

Alternative rare events probability estimation

Serena Arima, Giovanni Petris and Luca Tardella

Abstract Estimation of rare event probability is an old problem which is still a key issue in recent simulation literature. This topic is of major importance in queuing systems as well as extreme value analysis in many fields such as reliability, telecommunication and insurance risks. When closed form expression of the probability of a rare event is not available Monte Carlo approximation is one of the favorite solutions. However crude Monte Carlo yields very inefficient estimators and in order to overcome this problem several alternative solutions have been proposed based on the importance sampling technique. In this paper, after a brief review of some recent literature, we propose an alternative method for estimating rare events using simulation: we investigate simple deterministic transforms of the random sample drawn from the original distribution such that the rare events become not-so-rare. The new approach can be easily implemented and the resulting estimator can achieve a bounded relative error in some relevant cases such as exponential-like and regularly varying density functions. The approach is presented here in a very simple basic setting of a single random variable but it can be also extended to sum of random variables.

Key words: bounded relative error, heavy tails, importance sampling, Monte Carlo methods, rare event simulation, statistical computing

Serena Arima
Sapienza University of Rome, e-mail: serena.arima@uniroma1.it

Giovanni Petris
University of Arkansas e-mail: GPetris@uark.edu

Luca Tardella
Sapienza University of Rome e-mail: luca.tardella@uniroma1.it

1 Introduction

Estimation of rare event probability is one of the key issue of the recent simulation literature from both an applicative and methodological point of view. It is of practical interest in many areas: civil aircraft catastrophic failures, ruin probability for insurance company, overflow of memory buffers in telecommunication systems just to mention few typical examples. From a methodological point of view, the estimation of rare event probability is a challenging problem which has witnessed a remarkable increase of attention and developments in the last two decades ([6], [4]).

Although it is not possible to define universally (i.e. regardless on the problem at hand) how small the probability of an event must be in order to be considered rare, it is often argued that in many real contexts a convenient threshold can be set in the order of 10^{-9} . Indeed, the methodological core of the rare event probability estimation lies on the complexity of the random system which makes in the continuous case the corresponding integration problem analytically unfeasible or in the discrete case the involved enumeration problem beyond computing capabilities. To overcome such difficulties one of the most favorite strategies adopted so far consists in approximating the rare event probability through the simulation of the complex system. Monte Carlo approximation and importance sampling techniques via an exponential change of measure can be considered the most well-established current approaches for rare event estimation [7]. In this paper we propose alternative solutions which can give some positive solutions in some cases of heavy tailed systems for which the mainstream importance sampling strategy has not yet given a complete solution. In Section 2 we give a brief account of the main ideas of the mainstream importance sampling strategy for rare event estimation. In Section 3 we illustrate a different new naive idea which gives a fully satisfactory solution in some particular examples with one sample simulation. In Section 4 we sketch how the naive idea can be generalized further and we give a theoretical ground for a fully general approach. In Section 5 we give concluding remarks.

2 Rare event simulation strategies

Consider a complex systems formalized as a random variable X with distribution represented by a density function $f(X)$; we suppose that the event $X > c$ is considered rare when its probability α is smaller than a fixed threshold $\bar{\alpha}$. To fix ideas let us assume $\bar{\alpha} = 10^{-9}$ and write α , the quantity of interest, as follows

$$\alpha = \alpha(c) = Pr(X > c) = \int_c^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} I_{(-\infty, c)}(x) f(x) dx \quad (1)$$

Indeed we assume that $f(x)$ is such that the integral in (1) cannot be analytically solved and possibly not even $f(x)$ can be easily explicitly computed exactly as in the case when X corresponds to a sum of i.i.d. random variables. To cope with this

problem, Monte Carlo approximation techniques have been proposed [6]. Given a sample $X_1, X_2, \dots, X_n$ from f , a crude Monte Carlo estimation $\hat{\alpha}_{MC}$ of the probability α is obtained as

$$\hat{\alpha}_{MC} = \frac{1}{n} \sum_{i=1}^n I_{(c, +\infty)}(X_i) \quad (2)$$

where $I_{(c, +\infty)}(X_i)$ is the indicator function. The estimator $\hat{\alpha}_{MC}$ is an unbiased estimator of α and its variance is $Var(\hat{\alpha}_{MC}) = \frac{\alpha(1-\alpha)}{n}$. The relative error of an estimator $\hat{\alpha}$ is defined as

$$\delta_{rel}(\hat{\alpha}) = \frac{\sqrt{Var(\hat{\alpha})}}{E[\hat{\alpha}]} \quad (3)$$

