Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

RARE-EVENT SIMULATION WITHOUT VARIANCE REDUCTION: AN EXTREME VALUE THEORY APPROACH

Yuanlu Bai
Henry Lam

Industrial Engineering and Operations Research
Columbia University
500 West 120th Street
New York, NY 10027, USA

Sebastian Engelke

Research Center for Statistics
University of Geneva
Boulevard du Pont d'Arve 40
1205 Geneva, SWITZERLAND

ABSTRACT

In estimating probabilities of rare events, crude Monte Carlo (MC) simulation is inefficient which motivates the use of variance reduction techniques. However, these latter schemes rely heavily on delicate analyses of underlying simulation models, which are not always easy or even possible. We propose the use of extreme value analysis, in particular the peak-over-threshold (POT) method which is popularly employed for extremal estimation of real datasets, in the simulation setting. More specifically, we view crude MC samples as data to fit on a generalized Pareto distribution. We test this idea on several numerical examples. The results show that our POT estimator appears more accurate than crude MC and, while crude MC can easily give a trivial probability estimate 0, POT outputs a non-trivial estimate with a roughly correct magnitude. Therefore, in the absence of efficient variance reduction schemes, POT appears to offer potential benefits to enhance crude MC estimates.

1 INTRODUCTION

A major goal of rare-event simulation is to estimate tiny probabilities that are triggered by rare but catastrophic events (Bucklew 2004; Juneja and Shahabuddin 2006; Rubino and Tuffin 2009). This problem has been of wide interest to various application areas such as queueing systems (Dupuis et al. 2007; Dupuis and Wang 2009; Blanchet et al. 2009; Blanchet and Lam 2014; Kroese and Nicola 1999; Ridder 2009; Sadowsky 1991; Szechtman and Glynn 2002), communication networks (Kesidis et al. 1993), finance (Glasserman 2003; Glasserman and Li 2005; Glasserman et al. 2008) and insurance (Asmussen 1985; Asmussen and Albrecher 2010). In recent years, with the extensive development of machine learning and artificial intelligence, rare-event simulation is also applied to evaluate the robustness of machine learning predictors (Webb et al. 2018; Bai et al. 2022) or quantify the risk of intelligent physical systems (Huang et al. 2017; O'Kelly et al. 2018; Zhao et al. 2016; Zhao et al. 2017; Arief et al. 2021). In using Monte Carlo (MC) to estimate rare-event probabilities, a main challenge is that, by its own nature, the target rare events seldom occur in the simulation experiments. Since sufficient hits on the target events are required to achieve meaningfully accurate estimation, this makes crude MC computationally costly as the required simulation size to attain enough accuracy becomes enormous.

To address the inefficiency of crude MC, a range of variance reduction techniques have been developed. These techniques aim to alter the naive sampling procedure of crude MC to improve the error per simulation run, thus attaining adequate accuracy with less number of samples. Among them, importance sampling (IS) (Siegmund 1976; Sadowsky and Bucklew 1990; Juneja and Shahabuddin 2006) is one of the most popular methods. The idea is to utilize a change of measure to amplify the frequency of target events in the simulation, and then to correct this bias with the likelihood ratio. Conditional Monte Carlo (CMC)

978-1-6654-7661-4/22/$31.00 ©2022 IEEE

(Asmussen and Glynn 2007; Rubinstein and Kroese 2016) uses the conditional probability of the target events on certain information as an unbiased estimator, which is especially useful in the heavy-tailed case where the classical exponential tilting IS technique cannot be applied. Multi-level splitting or subset simulation (Glasserman et al. 1999; Villén-Altamirano and Villén-Altamirano 1994; Au and Beck 2001) chooses a sequence of nested subsets, factorizes the target probability into the product of conditional probabilities and then estimates each conditional probability by generating samples from the corresponding conditional distribution. Despite demonstratably powerful in many problems, in order to attain good performances, the variance reduction techniques described above often rely on tractable problem structures that allow careful algorithmic design. Unfortunately, this requirement could be difficult or even impossible to meet in complex practical applications. Thus, the main goal of this paper is to study *an approach to improve upon the efficiency of crude MC in the absence of variance reduction schemes*.

More specifically, we resort to extreme value theory (EVT) (Embrechts et al. 1997), which has been a prominent approach in extreme event analysis for real data. There, the main challenge is the scarcity of observations in the tail portion of the data that can directly infer the distributional extreme, and thus one needs to suitably extrapolate information from the "central" part of the distribution. One major approach, which we would borrow here, is the peak-over-threshold (POT) (Smith 1984) method. This method is based on the Pickands-Balkema-de Haan theorem (Balkema and De Haan 1974; Pickands III 1975) which states that, under suitable assumptions, the distribution function of the so-called excess loss above a threshold converges to the generalized Pareto distribution (GPD) as this threshold increases. In this sense, GPD is a justified model for tail data fitting. Analogically, we propose applying the POT method on the crude MC simulation data. That is, within an acceptable computational budget, we use the crude MC simulator to generate a simulation dataset, treat it as a real dataset, and fit the tail portion of the data with GPD. The target probability is then estimated with the fitted GPD. Although the POT method is well-established and this idea seems simple, to our best knowledge, POT has not been considered in the rare-event simulation context. As discussed above, the latter literature mainly focuses on developing algorithms to improve the efficiency of crude MC by utilizing tractable problem structures.

We apply this method to several numerical examples and compare the results with crude MC, including examples where variance reduction techniques such as IS cannot be applied easily. We find that, in a suitably wide parameter range, our POT estimator achieves smaller variance than crude MC. Moreover, with limited simulation samples, while crude MC often outputs a trivial estimate of 0, POT can output an estimate of a roughly correct magnitude.

The rest of this paper is organized as follows. Section 2 introduces relevant backgrounds about the challenge of crude MC and the theory of POT. Section 3 demonstrates the numerical results of several experiments. Finally, Section 4 concludes and discusses future work.

