi$. Our experiments also include different settings of γ and n. Figures 1 and 2 show the experiment results.

Before comparing light and heavy tails, we note that the tail part of the distribution is generally quite important to the probability estimation. This claim is supported by the gaps between the probability estimates using true distribution and truncated distributions in Figures 1 and 2. For instance, suppose we truncate at the 0.01 tail quantile. The gap between the estimate with the truth (between the blue solid line and the yellow dash) is roughly greater than one order of magnitude in almost all cases, which means we are not able to estimate a correct scale of the probability without the 0.01 tail information. Moreover, for fixed γ and n, we see a smooth trend of the estimates (in Figures 1 and 2) when the shape parameter increases from 0.

Next we compare the impacts between light and heavy tails. Although the trends in the figures seem to suggest smaller gaps as ξ increases (heavier tail), a more proper comparison should fix the target probability

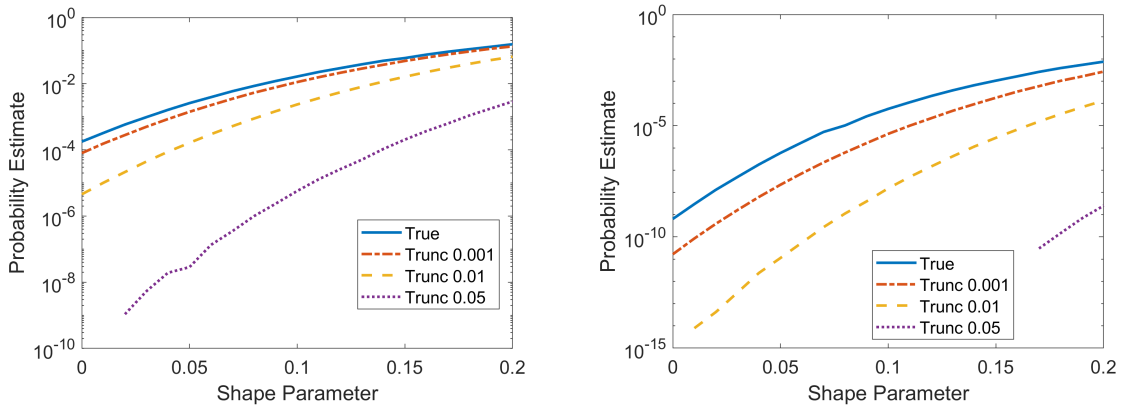


Figure 2: The probability estimation with untruncated and truncated distributions. "Trunc 0.001" denotes the probability estimate using distribution turncated at 0.001 tail quantile. (a) $\gamma = 40$, $n = 20$. (b) $\gamma = 60$, $n = 20$.

level and the truncation level. In this case, the impact of the heavier tail appears larger. In particular, we compare the gaps between the estimate with true distribution (blue solid line) and truncated distribution (orange dash-dot line) at the 0.001 tail quantile in Figure 1a at shape parameter value around 0 and in Figure 1b at value around 0.2. In these two cases, the objective probabilities have similar values and the truncated tail quantile are the same, and in the heavy-tail case, a larger gap can be seen. More specifically, the gap is smaller than one order of magnitude in light tail (Figure 1a at 0) and larger than one order of magnitude in heavy tail (Figure 1b at 0.2).

3.2 Data-driven Rare-Event Simulation

We consider the data-driven situation (i.e. when distribution of X is unknown but data is available) and use empirical distributions to drive the simulation of p. Here we set X_i as Gaussian distribution and generalized Pareto distribution with $\xi = 0.2$. We independently generate data sets of X_i's for 100 replications to construct empirical distributions to drive the simulation. Figures 3 and 4 show the true probability, the averaged estimates from all replications, and also the maximum and minimum estimates among the replications to provide a measure of variability.

In the light-tail cases, Figures 3 and 4a show similar trends in that the estimation variability reduces as the number of samples increases (the maximum and minimum values approach to the true probability as sample increases). The variability is higher when the rarity level grows, which can be observed from the slower convergence of the maximum and minimum estimates (Figure 4a versus Figure 3a).