A desirable property for rare event estimators is to possess a bounded relative error (BRE), that is

$$\lim_{c \rightarrow \infty} \delta_{rel}(\hat{\alpha}(c)) < k < \infty.$$

The relative error of the Monte Carlo estimator is $\delta_{rel}(\hat{\alpha}_{MC}) = \sqrt{\frac{1-\alpha}{n\alpha}}$. This establishes the starting point of the recent efforts in the literature of rare event estimation with MC method: when the event become rarer and rarer (i.e. the most interesting and problematic cases), that is when $\alpha \rightarrow 0$, the relative error $\delta_{rel}(\hat{\alpha}_{MC})$ tends to infinity, so that the relative error of the Monte Carlo approximation goes out of control unless n is huge enough. For instance, when $\alpha = 10^{-9}$ in order to obtain $\delta_{rel}(\hat{\alpha}_{MC}) < 0.10$, we need $n > 3.84 \times 10^{11}$, which is a prohibitive sample size even for simple systems. This simple example shows that $\hat{\alpha}_{MC}$ cannot guarantee BRE and motivates the study of alternative simulation-based approaches.

Importance sampling (IS) is certainly one of the most popular approaches in rare event analysis [2]. Suppose one wishes to estimate the expected value

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

where g is an arbitrary integrable function.

Importance Sampling is mainly based on a change of measure relying on the simulation from an alternative distribution $\tilde{f}$ (importance function). It can be implemented so that the variance of the resulting estimator is smaller than its Monte Carlo counterpart. In fact, the original target expected value can be rewritten as

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x)f(x) \frac{\tilde{f}(x)}{\tilde{f}(x)} dx = E_{\tilde{f}} \left[\frac{g(x)f(x)}{\tilde{f}(x)} \right] \quad (4)$$

When we are interested in estimating rare event probabilities, (4) corresponds to (1) with $g(X)$ replaced by $I_{(c, \infty)}(X)$ and the probability of interest is

$$\alpha = P(X > c) = \int I_{(c, \infty)}(x)f(x)dx = \int_c^{+\infty} f(x)dx.$$

An importance sampling approach, widely used in rare event estimation, is based on taking as importance function a parameterized distribution within an exponential family with density

$$\tilde{f}_\theta(x) = \frac{1}{m(\theta)} e^{\theta T(x)} f(x)$$

where $m(\theta) = \int e^{\theta T(x)} f(x) dx$. When $T(x) = x$, this is the well-known *exponential tilting* or *twisting* [7].

An importance sampling estimator for α is then obtained as

$$\hat{\alpha}_{ET} = \frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{\tilde{f}_\theta(X_i)} I(X_i > c) \quad (5)$$

where $X_1, X_2, \dots, X_n \sim \tilde{f}_\theta$. The parameter θ can be calibrated in order to minimize the variance of the estimator. It is shown [5] that under some specific conditions on the tail behavior, the *exponential twisting* trick can achieve bounded relative error for estimating the probability of tail events. However, the sufficient conditions on the tail behavior do not cover the case of heavy-tailed distributions [3].

3 An alternative approach

As already described in the previous section, the basic intuition underlying the exponential twisting is to use an importance density $\tilde{f}$ which is based on the original f but is biased towards the rare event (typically in the tails of the original distribution) so that the tails gets inflated and the rare event becomes not-so-rare. Here we suggest and explore a different idea. Instead of building up a shifted importance distribution moving some mass towards the rare event we modify of the original sample by censoring the rare event and moving the censored mass elsewhere. This results in a slight perturbation of the corresponding original censored-density obtained by appropriately moving rare event density into the body of the sample space. More formally, we fix $\lambda < c$ and instead of the original i.i.d. sample $X_1, \dots, X_n$ from f consider the λ -reflected sample $\tilde{X}_1, \dots, \tilde{X}_n$ where

$$\tilde{X}_i = X_i I_{(c, \infty)}(X_i) + (\lambda - X_i + c)(1 - I_{(c, \infty)}(X_i)) = \begin{cases} X_i & \text{if } X_i \leq c \\ \lambda - (X_i - c) & \text{if } X_i > c \end{cases} \quad (6)$$