2 BACKGROUND

2.1 Challenge of Crude MC

We consider a simulation model that outputs a random vector $X \in \mathbb{R}^d$ under probability measure P, and we are interested in estimating $p = P(X \in E)$ where $E \subset \mathbb{R}^d$ is a rare-event set, i.e., p is a tiny number. Suppose we generate n i.i.d. simulation samples $X_1, \dots, X_n$. Then the crude MC estimator is simply $\hat{p}_n^{MC} = \frac{1}{n}\sum_{i=1}^n I(X_i \in E)$. In this rare-event setting, we would like to control the discrepancy between $\hat{p}_n^{MC}$ and p, *relative* to the magnitude of p itself. In other words, we want to control the probability $P(|\hat{p}_n^{MC} - p| > \delta p)$ for some fixed $\delta < 1$.

Note that the crude MC estimator is clearly unbiased. By Chebyshev's inequality, we have for any $\delta > 0$,

$$P(|\hat{p}_n^{MC} - p| > \delta p) \leq \frac{var(\hat{p}_n^{MC})}{\delta^2 p^2} = \frac{var(I(X \in E))}{n\delta^2 p^2} = \frac{p(1-p)}{n\delta^2 p^2} \leq \frac{1}{n\delta^2 p}.$$

Thus, $n \geq \frac{1}{\varepsilon\delta^2 p}$ guarantees that $P(|\hat{p}_n^{MC} - p| > \delta p) \leq \varepsilon$ for any $\varepsilon > 0$. This reveals that, when p is tiny, the required simulation size which is reciprocal in p can be enormous. Therefore, as widely known in the literature, crude MC is inefficient for rare-event simulation.

Moreover, we note that $\hat{p}_n^{MC}$ can only take values in $0, 1/n, \ldots, (n-1)/n, 1$. In particular, $P(\hat{p}_n^{MC} = 0) = (1-p)^n \approx e^{-np}$. Thus, when n is not sufficiently large relative to p, there is a significant chance that the crude MC estimate is trivially 0.

As discussed in the introduction, we are interested in scenarios where crude MC is the only available simulation format. In this case, we study POT to obtain a better estimate from crude MC without overwhelming the computational effort.

2.2 Recap of POT

Now we briefly recap the background for POT. Define the generalized extreme value distribution (GEV) as

$$H_\xi(x) = \begin{cases} \exp\{-(1+\xi x)^{-1/\xi}\} & \text{if } \xi \neq 0; \\ \exp\{-\exp\{-x\}\} & \text{if } \xi = 0. \end{cases}$$

We say a distribution function F belongs to the maximum domain of attraction of H_ξ if there exist constants $c_n > 0, d_n \in \mathbb{R}$ such that $c_n^{-1}(M_n - d_n) \xrightarrow{d} H_\xi$ where M_n denotes the maximum of n i.i.d. samples from F. In this case we write $F \in MDA(H_\xi)$. This property is satisfied by a wide variety of distributions, including Cauchy, Pareto, Loggamma, Uniform, Beta, Exponential, Weibull, Gamma, Normal, Lognormal, etc.

Suppose that Y is a random variable with distribution function F. The right endpoint is defined as $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$. Here, x_F can be either infinite or finite. For a fixed threshold $u < x_F$, we call $F_u(x) = P(Y - u \leq x | Y > u), x \geq 0$ the excess distribution function of Y over the threshold u. Finally, we define the GPD as

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi} & \text{if } \xi \neq 0; \\ 1 - \exp\left\{-\frac{x}{\beta}\right\} & \text{if } \xi = 0 \end{cases}$$

where $\xi \in \mathbb{R}, \beta > 0$. The support of the distribution is $[0, \infty)$ if $\xi \geq 0$ and $[0, -\beta/\xi]$ if $\xi < 0$.

We have the following theorem (see, e.g., Embrechts et al. 1997):

Theorem 1 (Pickands–Balkema–de Haan Theorem) For any $\xi \in \mathbb{R}$, $F \in MDA(H_\xi)$ if and only if

$$\lim_{u \uparrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0$$

for some positive function β.

The theorem shows that under a mild assumption, the excess distribution converges to a GPD with certain parameters as the threshold increases. Therefore, given i.i.d. samples $Y_1, \ldots, Y_n$ from F, we can pick a large threshold u, fit the data exceeding u with GPD, and then use the fitted GPD to estimate the target tail quantity. This is called the POT method.

In this paper, we assume that the target rare event $\{X \in E\}$ can be formulated as $\{f(X) \geq a\}$ where f is a real-valued function and a is a large constant. Under this setting, we estimate the target probability p with the following procedure:

1. Generate i.i.d. simulation samples $X_1, \ldots, X_n$ and compute $Y_i = f(X_i), i = 1, \ldots, n$;
2. Pick a threshold $u < a$ following certain criterion;
3. Fit a GPD $G_{\hat{\xi},\hat{\beta}}$ with the excess data $\{Y_i - u : 1 \leq i \leq n, Y_i > u\}$;
4. Output $\hat{p}_n^{POT} = \left(\frac{1}{n}\sum_{i=1}^n I(Y_i > u)\right)\left(1 - G_{\hat{\xi},\hat{\beta}}(a-u)\right)$.

As discussed above, the theoretical guarantee for this POT procedure requires that the distribution function of $Y = f(X)$ belongs to $MDA(H_\xi)$ for some $\xi \in \mathbb{R}$. Though this assumption is not verifiable in practice, thanks to its generality the POT method is often applied to real datasets in the extreme event analysis literature. In this paper, we follow this reasoning and apply this method on simulation data.

We note that after generating some crude MC simulation samples and computing a crude MC estimate, we can always reuse these samples and apply the above procedure to obtain a POT estimate with little additional effort. Therefore, the POT estimate can always serve as a complement to the crude MC estimate, especially when the latter is trivially 0. In the next section, we will investigate under what conditions the POT estimator works well through multiple experiments.