On the other hand, the performance in the heavy-tail cases (Figure 4b) is more "abnormal". When the sample size is small, e.g. 10^4, the maximum estimate from the 100 replications is smaller than the true probability. Even with more samples (10^5 and 10^6), in most (92 and 75 out of the 100 respectively) replications, we obtain overly small estimates compared to the true probability (a difference of more than 5 orders of magnitude). These suggest a severe underestimation in the heavy-tail problem with limited data.

3.3 Using Nonparametric Bootstrap

Next we investigate the use of bootstrapping to assess input uncertainty. Such a technique has been studied in the simulation literature (e.g., Barton and Schruben 2001). In our experiment, we construct bootstrapped empirical distributions by repeatedly resampling with replacement and with full size from the data and using them to drive enough simulation runs per resample. The bootstrap size is $B = 100$. We use the empirical quantiles of the bootstrap estimates to construct a confidence interval for the estimate. We repeat our experiments 100 times. Here we again consider X_i with Gaussian distribution and generalized Pareto

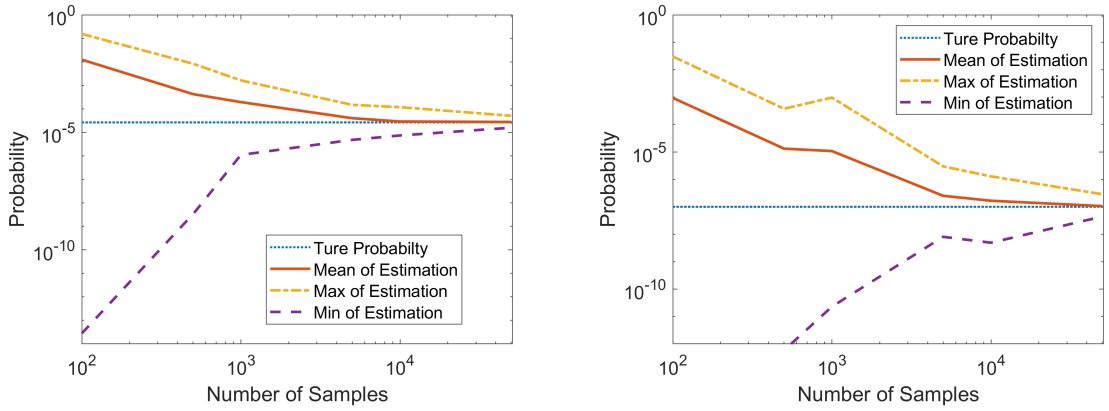


Figure 3: The estimation performance with different number of samples based on 100 replications. (a) Gaussian distribution, $\gamma = 70$. (b) Gaussian distribution, $\gamma = 90$.

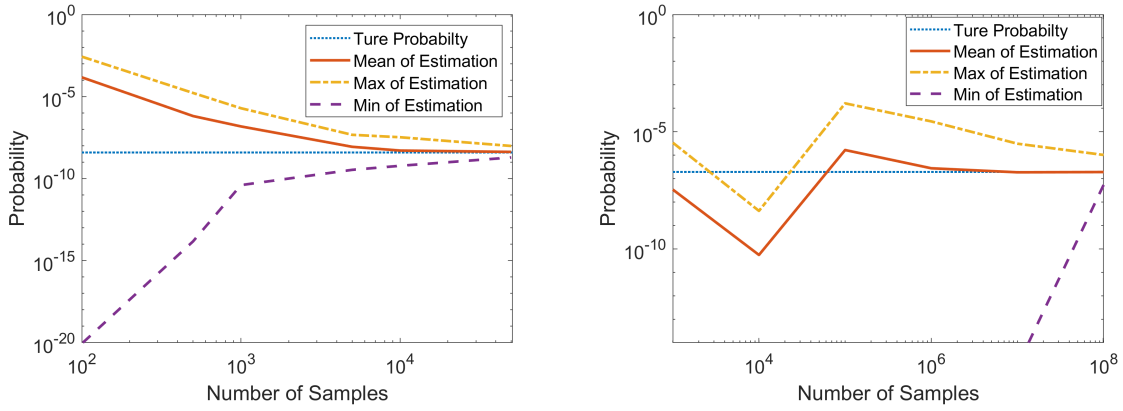


Figure 4: The estimation performance with different number of samples based on 100 replications. (a) Gaussian distribution, $\gamma = 100$. (b) Generalized Pareto distribution, $\gamma = 100$.