$\tilde{X}_i \sim \tilde{f}_\lambda$ and its corresponding density is

$$\tilde{f}_\lambda(x) = f(x) I_{(-\infty, c)}(x) + f(\lambda - x + c) I_{(-\infty, \lambda)}(x) \quad (7)$$

depending on a real parameter $\lambda \in (-\infty, c)$. Indeed, (6) and (7) tells us that, differently the reflected sample $\tilde{X}_i$ is easy to simulate directly as a function of the original simulation of the X_i 's and also the corresponding density is readily available. Graphically we show an example of the original f as well the $\tilde{f}_\lambda$ and g in Figure 1. Hence

we want to see the performance of the IS estimator based on this idea and possibly optimized with a convenient choice of λ

$$\hat{\alpha}_\lambda = 1 - \frac{1}{n} \sum_{i=1}^n \frac{f(\tilde{X}_i)}{\tilde{f}_\lambda(\tilde{X}_i)} = \frac{1}{n} \sum_{i=1}^n \frac{f(\lambda - \tilde{X}_i + c) I_{(-\infty, \lambda)}(\tilde{X}_i)}{\tilde{f}_\lambda(\tilde{X}_i)}$$

and we show that it can be usefully exploited for estimating rare events probabilities. The estimator $\hat{\alpha}_\lambda$ is indeed unbiased and consistent due to a simple importance sampling argument, namely

$$1 - \alpha = \int_{-\infty}^c f(x) dx = \int_{-\infty}^c \frac{f(x)}{\tilde{f}_\lambda(x)} \tilde{f}_\lambda(x) dx = E_{\tilde{f}_\lambda} \left[\frac{f(\tilde{X})}{\tilde{f}_\lambda(\tilde{X})} \right].$$

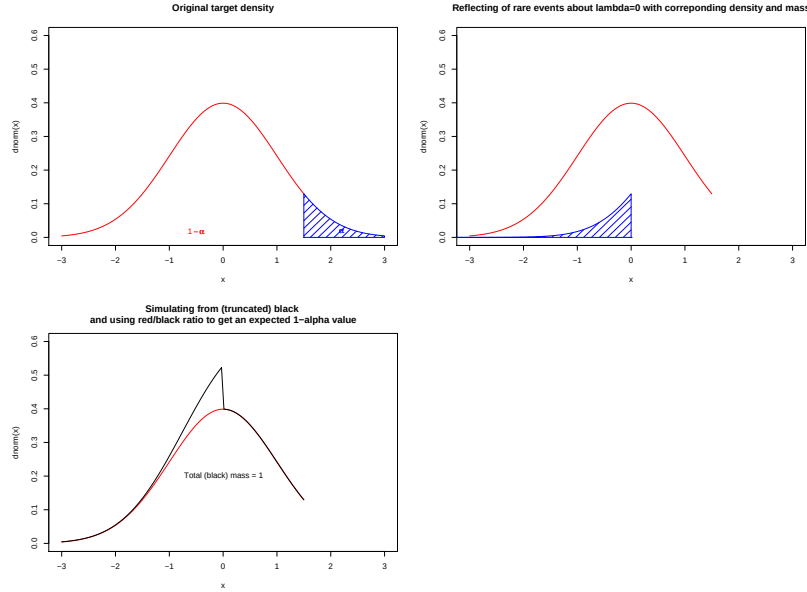


Fig. 1 Scheme of the proposed alternative method for estimating the rare event probability.

A sufficient condition in order to guarantee that $\hat{\alpha}_\lambda$ enjoys the BRE property is to show that its second moment given by

$$E[\hat{\alpha}_\lambda^2] = \text{Var}[\hat{\alpha}_\lambda] + \alpha^2 = \int_{-\infty}^{\lambda} \left(\frac{f(\lambda - x + c)}{f(x) + f(\lambda - x + c)} \right)^2 \tilde{f}_\lambda(x) dx \quad (8)$$

is $O(\alpha^2)$.

We can show that for some probability distributions $f(x)$ our new strategy yields a simple estimator with BRE.