3 NUMERICAL EXPERIMENTS

3.1 Example 1: Sample Mean

We start from an easy example. Suppose that $X \in \mathbb{R}^d$, the simulation output, consists of d i.i.d. random variables. That is, $X^{(1)}, \ldots, X^{(d)}$ are i.i.d. from certain distribution where $X^{(j)}$ denotes the j-th element of the vector X. We consider $f(X) = \frac{1}{d}\sum_{j=1}^{d} X^{(j)}$, and hence we aim to estimate $P\left(\frac{1}{d}\sum_{j=1}^{d} X^{(j)} \geq a\right)$. With i.i.d. simulation samples $X_1, \ldots, X_n$, the crude MC estimator is simply $\hat{p}_n^{MC} = \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{d}\sum_{j=1}^{d} X_i^{(j)} \geq a\right)$.Alternatively, we can use the POT procedure described in Section 2.2 to compute $\hat{p}_n^{POT}$.

In order to apply POT, we need to select the threshold u properly, which involves an intrinsic bias-variance tradeoff. Intuitively, if u is too small, then approximating the excess distribution with GPD may bear large bias; on the other hand, if u is too large, then there are too few data exceeding this threshold, which may result in large variance in fitting the GPD parameters. In the literature, some useful tools are proposed to facilitate this selection, such as the mean residual plot (Embrechts et al. 1997), but its performance is arguably still case-by-case. In our experiments, for simplicity, we choose u as a certain sample quantile such that it is large but there are still sufficient data above it. In this sample mean example, we set $d = 10, n = 10^6$ and we compare the performance of u being the 0.99/0.999/0.9999-th sample quantile. For instance, if u is chosen as the 0.999-th sample quantile, then there are still $10^6 \times (1 - 0.999) = 1000$ samples above this threshold to help us fit the GPD parameters. Moreover, it remains an open problem what is the best way to fit the GPD parameters. Throughout this paper, we use maximum likelihood estimation.

First, we test the methods on a light-tailed distribution, more specifically, normal distribution. We let $X_i^{(j)} \sim N(0,1)$. Under this setting, a highly efficient IS scheme is available. That is, we sample $\tilde{X}_1, \ldots, \tilde{X}_n$ such that $\tilde{X}_i^{(j)}$'s are i.i.d. from $N(a,1)$ instead of $N(0,1)$. Then the IS estimator is computed by $\hat{p}_n^{IS} = \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{1}{d}\sum_{j=1}^{d} \tilde{X}_i^{(j)} \geq a\right) \exp\left\{-a\sum_{j=1}^{d} \tilde{X}_i^{(j)} + da^2/2\right\}$.We note that $Y = f(X) \sim N(0, 1/d)$, so actually the true probability $p = P(Y \geq a)$ is known, which helps us compare the performance of all these estimators and evaluate the GPD fitting performance.

In order to evaluate these estimators, we generate n simulation samples, compute each estimator, and repeat this process $N = 1000$ times. Table 1 shows the descriptive statistics of these 1000 estimations with $a = 1.5$, where the true probability $p = 1.05 \times 10^{-6}$. We see that the carefully designed IS estimator is highly accurate, i.e. the mean is nearly the truth and the standard deviation is small relative to the truth. The mean of the crude MC estimates is close to the truth, which is reasonable since it is actually the mean value of $n \times N = 10^9$ simulation results. However, the standard deviation of the crude MC estimator is nearly the same as the true probability. In fact, we see that at least 25% of the crude MC estimates are trivially 0. For the POT method, we see that POT-0.99 (i.e. u is chosen as the 0.99 sample quantile) has large bias but small variance (or standard deviation); POT-0.999 has small bias and the variance is smaller than crude MC; and POT-0.9999 is more unstable than POT-0.999 as it has larger variance and also gives 0 occasionally. Overall, POT-0.999 appears the best among the three.

Table 1: Statistics of each estimator in Example 1 with $N(0,1)$ distribution and $a = 1.5$. $p = 1.05 \times 10^{-6}$.

method	mean	std	min	25%	50%	75%	max
MC	1.13E-06	1.05E-06	0	0	1.00E-06	2.00E-06	6.00E-06
IS	1.05E-06	2.39E-09	1.04E-06	1.05E-06	1.05E-06	1.05E-06	1.06E-06
POT-0.99	5.88E-07	2.74E-07	6.34E-08	3.88E-07	5.52E-07	7.36E-07	1.95E-06
POT-0.999	1.01E-06	6.14E-07	1.71E-08	5.47E-07	8.79E-07	1.38E-06	3.48E-06
POT-0.9999	1.00E-06	7.53E-07	0	4.05E-07	8.77E-07	1.48E-06	3.77E-06

Although POT-0.999 still cannot compete with IS, it indeed performs better than crude MC using the same simulation samples. To achieve the same standard deviation as POT-0.999, crude MC requires $n' = \lceil p(1-p)/std^2 \rceil \approx 2.79 \times 10^6$, which is almost three times the actual simulation size. Moreover, as mentioned before, crude MC estimator gives 0 frequently, while POT-0.999 can at least give an estimate of a roughly correct magnitude. To illustrate this difference better, we plot the histograms of these two estimators in Figure 1, where we find that POT-0.999 is more concentrated around the true probability while over 30% of the crude MC estimates are 0.

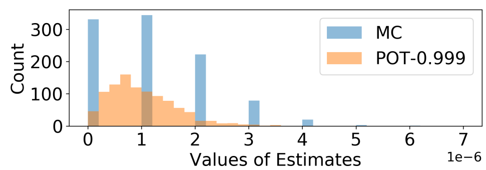


Figure 1: Histograms of crude MC and POT-0.999 estimates in Example 1 with $N(0,1)$ distribution and $a = 1.5$.