Table 1: The coverage and the width of plain bootstrap confidence interval. The results are computed from 30 replications. The rare event problem is defined by the sum of standard Gaussian variables.

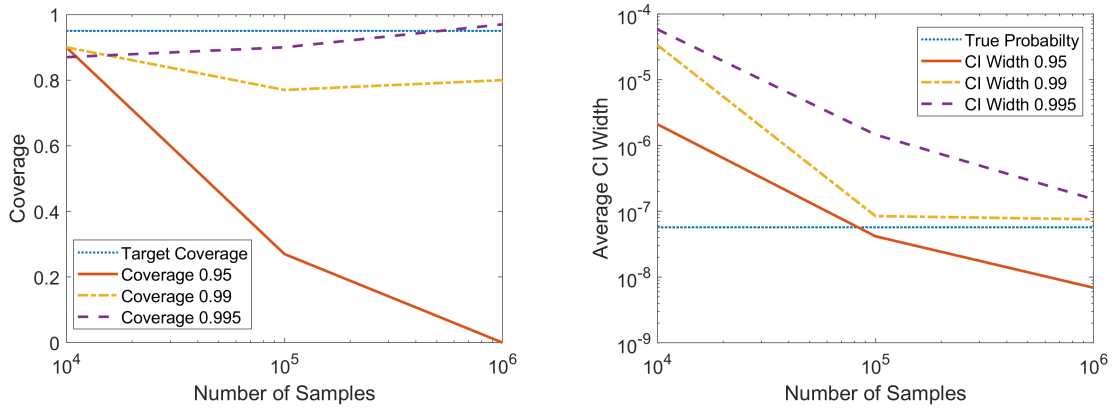
	$\gamma = 70$, $p = 2.6561 \times 10^{-5}$		$\gamma = 100$, $p = 3.882 \times 10^{-9}$		$\gamma = 115$, $p = 1.5734 \times 10^{-11}$	
Samples	Coverage	CI Width	Coverage	CI Width	Coverage	CI Width
100	0.9	0.4019	0.9	0.07700	0.9	0.024
500	0.93	0.0203	0.93	2.75×10^{-4}	0.93	1.20×10^{-5}
1000	0.97	0.0051	0.97	1.89×10^{-5}	0.97	4.25×10^{-7}
5000	0.97	2.33×10^{-4}	0.97	1.155×10^{-7}	0.97	9.36×10^{-10}

distribution with $\xi = 0.2$. We examine whether the confidence intervals constructed from the bootstrap scheme provide the target coverage (95% in our experiment).

Tables 1 and 2 suggest that the bootstrap works well in light-tailed problems, but fails in heavy-tail problems. This ties to our explanation in Section 2 that the impact from tail uncertainty is more profound in the heavy-tail case, which cannot be captured through the standard bootstrap. Table 1 shows that, for light-tail problems., the coverages of the confidence intervals are above 90% in all the considered cases (different numbers of samples and rarity levels). Note that when the number of samples is small (e.g. 100)

Table 2: The coverage and the width of bootstrap confidence interval. The results are computed from 100 replications. The problem is defined by the sum of generalized Pareto variables. The true probability is 1.9195×10^{-7}. "# of 0 Width" presents the number of replications with 0 confidence interval width.

Sample Size	Coverage	CI Width (exclude 0)	# of 0 Width
10^4	0.02	1.15×10^{-5}	97
10^6	0.02	2.40×10^{-6}	80
10^7	0.04	2.00×10^{-7}	3



(a) (b)

Figure 5: The confidence interval coverage and width of the generalized Pareto tail bootstrap scheme (fitted using MLE) on the problem with $p = 5.7095 \times 10^{-8}$. (a) Coverage. (b) Average CI Width.

the confidence interval width is relatively big compared to the estimated probability (0.4 to 0.024 when $p = 2.66 \times 10^{-5}$ to 1.57×10^{-11}). Though this could be pessimistic, these wide intervals successfully detect the unreliable probability estimate (see experiments in Section 3.2).

