

# Extreme Value Importance Sampling for Rare Event Risk Measurement

D.L. McLeish and Zhongxian Men

**Abstract** We suggest practical and simple methods for Monte Carlo estimation of the (small) probabilities of large losses using importance sampling. We argue that a simple optimal choice of importance sampling distribution is a member of the generalized extreme value distribution and, unlike the common alternatives such as Esscher transform, this family achieves bounded relative error in the tail. Examples of simulating rare event probabilities and conditional tail expectations are given and very large efficiency gains are achieved.

**Keywords** Rare event simulation · Risk measurement · Relative error · Monte Carlo methods · Importance sampling

## 1 Introduction

Suppose  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$  is a vector of independent random variables each with cumulative distribution function (cdf)  $F$  and probability density function (pdf)  $f$  with respect to Lebesgue measure. Suppose we wish to estimate the probability of a large loss,  $p_t = P(L(\mathbf{Y}) > t)$  where  $L(\mathbf{Y})$  is the loss determined by the realization  $\mathbf{Y}$  (usually assumed to be monotonic in its components) and  $t$  is some predetermined threshold. There are many different loss functions  $L(\mathbf{Y})$  used in rare event simulation, including barrier hitting probabilities of sums or averages of independent random variables, or of processes such as an Ornstein–Uhlenbeck or Feller process. The methods discussed here are designed for problems in which a small number of continuous factors are the primary contributors to large losses. We wish to use *importance sampling* (IS) (see [3], Sect.4.6 or [12] p. 183): generate independent replications of  $\mathbf{Y}$  repeatedly, say  $n$  times, from an alternative distribution, say one with pdf  $f_{\text{IS}}(\mathbf{y})$  and then estimate the above expected value using the IS estimator

---

D.L. McLeish (✉)  
University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada  
e-mail: dlmcleis@uwaterloo.ca

Z. Men  
e-mail: zmen@uwaterloo.ca

© The Author(s) 2015  
K. Glau et al. (eds.), *Innovations in Quantitative Risk Management*,  
Springer Proceedings in Mathematics & Statistics 99,  
DOI 10.1007/978-3-319-09114-3\_18

$E_{\theta}([I(L(\mathbf{Y}) > t)f(\mathbf{Y})/f_{IS}(\mathbf{Y})])$ , where  $I(*)$  denotes the indicator function. We use

$$\hat{p}_t = \frac{1}{n} \sum_{i=1}^n I(L(\mathbf{Y}_i) > t) \frac{f(\mathbf{Y}_i)}{f_{IS}(\mathbf{Y}_i)}, \quad \text{where } \mathbf{Y}_j \sim f_{IS}(\mathbf{y}). \tag{1}$$

If we denote by  $E_{IS}$  the expected value under the IS distribution  $f_{IS}$ , and by  $E$  the expected value under the original distribution  $f$ , then

$$\begin{aligned} E_{IS}(\hat{p}_t) &= E_{IS}\left(I(L(\mathbf{Y}) > t) \frac{f(\mathbf{Y})}{f_{IS}(\mathbf{Y})}\right) \\ &= p_t \end{aligned}$$

confirming that this is an *unbiased* estimator. There is a great deal of the literature on such problems when the event of interest is “rare”, i.e., when  $p_t$  is very small, and many different approaches depending on the underlying loss function and distribution. We do not attempt a review of the literature in the limited space available. Excellent reviews of the methods and applications are given in Chap.6 of [1] and Chap. 10 of [9]. Highly efficient methods have been developed for tail estimation in very simple problems, such as when the loss function consists of a sum of independent identically distributed increments. In this paper, we will provide practical tools for simulation of such problems in many examples of common interest. For rare events, the variance or standard error is less suitable as a performance measure than a version scaled by the mean because in estimating very small probabilities such as 0.0001, it is not the absolute size of the error that matters but its size relative to the true value.

**Definition 1** The relative error (RE) of the importance sample estimator is the ratio of the estimator’s standard deviation to its mean.

Simulation is made more difficult for rare events because crude Monte Carlo fails. As a simple illustration, suppose we wish to estimate a very small probability  $p_t$ . To this end, we generate  $n$  values of  $L(\mathbf{Y}_i)$  and estimate this probability with  $\hat{p} = X/n$  where  $X$  is the number of times that  $L(\mathbf{Y}_i) > t$  and  $X$  has a Binomial( $n, p_t$ ) distribution. In this case, the relative error is

$$RE = \frac{\sqrt{\text{Var}(\frac{X}{n})}}{E(\frac{X}{n})} = \frac{\sqrt{\frac{1}{n} p_t (1 - p_t)}}{p_t} = n^{-1/2} \sqrt{\frac{1 - p_t}{p_t}}.$$

For rare events,  $p_t$  is small and the relative error is very large. If we wish a normal-based confidence interval for  $p_t$  of the form  $(\hat{p}_t - 0.1\hat{p}_t, \hat{p}_t + 0.1\hat{p}_t)$  for example, we are essentially stipulating a certain relative error (RE = 0.05102) whatever the value of  $p_t$ . In order to achieve a reasonable bound on the relative error, we would need to use sample sizes that were of the order of  $p_t^{-1}$ , i.e., larger and larger sample sizes for rarer events.