Now we move a to 2, so the target event is even rarer ($p = 1.27 \times 10^{-10}$). Table 2 shows the descriptive statistics. In this case, since p is too small, crude MC completely fails, in the sense that all the 1000 estimates are 0. The IS estimator still performs well. We find that the POT estimators cannot perform well. In order to explain this phenomenon, in Figure 2, we plot the fitted GPD tails in five random replications against the true tail. We find that when a is not far from the threshold u (approximately 0.7, 1.0, 1.2 respectively for the 0.99, 0.999, 0.9999-th sample quantiles), the fitted values are generally close to the true values. When a increases, the fitting becomes more and more inaccurate and unstable, which coincides with our observations on $a = 1.5$ and $a = 2$. This phenomenon is reasonable, considering that in the POT method, we are leveraging the limited data above u, most of which should not be far from u, to infer the whole tail. That is, POT may not be suitable if a is too far from the threshold u.

Table 2: Statistics of each estimator in Example 1 with $N(0,1)$ distribution and $a = 2$. $p = 1.27 \times 10^{-10}$.

method	mean	std	min	25%	50%	75%	max
MC	0	0	0	0	0	0	0
IS	1.27E-10	3.37E-13	1.26E-10	1.27E-10	1.27E-10	1.27E-10	1.28E-10
POT-0.99	5.72E-13	3.68E-12	0	0	1.11E-18	7.58E-15	7.49E-11
POT-0.999	5.65E-10	2.25E-09	0	3.19E-18	1.73E-12	1.99E-10	3.77E-08
POT-0.9999	5.29E-09	1.72E-08	0	0	1.83E-12	1.95E-09	1.94E-07

We also investigate the performance in the heavy-tailed case. Let $X_i^{(j)} \sim Pareto(t,\alpha)$ where $t,\alpha > 0$ are the scale and shape parameters. That is, the CDF is $P(X_i^{(j)} \leq x) = 1 - (t/x)^\alpha, x \geq t$. In this case, IS cannot

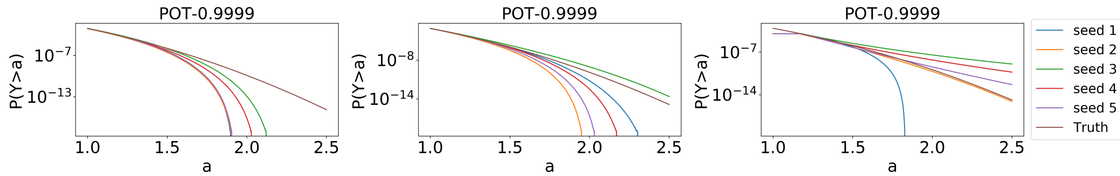


Figure 2: GPD fitting performance in Example 1 with $N(0,1)$ distribution.

be easily applied. Instead, the CMC estimator defined by $\hat{p}_n^{CMC} = \frac{1}{n}\sum_{i=1}^{n} nP(X_i^{(d)} > (\max_{j=1,\dots,d-1} X_i^{(j)}) \vee (da - \sum_{j=1}^{d-1} X_i^{(j)}))$ is known to be unbiased and efficient. We still repeat each estimator $N = 1000$ times to make a comparison.

Tables 3 to 6 show the descriptive statistics where $t = 1$ and $\alpha = 1.5, 2, 3, 5$. In these experiments, a is tuned manually such that the true probability is close to 10^{-7}, and hence we could compare the influence of the shape parameter α. In all the four experiments, the CMC estimator is highly accurate, so we use the mean of CMC as an approximate to the truth. By contrast, the crude MC estimator performs badly. In fact, in every experiment, over 75% of the crude MC estimates are 0. Among the three POT estimators, POT-0.999 still performs generally well, as the mean is close to the truth and the standard deviation is much smaller than crude MC. From the histograms in Figure 3, we see that crude MC estimates are almost all 0 while POT-0.999 estimates are more concentrated. Finally, we note that smaller α implies heavier tail, and we observe from these results that POT works better in this case. Especially, when $\alpha = 1.5$, the standard deviation of POT-0.999 is only 6.32×10^{-8}, while it requires $n' = \lceil p(1-p)/std^2 \rceil \approx 2.50 \times 10^7$ samples for crude MC to achieve the same standard deviation, which is 25 times the actual n.

Table 3: Statistics of each estimator in Example 1 with $Pareto(1, 1.5)$ distribution and $a = 21601.56$.

method	mean	std	min	25%	50%	75%	max
MC	9.60E-08	3.21E-07	0	0	0	0	2.00E-06
CMC	9.96E-08	1.07E-13	9.96E-08	9.96E-08	9.96E-08	9.96E-08	9.96E-08
POT-0.99	9.68E-08	2.29E-08	4.54E-08	8.09E-08	9.45E-08	1.11E-07	1.88E-07
POT-0.999	1.09E-07	6.32E-08	1.35E-08	6.44E-08	9.58E-08	1.39E-07	4.46E-07
POT-0.9999	1.78E-07	1.75E-06	1.11E-20	1.58E-08	4.64E-08	1.11E-07	5.13E-05

Table 4: Statistics of each estimator in Example 1 with $Pareto(1, 2)$ distribution and $a = 1000$.

method	mean	std	min	25%	50%	75%	max
MC	9.70E-08	3.22E-07	0	0	0	0	2.00E-06
CMC	1.00E-07	2.98E-13	1.00E-07	1.00E-07	1.00E-07	1.00E-07	1.00E-07
POT-0.99	9.47E-08	2.54E-08	4.04E-08	7.65E-08	9.23E-08	1.10E-07	2.25E-07
POT-0.999	1.12E-07	7.22E-08	8.15E-09	6.10E-08	9.41E-08	1.45E-07	5.01E-07
POT-0.9999	1.36E-07	1.66E-07	0	2.17E-08	8.31E-08	1.82E-07	1.52E-06

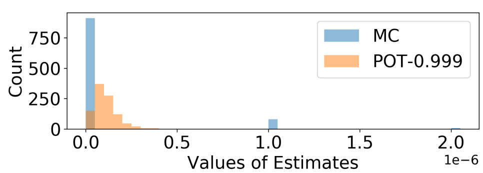
From the observations drawn from this sample mean example, we conclude that POT helps us refine crude MC estimation, especially when the simulation size is relatively large so that there are abundant data above the threshold, the rare-event set boundary a is not far from the threshold, and the tail of the distribution is heavy. In the following subsections, we compare crude MC and POT-0.999 on several other examples to examine the generality of this conclusion.