On the other hand, Table 2 shows that the coverage from the standard bootstrap are close to 0 in all considered cases, including the case of using 10^7 samples to simulate a rare-event probability of order 10^{-7}. Also, the last column shows that when the sample size is smaller than 10^6, most of the constructed confidence intervals have a 0 width. This suggests that in heavy-tail problems, the lack of tail information not only causes problems in estimating the probability itself, but can also deem the assessment of input uncertainty very challenging.

3.4 Bootstrap using Generalized Paerto Distribution

Lastly, we attempt to overcome the challenges in Section 3.3 by fitting a generalized Pareto distribution to the tail of the data. We then run the bootstrap similar to before taking into the fitted tail from each resample. We again consider different truncation points with 0.05, 0.01 and 0.005 tail quantile in our experiment. Here we use X_i with a t-distribution with degree of freedom $\nu = 4$. The considered numbers of samples (from 10^4 to 10^6) are not enough for standard bootstrap to work (see Table 2). Similar to the experiments in Section 3.3, each confidence interval is calculated from 100 bootstrap size, and the reported results are based on 30 experiments. To fit the generalized Pareto, we implement the maximum likelihood estimation (MLE), the method-of-moments (MOM) and probability-weighted moments (PWM) (Castillo and Hadi 1997; Hosking and Wallis 1987).

The experiment results show that although the overall performance of this Pareto tail bootstrap scheme is better than the standard bootstrap, the obtained confidence interval can still be misleading. The latter is caused by the model biasedness from the generalized Pareto in finite sample that the bootstrap cannot overcome. As shown in Figure 5, the coverage could drop to 0 as we increase the number of samples.

Table 3: The coverage and the width of confidence interval from bootstrap using generalized Pareto distribution. "# Spl" represents the number of samples and "Tail Qtl" represents the tail quantile of the truncation points. The problem has a rare event probability $p = 5.7095 \times 10^{-8}$.

	# Spl	10^4		10^5		10^6	
Tail Qtl	Method	Coverage	CI Width	Coverage	CI Width	Coverage	CI Width
0.05	MLE	0.9	2.10×10^{-6}	0.27	4.17×10^{-8}	0	6.89×10^{-9}
	MOM	0.67	1.23×10^{-6}	0.53	4.89×10^{-7}	0.30	3.19×10^{-7}
	PWM	0.87	1.81×10^{-6}	0.30	4.57×10^{-8}	0	7.04×10^{-9}
0.01	MLE	0.90	3.36×10^{-5}	0.77	8.46×10^{-7}	0.80	7.56×10^{-8}
	MOM	0.60	1.26×10^{-6}	0.70	8.10×10^{-7}	0.77	5.61×10^{-7}
	PWM	0.90	1.09×10^{-5}	0.77	8.13×10^{-7}	0.77	8.84×10^{-8}
0.005	MLE	0.87	5.81×10^{-5}	0.90	1.46×10^{-6}	0.97	1.42×10^{-7}
	MOM	0.63	1.17×10^{-6}	0.67	8.43×10^{-7}	0.87	6.10×10^{-7}
	PWM	0.87	1.25×10^{-5}	0.83	1.58×10^{-6}	0.97	1.70×10^{-7}

This is because when the interval width shrinks as the number of samples increases, the model biasedness starts to surface.

Among the three approaches for fitting generalized Pareto distribution, MLE and PWM turn out to be more reliable than MOM (see Table 3, where the coverage of MOM is less than the other two approaches in most cases). The performance matches the documented fact that MOM is unreliable when the shape parameter $\xi > 0.2$. When the sample size is smaller (with 10^4 samples), PWM gives a smaller average confidence interval width than MLE, while providing similar coverage (e.g. with 0.01 tail quantile the widths are 3.36×10^{-5} for MLE and 1.09×10^{-5} for PWM). It therefore suggests that PWM is more suitable for smaller samples. When the sample size is large (10^6), MLE has an upper hand in terms of confidence interval width (e.g. with 0.005 tail quantile the widths are 1.42×10^{-7} for MLE and 1.70×10^{-7} for PWM).

ACKNOWLEDGMENTS

We gratefully acknowledge support from the National Science Foundation under grants CAREER CMMI-1653339/1834710 and IIS-1849280.

REFERENCES

Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*, Volume 57. Springer Science & Business Media.