Returning to the estimator (1), if we take its variance, we obtain

$$n^{-1} \left[ E_{IS} \left( I(L(\mathbf{Y}) > t) \frac{f^2(\mathbf{Y})}{f_{IS}^2(\mathbf{Y})} \right) - p_t^2 \right] = n^{-1} \left[ E \left( I(L(\mathbf{Y}) > t) \frac{f(\mathbf{Y})}{f_{IS}(\mathbf{Y})} \right) - p_t^2 \right]$$

the relative error is

$$RE(f_{IS}; t, n) = n^{-1/2} \left[ p_t^{-2} E \left( I(L(\mathbf{Y}) > t) \frac{f(\mathbf{Y})}{f_{IS}(\mathbf{Y})} \right) - 1 \right]^{1/2}.$$

The relative error decreases linearly in  $n^{-1/2}$  but it is the factor

$$p_t^{-2} \left[ E \left( I(L(\mathbf{Y}) > t) \frac{f(\mathbf{Y})}{f_{IS}(\mathbf{Y})} \right) \right]^{1/2},$$

highly sensitive to  $t$  when  $p_t$  is small, that determines whether an IS distribution is good or bad for a given problem. There is a large literature regarding the use of importance sampling for such problems, much recommending the use of the exponential tilt or *Esscher transform*. The suggestion is to adopt an IS distribution of the form

$$f_{IS}(y) = \text{constant} \times e^{\theta y} f(y) \tag{2}$$

and then tune the parameter  $\theta$  so that the IS estimator is as efficient as possible (see, for example [1, 7, 15]). Chapter 10 of [9] provides a detailed discussion of methods and applications as well as a discussion of the boundedness of relative error. McLeish [11] demonstrates that the IS distribution (2) is suboptimal and unlike the alternatives we explore there, does not typically achieve bounded relative error. We argue for the use of the generalized extreme value (GEV) family of distributions for such problems. A loose paraphrase of the theme of the current paper is “all you really need<sup>1</sup> is GEV”. Indeed in Appendix A, we prove a result (Proposition 1) which shows that, under some conditions, there is always an importance sampling estimator whose relative error is bounded in the tail obtained by generating the distance along one principal axis from an extreme value distribution, while leaving the other coordinates unchanged in distribution. We now consider some one-dimensional problems.

## 2 The One-Dimensional Case

Consider estimating  $P(L(Y) > t)$  where the value of  $t$  is large, the random variable  $Y$  is one-dimensional and  $L(y)$  is monotonically increasing. We would like to use an importance sample distribution for which, by adjusting the values of the parameters,

---

<sup>1</sup> For importance sampling estimates of rare events, at least, with apologies to the Beatles.

we can have small relative error for any large  $t$ . We seek a parametric family  $\{f_\theta; \theta \in \Theta\}$  of importance sample estimators which have bounded relative error as follows:

**Definition 2** Suppose  $\mathcal{H}$  is the class of non-negative integrable functions  $\{I(L(Y) > t); \text{ for } t > T\}$ . We say a parametric family  $\{f_\theta; \theta \in \Theta\}$  has **bounded relative error** for the class  $\mathcal{H}$  if

$$\sup_{t > T} \inf_{\theta \in \Theta} \text{RE}(f_\theta; t, n) < \infty.$$

A parametric family has bounded relative error for estimating functions in a class  $\mathcal{H}$  if, for each  $t > T$ , there exists a parameter value  $\theta$  which provides bounded relative error. Indeed, a bound on the relative error of approximately  $0.738n^{-1/2}$  can be achieved by importance sampling if we know the tail behavior of the distribution. There are very few circumstances under which the exponential tilt, families of continuous densities of the form

$$f_\theta(x) = \text{constant} \times e^{\theta y} f(y),$$

provides bounded relative error. The literature recommending the exponential tilt usually rests on demonstrating *logarithmic efficiency* (see [1], p. 159 or Sect. 10.1 of [9]), a substantially weaker condition that does not guarantee a bound in the relative error. Although we may design a simulation optimally for a specific criterion such as achieving small relative error in the estimation of  $P(L(Y) > t)$ , we are more often interested in the nature of the whole tail beyond  $t$ . For example, we may be interested in  $E[(L(Y) - t) I(L(Y) > t)] = \int_t^\infty P(L(Y) > s) ds$  and this would require that a single simulation be efficient for estimating all parameters  $P(L(Y) > s)$ ,  $s > t$ . The property of bounded relative error provides some assurance that the family used adapts to the whole tail, rather than a single quantile.