Table 5: Statistics of each estimator in Example 1 with $Pareto(1,3)$ distribution and $a = 47.625$.

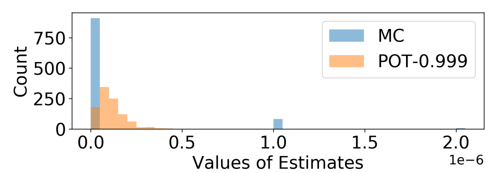
method	mean	std	min	25%	50%	75%	max
MC	9.90E-08	3.25E-07	0	0	0	0	2.00E-06
CMC	1.02E-07	2.00E-12	1.02E-07	1.02E-07	1.02E-07	1.02E-07	1.02E-07
POT-0.99	7.54E-08	2.48E-08	2.27E-08	5.79E-08	7.22E-08	8.98E-08	1.97E-07
POT-0.999	1.20E-07	8.48E-08	3.66E-09	5.82E-08	1.00E-07	1.58E-07	5.07E-07
POT-0.9999	1.40E-07	1.84E-07	0	1.60E-08	7.63E-08	1.93E-07	1.80E-06

Table 6: Statistics of each estimator in Example 1 with $Pareto(1,5)$ distribution and $a = 5.234375$.

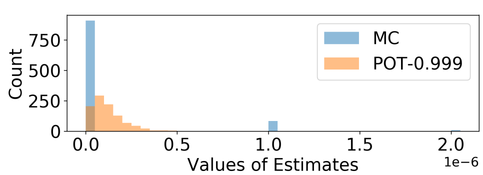
method	mean	std	min	25%	50%	75%	max
MC	8.60E-08	3.01E-07	0	0	0	0	2.00E-06
CMC	1.01E-07	1.57E-11	1.01E-07	1.01E-07	1.01E-07	1.01E-07	1.01E-07
POT-0.99	2.23E-08	1.22E-08	2.71E-09	1.39E-08	1.99E-08	2.76E-08	9.95E-08
POT-0.999	9.55E-08	8.17E-08	2.29E-10	3.81E-08	7.39E-08	1.29E-07	6.42E-07
POT-0.9999	1.29E-07	1.82E-07	0	8.10E-09	5.87E-08	1.79E-07	1.58E-06



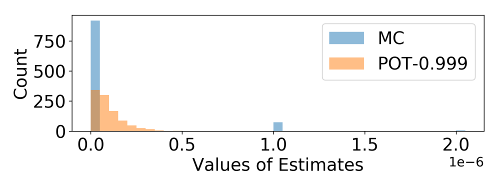
(a) $\alpha = 1.5$



(b) $\alpha = 2$



(c) $\alpha = 3$



(d) $\alpha = 5$

Figure 3: Histograms of crude MC and POT-0.999 estimates in Example 1 with $Pareto(1,\alpha)$ distribution.

3.2 Example 2: Random Walk

In this subsection, we consider the overflow of a random walk process. We still suppose that $X^{(1)},\dots,X^{(d)}$ are i.i.d. random variables. Then the process $\{\sum_{j=1}^{k}\left(X^{(j)} - EX^{(j)}\right) : k = 1,\dots,d\}$ is a random walk where the increment at each step has zero mean. We set $f(X) = \max_{k=1,\dots,d}\sum_{j=1}^{k}\left(X^{(j)} - EX^{(j)}\right)$, and the rare event $\{f(X) \geq a\}$ represents the excursion of this random walk. With n simulation samples $X_1,\dots,X_n$, we can compute $\hat{p}_n^{MC}$ and $\hat{p}_n^{POT}$. We set $d = 10$ and $n = 10^6$.

Like in Section 3.1, we repeat evaluating each estimator for $N = 1000$ times. However, without accurate IS or CMC estimators, now we need to estimate the truth in another way. As mentioned before, the average of the N crude MC estimates, denoted by $\overline{\hat{p}_n^{MC}}$, is actually the average of $n \times N$ simulation samples from all the N repetitions. Therefore, we can pool these $n \times N$ samples to compute a confidence interval (CI). In particular, in this as well as the following subsections, we use the 95% CI given by $\overline{\hat{p}_n^{MC}} \pm 1.96\sqrt{\frac{\overline{\hat{p}_n^{MC}}(1-\overline{\hat{p}_n^{MC}})}{n\times N}}$ as an estimate for the true target probability.

Table 7 presents the descriptive statistics where $X^{(j)} \sim N(0,1)$ and Tables 8 to 10 present the statistics where $X^{(j)} \sim Pareto(1,\alpha)$ for $\alpha = 2,3,5$. The rare-event boundary a is tuned such that the true probability is close to 10^{-6} for better comparison. Figure 4 shows the histograms of the two estimators in these experiments. These results show consistent observations with the previous example. The POT estimator is only slightly biased but the standard deviation is nearly half of that of the crude MC estimator. It also gives non-trivial estimates while about 30% to 40% of the crude MC estimates are 0. Also the POT estimator performs better for heavier-tailed distributions.

Table 7: Statistics of each estimator in Example 2 with $N(0,1)$ distribution and $a = 15$. CI: $1.23\times 10^{-6} \pm 6.86\times 10^{-8}$.

method	mean	std	min	25%	50%	75%	max
MC	1.23E-06	1.09E-06	0	0	1.00E-06	2.00E-06	6.00E-06
POT-0.999	1.14E-06	6.64E-07	2.13E-08	6.23E-07	1.05E-06	1.54E-06	3.99E-06

Table 8: Statistics of each estimator in Example 2 with $Pareto(1,2)$ distribution and $a = 3000$. CI: $1.06\times 10^{-6} \pm 6.40\times 10^{-8}$.

method	mean	std	min	25%	50%	75%	max
MC	1.07E-06	1.09E-06	0	0	1.00E-06	2.00E-06	6.00E-06
POT-0.999	1.15E-06	4.74E-07	2.17E-07	8.03E-07	1.08E-06	1.42E-06	3.15E-06

Table 9: Statistics of each estimator in Example 2 with $Pareto(1,3)$ distribution and $a = 200$. CI: $1.17\times 10^{-6} \pm 6.71\times 10^{-8}$.