Atar, R., K. Chowdhary, and P. Dupuis. 2015. "Robust Bounds on Risk-sensitive Functionals via Rényi Divergence". *SIAM/ASA Journal on Uncertainty Quantification* 3(1):18–33.

Barton, R., S. Chick, R. Cheng, S. Henderson, A. Law, B. Schmeiser, L. Leemis, L. Schruben, and J. Wilson. 2002. "Panel Discussion on Current Issues in Input Modeling". In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Ycesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 353–369. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Barton, R. R. 2012. "Input Uncertainty in Outout Analysis". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. Uhrmacher, 67–78. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Barton, R. R., and L. W. Schruben. 2001. "Resampling Methods for Input Modeling". In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 372–378. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Ben-Tal, A., D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. 2013. "Robust Solutions of Optimization Problems Affected by Uncertain Probabilities". *Management Science* 59(2):341–357.

Blanchet, J., C. Dolan, and H. Lam. 2014. "Robust Rare-event Performance Analysis with Natural Non-convex Constraints". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O.

Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 595–603. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Blanchet, J., and K. R. Murthy. 2016. "On Distributionally Robust Extreme Value Analysis". *arXiv preprint arXiv:1601.06858*.

Bucklew, J. 2013. *Introduction to Rare Event Simulation*. Springer Science & Business Media.

Castillo, E., and A. S. Hadi. 1997. "Fitting the Generalized Pareto Distribution to Data". *Journal of the American Statistical Association* 92(440):1609–1620.

Chick, S. E. 2006. "Bayesian Ideas and Discrete Event Simulation: Why, What and How". In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 96–105. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Clarke, F. H. 1975. "Generalized Gradients and Applications". *Transactions of the American Mathematical Society* 205:247–262.

Delage, E., and Y. Ye. 2010. "Distributionally Robust Optimization under Moment Uncertainty with Application to Data-driven Problems". *Operations Research* 58(3):595–612.

Dembo, A., and O. Zeitouni. 1998. *Large Deviations Techniques and Applications*. 2nd ed. Springer Science & Business Media.

Denisov, D., A. B. Dieker, V. Shneer et al. 2008. "Large Deviations for Random Walks Under Subexponentiality: the Big-jump Domain". *The Annals of Probability* 36(5):1946–1991.

Dhara, A., B. Das, and K. Natarajan. 2017. "Worst-Case Expected Shortfall with Univariate and Bivariate Marginals". *arXiv preprint arXiv:1701.04167*.

Dupuis, P., M. A. Katsoulakis, Y. Pantazis, and L. Rey-Bellet. 2018. "Sensitivity Analysis for Rare Events based on Rényi Divergence". *arXiv preprint arXiv:1805.06917*.

Embrechts, P., R. Frey, and A. McNeil. 2005. "Quantitative Risk Management". *Princeton Series in Finance, Princeton* 10.

Embrechts, P., C. Klüppelberg, and T. Mikosch. 2013. *Modelling Extremal Events: For Insurance and Finance*, Volume 33. Springer Science & Business Media.

Embrechts, P., and G. Puccetti. 2006. "Bounds for Functions of Multivariate Risks". *Journal of Multivariate Analysis* 97(2):526–547.

Embrechts, P., G. Puccetti, and L. Rüschendorf. 2013. "Model Uncertainty and VaR Aggregation". *Journal of Banking & Finance* 37(8):2750–2764.

Ghaoui, L. E., M. Oks, and F. Oustry. 2003. "Worst-case Value-at-risk and Robust Portfolio Optimization: a Conic Programming Approach". *Operations Research* 51(4):543–556.

Glasserman, P., and X. Xu. 2014. "Robust Risk Measurement and Model Risk". *Quantitative Finance* 14(1):29–58.

Hansen, L. P., and T. J. Sargent. 2008. *Robustness*. Princeton university press.

Henderson, S. G. 2003. "Input Model Uncertainty: Why Do We Care and What Should We Do About It?". In *Proceedings of the 2003 Winter Simulation Conference*, edited by S. Chick, P. J. Snchez, D. Ferrin, and D. J. Morrice, Volume 1, 90–100. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hosking, J. R., and J. R. Wallis. 1987. "Parameter and Quantile Estimation for the Generalized Pareto Distribution". *Technometrics* 29(3):339–349.