For example, for the Laplace distribution with arbitrary parameter β we have that:

$$E[\hat{\alpha}_\lambda^2] = \frac{e^{-\frac{c}{\beta}}}{2+2e^{\frac{c}{\beta}}} \quad E[\hat{\alpha}_\lambda] = \frac{e^{-\frac{c}{\beta}}}{2}$$

$$\lim_{c \rightarrow +\infty} \sqrt{\frac{E[\hat{\alpha}_\lambda^2]}{(E[\hat{\alpha}_\lambda])^2} - 1} = 1$$

The estimator $\hat{\alpha}_\lambda$ turns out to have BRE also for the logistic distribution. However we cannot get a closed form expression for $\delta_{rel}(\hat{\alpha}_\lambda)$ but it can be computed numerically and it can be shown to converge to 2.33 as $c \rightarrow \infty$ in the standard logistic case. However, it is possible to verify that the simple intuitive reflection idea, as it stands, does not yield estimators with BRE in general. For instance one can see that $\delta_{rel}(\hat{\alpha}_\lambda)$ is unbounded in the Cauchy case as well as in the double Pareto. We then explain in the next section how to generalize the original simple idea underlying $\hat{\alpha}_\lambda$ so that it can help building new classes of IS estimators with BRE for other relevant classes of distributions.

4 Generalizing the alternative approach

In this section we provide two main ideas to generalize the previous simple alternative IS estimation of the rare event probability α . A first generalization of the trick consists in considering the construction of the importance density $\tilde{f}_\lambda(x)$ corresponding to a *reflected and rescaled* X_i 's exceeding the threshold with a suitable change of scale say $s(c) > 0$ possibly depending on c . More precisely, the transformed X_i 's become $\tilde{X}_1, \dots, \tilde{X}_n$ where

$$\tilde{X}_i = X_i I_{(c, \infty)}(X_i) + \left(\lambda + \frac{c - X_i}{s(c)} \right) (1 - I_{(c, \infty)}(X_i)) = \begin{cases} X_i & \text{if } X_i \leq c \\ \lambda - \frac{X_i - c}{s(c)} & \text{if } X_i > c \end{cases}$$

$\tilde{X}_i \sim \tilde{f}_\lambda$ and their corresponding density is

$$\tilde{f}_\lambda(x) = f(x) I_{(-\infty, c)}(x) + s(c) f(s(c)(\lambda - x) + c) I_{(-\infty, \lambda)}(x)$$

depending on a real parameter $\lambda \in (-\infty, c)$ and a suitable change of scale $s(c) > 0$. To appreciate the potential usefulness of the generalized approach we can show that for $s(c) = c$ the generalized estimator

$$\begin{aligned} \hat{\alpha}_{\lambda, s(c)} &= 1 - \frac{1}{n} \sum_{i=1}^n \frac{f(\tilde{X}_i)}{\tilde{f}_\lambda(\tilde{X}_i)} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{s(c) f(s(c)(\lambda - X_i) + c) I_{(-\infty, \lambda)}(X_i)}{\tilde{f}_\lambda(\tilde{X}_i)} \end{aligned}$$

yields an estimator with BRE when f follows a Cauchy distribution with arbitrary location and scale parameter. For a standard Cauchy distribution, we have that:

$$E[\hat{\alpha}_\lambda^2] = \frac{1}{2\pi} \left(\pi - \sqrt{\frac{c}{1+2c(1+c)}} \pi - 2\text{ArcTan}(c) + 2\sqrt{\frac{c}{1+2c(1+c)}} \text{ArcTan}\left(\frac{c^{\frac{3}{2}}}{\sqrt{1+2c(1+c)}}\right) \right)$$

$$E[\hat{\alpha}] = \frac{\text{ArcTan}(\frac{1}{c})}{\pi} \quad \lim_{c \rightarrow +\infty} \sqrt{\frac{E[\hat{\alpha}_\lambda^2]}{(E[\hat{\alpha}_\lambda])^2}} - 1 = 1.04613$$

The estimator $\hat{\alpha}_\lambda$ turns out to have BRE also for the Pareto distribution for which the relative error can be computed numerically: for a standard Pareto distribution the relative error is 2.33.