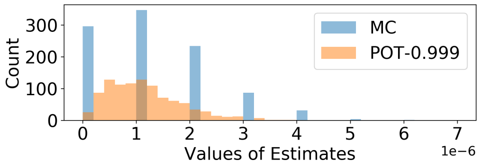
method	mean	std	min	25%	50%	75%	max
MC	1.17E-06	1.14E-06	0	0	1.00E-06	2.00E-06	6.00E-06
POT-0.999	1.29E-06	5.61E-07	1.91E-07	8.74E-07	1.19E-06	1.61E-06	3.27E-06

Table 10: Statistics of each estimator in Example 2 with $Pareto(1,5)$ distribution and $a = 25$. CI: $8.26\times 10^{-7} \pm 5.63\times 10^{-8}$.

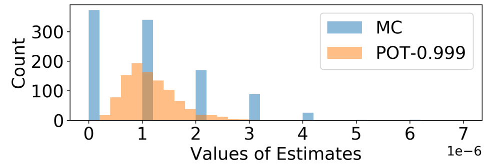
method	mean	std	min	25%	50%	75%	max
MC	8.26E-07	9.23E-07	0	0	1.00E-06	1.00E-06	6.00E-06
POT-0.999	8.72E-07	4.60E-07	6.98E-08	5.39E-07	7.86E-07	1.10E-06	2.75E-06

3.3 Example 3: Neural Network

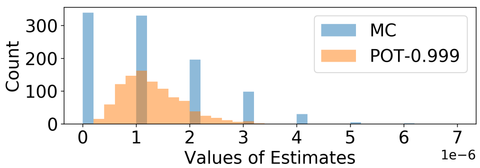
In this example, we consider an example that uses rare-event simulation to evaluate the robustness of neural network classification, a problem that has been recently studied in Webb et al. (2018), Bai et al. (2022). We consider the classification problem on the MNIST dataset which contains 70,000 images of handwritten digits. Each image is encoded as a vector $x \in \mathbb{R}^{784}$, and belongs to a class in $\{1,\ldots,10\}$ (we assume $j = 1,\ldots,9$ corresponds to digit j and $j = 10$ corresponds to digit 0). The goal of the classification task is to predict the class given the input x accurately. We train a simple neural network with an accuracy of over 97%. Given an input image (encoded as a vector) x, this trained neural network outputs the predicted logits



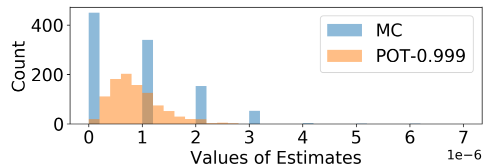
(a) $N(0,1)$



(b) $Pareto(1,2)$



(c) $Pareto(1,3)$



(d) $Pareto(1,5)$

Figure 4: Histograms of crude MC and POT-0.999 estimates in Example 2.

$z_j(x), j=1,\dots,10$, and then this input is classified as $\arg\max_j z_j(x)$. Suppose that the neural network is able to correctly classify the input x_0 to its true class j_0, and we aim to evaluate the robustness of this classification model via simulation. That is, we exert a random perturbation ε from certain distribution on the input and then compute the simulation output $X=(z_j(x_0+\varepsilon))_{j=1,\dots,10}$. Let $f(X)=\max_{j\neq j_0} X^{(j)}-X^{(j_0)}$, and $a=0$. Thus, the target rare event $\{f(X)\geq a\}$ is the event that the classification result changes after the perturbation. We choose the simulation size $n=10^6$ and repeat each estimator $N=1000$ times.

We suppose that each element of the random perturbation ε is i.i.d. with zero mean. For the light-tailed case, we try $N(0,0.17^2)$; for the heavy-tailed case, we try $Pareto(0.0008,2)$ and $Pareto(0.02,3)$ minus the mean. From Tables 11 to 13, we find that the POT estimator still performs similarly to the previous examples. We skip the histograms due to paper length limitation, but they look similar to the previous ones.

Table 11: Statistics of each estimator in Example 3 with $N(0,0.17^2)$ distribution. CI: $1.10\times10^{-6}\pm6.49\times10^{-8}$.

method	mean	std	min	25%	50%	75%	max
MC	1.10E-06	1.04E-06	0	0	1.00E-06	2.00E-06	7.00E-06
POT-0.999	1.03E-06	5.92E-07	3.87E-11	5.75E-07	9.42E-07	1.41E-06	3.98E-06

Table 12: Statistics of each estimator in Example 3 with $Pareto(0.0008,2)$ distribution. CI: $1.07\times10^{-6}\pm6.42\times10^{-8}$.

method	mean	std	min	25%	50%	75%	max
MC	1.07E-06	1.03E-06	0	0	1.00E-06	2.00E-06	6.00E-06
POT-0.999	1.22E-06	4.46E-07	1.66E-07	9.06E-07	1.16E-06	1.49E-06	3.14E-06

3.4 Example 4: Queueing System

In this example, we consider a complicated queueing system described as follows. Suppose that customers arrive to the system following a Poisson process with rate λ. There are m sequential first-in-first-out queues in the system, and the service time of each queue follows a certain distribution. For $k=1,\dots,m-1$, after

Table 13: Statistics of each estimator in Example 3 with $Pareto(0.02,3)$ distribution. CI: $1.03\times10^{-6}\pm 6.28\times10^{-8}$.

method	mean	std	min	25%	50%	75%	max
MC	1.03E-06	1.02E-06	0	0	1.00E-06	2.00E-06	5.00E-06
POT-0.999	1.33E-06	5.19E-07	2.43E-07	9.59E-07	1.27E-06	1.63E-06	3.66E-06

the service at the k-th queue, each customer joins the $(k+1)$-th queue with probability p and directly leaves the system with probability $1-p$. After the service at the m-th queue, each customer leaves the system. Suppose that this simulation model outputs the sojourn times of the first d customers, i.e. $X^{(j)}$ is the sojourn time of the j-th customer. We aim to estimate the probability that the maximum sojourn time exceeds a, i.e. $f(X)=\max_{j=1,\dots,d}X^{(j)}$. In the experiments, we fix $\lambda=1, m=10, p=0.8, d=10$. We choose the simulation size $n=10^6$ and repeat each estimator $N=100$ times.