Jelenkovic, P. R. 1999. "Network Multiplexer with Truncated Heavy-tailed Arrival Streams". In *Proceedings of the Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, Volume 2, 625–632. IEEE.

Lam, H. 2016. "Advanced Tutorial: Input Uncertainty and Robust Analysis in Stochastic Simulation". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, and E. Zhou, 178–192. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lam, H., and C. Mottet. 2015. "Simulating Tail Events with Unspecified Tail Models". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 392–402. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lam, H., and C. Mottet. 2017. "Tail Analysis without Parametric Models: A Worst-case Perspective". *Operations Research* 65(6):1696–1711.

Leadbetter, M. R. 1991. "On a Basis for Peaks over Threshold Modeling". *Statistics & Probability Letters* 12(4):357–362.

Li, B., R. Jiang, and J. L. Mathieu. 2019. "Ambiguous Risk Constraints with Moment and Unimodality Information". *Mathematical Programming* 173(1-2):151–192.

Lim, A. E., and J. G. Shanthikumar. 2007. "Relative Entropy, Exponential Utility, and Robust Dynamic Pricing". *Operations Research* 55(2):198–214.

Nagaev, S. V. et al. 1979. "Large Deviations of Sums of Independent Random Variables". *The Annals of Probability* 7(5):745–789.

Nakayama, M. K. 1995. "Asymptotics of Likelihood Ratio Derivative Estimators in Simulations of Highly Reliable Markovian Systems". *Management Science* 41(3):524–554.

Nakayama, M. K. 1998. "On Derivative Estimation of The Mean Time to Failure in Simulations of Highly Reliable Markovian Systems". *Operations Research* 46(2):285–290.

Nelson, B. L., A. T. K. Wan, S. Fan, and X. Zhang. 2019. "Reducing Simulation Input-model Risk via Input Model Averaging". *working paper*.

Olvera-Cravioto, M. 2006. "The Single-server Queue with Heavy Tails". *PhD dissertation*.

Petersen, I. R., M. R. James, and P. Dupuis. 2000. "Minimax Optimal Control of Stochastic Uncertain Systems with Relative Entropy Constraints". *IEEE Transactions on Automatic Control* 45(3):398–412.

Puccetti, G., and L. Rüschendorf. 2012. "Computation of Sharp Bounds on The Distribution of a Function of Dependent Risks". *Journal of Computational and Applied Mathematics* 236(7):1833–1840.

Puccetti, G., and L. Rüschendorf. 2013. "Sharp Bounds for Sums of Dependent Risks". *Journal of Applied Probability* 50(01):42–53.

Rhee, C.-H., J. Blanchet, and B. Zwart. 2016. "Sample Path Large Deviations for Lèvy Processes and Random Walks with Regularly Varying Increments". *arXiv preprint arXiv:1606.02795*.

Song, E., B. L. Nelson, and C. D. Pegden. 2014. "Advanced Tutorial: Input Uncertainty Quantification". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 162–176. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Van Parys, B. P., P. J. Goulart, and M. Morari. 2019. "Distributionally Robust Expectation Inequalities for Structured Distributions". *Mathematical Programming* 173(1-2):251–280.

Wang, B., and R. Wang. 2011. "The Complete Mixability and Convex Minimization Problems with Monotone Marginal Densities". *Journal of Multivariate Analysis* 102(10):1344–1360.

Wiesemann, W., D. Kuhn, and M. Sim. 2014. "Distributionally Robust Convex Optimization". *Operations Research* 62(6):1358–1376.

Yuen, R., S. Stoev, and D. Cooley. 2019. "Distributionally Robust Inference for Extreme Value-at-risk". *working paper*.

AUTHOR BIOGRAPHIES

ZHIYUAN HUANG is a fourth-year Ph.D. student in the Department of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. His research interests include simulation and stochastic optimization. His email address is zhyhuang@umich.edu.

HENRY LAM is an Associate Professor in the Department of Industrial Engineering and Operations Research at Columbia University. His research focuses on Monte Carlo simulation, uncertainty quantification, risk analysis, and stochastic and robust optimization. His email address is khl2114@columbia.edu.