In order to yield a general perspective one can consider the following strategy. The idea is to reflect the X_i simulated over the threshold c in such a way that the reflected point $\tilde{X}_i$ as a function of the original one $\tilde{X}_i = H(X_i)$ has a distribution similar to the original one so that the corresponding importance density $\tilde{f}_\lambda(x)$ and the difference $\tilde{f}_\lambda(x) - f(x)$ can yield a convenient reduced variance.

We will argue that the following function $H(\cdot)$

$$H(x) = F^{-1} \left(F(\lambda) \frac{F(x) - (1 - \alpha)}{\alpha} \right)$$

would be a convenient choice. In fact, if $X_i \sim F$ it is well known that $F(X_i) \sim \text{Unif}[0, 1]$, so that, conditionally on $X_i > c$, $F(X_i) \sim \text{Unif}[1 - \alpha, 1]$, hence

$$\frac{F(X_i) - (1 - \alpha)}{\alpha} \sim \text{Unif}[0, 1]$$

and finally

$$F^{-1} \left(F(\lambda) \frac{F(X_i) - (1 - \alpha)}{\alpha} \right) \sim Y$$

where

$$P(Y < y) = P(X_i < y | X_i < \lambda).$$

Now let us consider the density function of the following H_λ reflected sample.

$$\begin{aligned} \tilde{X}_i &= X_i I_{(c, \infty)}(X_i) + F^{-1} \left(F(\lambda) \frac{F(X_i) - (1 - \alpha)}{\alpha} \right) (1 - I_{(c, \infty)}(X_i)) = \\ &= \begin{cases} X_i & \text{if } X_i \leq c \\ F^{-1} \left(F(\lambda) \frac{F(X_i) - (1 - \alpha)}{\alpha} \right) & \text{if } X_i > c \end{cases} \end{aligned}$$

Now one can easily argue that the above idea is pointless since it apparently requires the exact knowledge of α . On the other hand one can easily show that in some relevant cases (exponential-like and regularly varying functions) it yields expressions

which seemingly do not require the exact knowledge of α and turn out to suggest an importance sampling estimator which enjoys the BRE property.

5 Discussion

In this paper, we propose an alternative simulation-based method for estimating the probability of rare events. Our main idea is to reverse the intuition of the standard exponential tilting approach: instead of using an instrumental importance density which shifts probability mass towards the rare event we propose to move the rare event towards the body of the original distribution. Similarly to the exponential tilting this idea rely a new instrumental importance density $\tilde{f}$ which is a sort of perturbation of the original density but differently from the exponential tilting the simulation $\tilde{X}_i \sim \tilde{f}$ can be explicitly and easily derived from the original sampling of X_i . A referee noted that, surprisingly, two events in the neighborhood of the threshold such that $x_1 < c < x_2$, originally having a very similar density f end up having a very different density under the new importance $\tilde{f}$. However although it sounds counterintuitive we have shown that this transformation can be appropriate since in the examples considered in Section 3 and 4 it yields BRE estimators in some relevant cases such as exponential-like and regularly varying functions. The method is simple and easy to apply with a very low computational cost. We have shown in [1] that the approach can be effectively extended to systems consisting of sum of random variables. The proposed method deserves to be investigated further to find out sufficient conditions on classes of densities which allow to extend the BRE property derived for a specific class.

References

1. Arima, S., Petris, G. and Tardella, L.: An alternative rare events probability estimation, Working Paper (2010)
2. Asmussen, S. and Kroese, D. and Rubinstein, R.Y.: Heavy tails, importance sampling and cross-entropy. *Stochastic Models* . **21** (1), 1532–6349 (2005)
3. Asmussen, S. and Kroese, D.: Improved algorithms for rare event simulation with heavy tails. *Advances in Applied Probability*. **38** (2), 545–558 (2006)
4. Asmussen, S.: Importance sampling for failure probabilities in computing and data transmission. *Journal of Applied Probability*. **46**, 768–790 (2009)
5. Juneja, S.: Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error . *Queueing Systems. Theory and Applications*. **57** (2), 115–128 (2007)
6. Rubino, G. and Tuffin, B.: *Rare Event Simulation using Monte Carlo Methods*. J. Wiley (eds.) (2009)
7. Siegmund, D. : Importance Sampling in the Monte Carlo Study of Sequential Tests. *Annals of Statistics*, **4** (4), 673–684 (1976)