Tables 14 to 17 present the descriptive statistics where the service distribution is chosen as exponential or Pareto distributions. The value of a is not carefully tuned in this example, but we see that the POT estimate still performs generally better than crude MC.

Table 14: Statistics of each estimator in Example 4 with $Exp(1)$ distribution and $a=45$. CI: $3.50\times10^{-7}\pm 1.16\times10^{-7}$.

method	mean	std	min	25%	50%	75%	max
MC	3.50E-07	5.39E-07	0	0	0	1.00E-06	2.00E-06
POT-0.999	2.95E-07	1.93E-07	6.35E-09	1.23E-07	2.75E-07	4.41E-07	8.34E-07

Table 15: Statistics of each estimator in Example 4 with $Pareto(1,2)$ distribution and $a=1000$. CI: $5.30\times10^{-7}\pm 1.43\times10^{-7}$.

method	mean	std	min	25%	50%	75%	max
MC	5.30E-07	8.22E-07	0	0	0	1.00E-06	4.00E-06
POT-0.999	4.98E-07	2.34E-07	1.20E-07	3.25E-07	4.53E-07	6.04E-07	1.12E-06

Table 16: Statistics of each estimator in Example 4 with $Pareto(1,2)$ distribution and $a=2000$. CI: $1.40\times10^{-7}\pm 7.33\times10^{-8}$.

method	mean	std	min	25%	50%	75%	max
MC	1.40E-07	3.77E-07	0	0	0	0	2.00E-06
POT-0.999	1.31E-07	7.75E-08	2.08E-08	7.37E-08	1.13E-07	1.63E-07	3.46E-07

Table 17: Statistics of each estimator in Example 4 with $Pareto(1,3)$ distribution and $a=50$. CI: $4.40\times10^{-7}\pm 1.30\times10^{-7}$.

method	mean	std	min	25%	50%	75%	max
MC	4.40E-07	7.43E-07	0	0	0	1.00E-06	3.00E-06
POT-0.999	4.50E-07	2.28E-07	1.07E-07	2.89E-07	3.85E-07	5.68E-07	1.10E-06

4 CONCLUSION

While variance reduction techniques are powerful tools to increase the efficiency of crude MC in rare-event simulation, they often heavily rely on tractable problem structures and careful algorithmic design, which may not be possible for complex practical problems. This motivates us to study the use of POT, a prominent method in extreme event analysis, in rare-event simulation as an alternative to variance reduction. We formulate and test our POT approach on four rare-event simulation examples. Naturally, POT may not be as efficient as carefully designed algorithms such as IS or CMC when they are available. However, it outperforms crude MC in a reasonably wide spectrum of problems. It performs especially well when the simulation size is relatively large, the rare event boundary is not far from the threshold, and the tail of the distribution is heavy. Compared to crude MC, the POT estimator usually has smaller standard deviation with equal simulation size, and it gives an estimate of a roughly correct magnitude instead of a trivial estimate 0. Therefore, if efficient variance reduction techniques are not available, then our POT procedure can be used to refine and improve the crude MC estimate. For instance, if after generating some simulation samples, we find that none of them hits the target event but we cannot afford generating more samples, then we can compute the POT estimator with these existing samples to get a non-trivial estimate. In the future, we will investigate how to further improve the performance of our POT procedure, including threshold selection and fitting of GPD parameters.

ACKNOWLEDGMENTS

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

REFERENCES

Arief, M., Z. Huang, G. K. S. Kumar, Y. Bai, S. He, W. Ding, H. Lam, and D. Zhao. 2021. “Deep Probabilistic Accelerated Evaluation: A Robust Certifiable Rare-Event Simulation Methodology for Black-Box Safety-Critical Systems”. In *International Conference on Artificial Intelligence and Statistics*, edited by A. Banerjee and K. Fukumizu, 595–603. Proceedings of Machine Learning Research.

Asmussen, S. 1985. “Conjugate Processes and the Simulation of Ruin Problems”. *Stochastic Processes and Their Applications* 20(2):213–229.

Asmussen, S., and H. Albrecher. 2010. *Ruin Probabilities*, Volume 14. Singapore: World Scientific.

Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*, Volume 57. New York: Springer.

Au, S.-K., and J. L. Beck. 2001. “Estimation of Small Failure Probabilities in High Dimensions by Subset Simulation”. *Probabilistic Engineering Mechanics* 16(4):263–277.

Bai, Y., Z. Huang, H. Lam, and D. Zhao. 2022. “Rare-event simulation for neural network and random forest predictors”. *ACM Transactions on Modeling and Computer Simulation* 32(3):1–33.

Balkema, A. A., and L. De Haan. 1974. “Residual Life Time at Great Age”. *The Annals of Probability* 2(5):792–804.

Blanchet, J., P. Glynn, and H. Lam. 2009. “Rare Event Simulation for a Slotted Time $M/G/s$ Model”. *Queueing Systems* 63(1):33–57.

Blanchet, J., and H. Lam. 2014. “Rare-Event Simulation for Many-Server Queues”. *Mathematics of Operations Research* 39(4):1142–1178.

Bucklew, J. A. 2004. *Introduction to Rare Event Simulation*, Volume 5. New York: Springer.

Dupuis, P., A. D. Sezer, and H. Wang. 2007. “Dynamic Importance Sampling for Queueing Networks”. *The Annals of Applied Probability* 17(4):1306–1346.

Dupuis, P., and H. Wang. 2009. “Importance Sampling for Jackson Networks”. *Queueing Systems* 62(1):113–157.

Embrechts, P., T. Mikosch, and C. Klüppelberg. 1997. *Modelling Extremal Events: for Insurance and Finance*. New York: Springer-Verlag.

Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*, Volume 53. New York: Springer.

Glasserman, P., P. Heidelberger, P. Shahabuddin, and T. Zajic. 1999. “Multilevel Splitting for Estimating Rare Event Probabilities”. *Operations Research* 47(4):585–600.

Glasserman, P., W. Kang, and P. Shahabuddin. 2008. "Fast Simulation of Multifactor Portfolio Credit Risk". *Operations Research* 56(5):1200–1217.

Glasserman, P., and J. Li. 2005. "Importance Sampling for Portfolio Credit Risk". *Management Science* 51(11):1643–1656.

Huang, Z., H. Lam, D. J. LeBlanc, and D. Zhao. 2017. "Accelerated Evaluation of Automated Vehicles Using Piecewise Mixture Models". *IEEE Transactions on Intelligent Transportation Systems* 19(9):2845–2855.

Juneja, S., and P. Shahabuddin. 2006. "Rare-Event Simulation Techniques: An Introduction and Recent Advances". *Handbooks in Operations Research and Management Science* 13:291–350.

Kesidis, G., J. Walrand, and C.-S. Chang. 1993. "Effective Bandwidths for Multiclass Markov Fluids and Other ATM Sources". *IEEE/ACM Transactions On Networking* 1(4):424–428.

Kroese, D. P., and V. F. Nicola. 1999. "Efficient Estimation of Overflow Probabilities in Queues with Breakdowns". *Performance Evaluation* 36:471–484.

O'Kelly, M., A. Sinha, H. Namkoong, R. Tedrake, and J. C. Duchi. 2018. "Scalable End-to-End Autonomous Vehicle Testing via Rare-Event Simulation". *Advances in Neural Information Processing Systems* 31.

Pickands III, J. 1975. "Statistical Inference Using Extreme Order Statistics". *The Annals of Statistics*:119–131.

Ridder, A. 2009. "Importance Sampling Algorithms for First Passage Time Probabilities in the Infinite Server Queue". *European Journal of Operational Research* 199(1):176–186.

Rubino, G., and B. Tuffin. 2009. *Rare Event Simulation Using Monte Carlo Methods*. Hoboken, New Jersey: John Wiley & Sons.

Rubinstein, R. Y., and D. P. Kroese. 2016. *Simulation and the Monte Carlo Method*. Hoboken, New Jersey: John Wiley & Sons.

Sadowsky, J. S. 1991. "Large Deviations Theory and Efficient Simulation of Excessive Backlogs in a $GI/GI/m$ Queue". *IEEE Transactions on Automatic Control* 36(12):1383–1394.

Sadowsky, J. S., and J. A. Bucklew. 1990. "On Large Deviations Theory and Asymptotically Efficient Monte Carlo Estimation". *IEEE Transactions on Information Theory* 36(3):579–588.

Siegmund, D. 1976. "Importance Sampling in the Monte Carlo Study of Sequential Tests". *The Annals of Statistics*:673–684.

Smith, R. L. 1984. "Threshold Methods for Sample Extremes". In *Statistical Extremes and Applications*, edited by J. T. de Oliveira, 621–638. Dordrecht: Springer.

Szechtman, R., and P. W. Glynn. 2002. "Rare-Event Simulation for Infinite Server Queues". In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, Volume 1, 416–423. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Villén-Altamirano, M., and J. Villén-Altamirano. 1994. "RESTART: a Straightforward Method for Fast Simulation of Rare Events". In *Proceedings of the 1994 Winter Simulation Conference*, edited by J. D. Tew, M. S. Manivannan, D. A. Sadowski, and A. F. Seila, 282–289. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Webb, S., T. Rainforth, Y. W. Teh, and M. P. Kumar. 2018. "A Statistical Approach to Assessing Neural Network Robustness". *arXiv preprint arXiv:1811.07209*. https://arxiv.org/abs/1811.07209, accessed 22nd April, 2022.

Zhao, D., X. Huang, H. Peng, H. Lam, and D. J. LeBlanc. 2017. "Accelerated Evaluation of Automated Vehicles in Car-Following Maneuvers". *IEEE Transactions on Intelligent Transportation Systems* 19(3):733–744.

Zhao, D., H. Lam, H. Peng, S. Bao, D. J. LeBlanc, K. Nobukawa, and C. S. Pan. 2016. "Accelerated Evaluation of Automated Vehicles Safety in Lane-Change Scenarios Based on Importance Sampling Techniques". *IEEE Transactions on Intelligent Transportation Systems* 18(3):595–607.

AUTHOR BIOGRAPHIES

YUANLU BAI is a PhD candidate in the Department of IEOR at Columbia University. She received an M.S. in operations research from Columbia University, a B.S. in statistics and a B. Ec. in economics from Peking University. Her research interests are uncertainty quantification and rare-event simulation. Her email address is yb2436@columbia.edu.

SEBASTIAN ENGELKE is an Associate Professor at the Research Center for Statistics at the University of Geneva, where he is holding an Eccellenza grant. His research interests are in extreme value theory, spatial statistics, graphical models and data science. Since 2018, Sebastian is Associate Editor of the Springer journal Extremes and the Scandinavian Journal of Statistics. His email is sebastian.engelke@unige.ch.

HENRY LAM is an Associate Professor in the Department of Industrial Engineering and Operations Research at Columbia University. He received his Ph.D. degree in statistics from Harvard University in 2011, and was on the faculty of Boston University and the University of Michigan before joining Columbia in 2017. His research focuses on Monte Carlo simulation, uncertainty quantification, risk analysis, and stochastic and robust optimization. His email address is henry.lam@columbia.edu.