For simplicity, we assume for the present that  $Y$  is univariate, has a continuous distribution, and  $L(Y)$  is a strictly increasing function of  $Y$ . Then

$$P(L(Y) > t) = \int_{L^{-1}(t)}^\infty f(y) dy$$

and we can achieve bounded relative error if we use an importance sample distribution drawn from the family

$$f_\theta(y) = \text{constant} \times e^{\theta T(y)} f(y) \tag{3}$$

where  $T(y)$  behaves, for large values of  $y$ , roughly like a linear function of  $\bar{F}(y) = 1 - F(y)$ . If  $T(y) \sim -\bar{F}(y)$  as  $y \rightarrow \infty$ , the optimal parameter  $\theta$  is  $\theta_t = \frac{k_2}{p_t} \simeq \frac{1.59362}{p_t}$  and the limit of the relative error of the IS estimator is  $\simeq 0.738n^{-1/2}$  (see Appendix

---

<sup>2</sup> Here  $k_2 \simeq 1.5936$  is the unique positive solution to the equation  $e^{-k} + \frac{k}{2} = 1$ .

A). The simplest and most tractable family of distributions with appropriate tail behavior is the GEV distribution associated with the density  $f(y)$ .

We now provide an intuitive argument in favor of the use of the GEV family of IS distributions. For a more rigorous justification, see Appendix A.

The choice  $T(y) = -\bar{F}(y)$  provides asymptotic bounded relative error [11]. Consider a family of cumulative distribution functions

$$F_\theta(y) = \frac{e^{-\theta\bar{F}(y)} - e^{-\theta}}{1 - e^{-\theta}}.$$

The corresponding probability density function  $F'_\theta(y)$  is of the form (3). As  $\theta \rightarrow \infty$  and  $y \rightarrow \infty$  in such a way that  $\theta\bar{F}(y)$  converges to a nonzero constant, then

$$(F(y))^\theta = (1 - \bar{F}(y))^\theta \sim e^{-\theta\bar{F}(y)}, \tag{4}$$

so that

$$F_\theta(y) \sim (F(y))^\theta. \tag{5}$$

Therefore,  $F_\theta(y)$  is asymptotically equivalent to the distribution of the maximum of  $\theta$  observations from the original target distribution  $F$ . This, suitably normalized, converges to a member of the GEV family of distributions. We also show in [11] that the optimal parameter is asymptotic to  $\theta = k_2/p_t$  as  $p_t \rightarrow 0$ . Consequently,

$$F_\theta(t) \sim (1 - \bar{F}(t))^\theta \sim (1 - p_t)^{k_2/p_t} \sim e^{-k_2} \simeq 0.203.$$

Thus, when we use the corresponding extreme value importance sample distribution, about 20.3 % of the observations will fall below  $t$  and the other 79.7 % will fall above, and this can be used to identify one of the parameters of the IS distribution. Of course, only the observations greater than  $t$  are directly relevant to estimating quantities like  $P(L > t)$ . This leads to the option of conditioning on the event  $L > t$  and using the generalized Pareto family (see Appendix B).

The three distinct classes of extreme value distributions and some of their basic properties are outlined in Appendix B. All have simple closed forms for their pdf, cdf, and inverse cdf and can be easily and efficiently generated. In addition to a shape parameter  $\xi$ , they have location and scale parameters  $d, c$  so the cdf is  $H_\xi(\frac{x-d}{c})$ , where  $H_\xi(x)$  is the cdf for unit scale and 0 location parameter. We say that a given continuous cdf  $F$  falls in the maximum domain of attraction of an extreme value cdf  $H_\xi(x)$  if there exist sequences  $c_n, d_n$  such that

$$F^n(d_n + c_n x) \rightarrow H_\xi(x) \text{ as } n \rightarrow \infty.$$

We will choose an extreme value distribution with parameters that approximate the distribution of the maximum of  $\theta_t = k_2/p_t$  random variables from the original density  $f(y)$ . Further properties of the extreme value distributions and detail on the

choice of parameters is given in Appendix B. Proposition 1 in Appendix A shows that if  $F$  is in the domain of attraction of  $H_0$ , then  $H_0$  provides a family of IS distributions with bounded relative error. It is also unique in the sense that any other such IS distribution has tails essentially equivalent to those of  $H_0$ . A similar result can be proved for  $\xi \neq 0$ . The superiority of the extreme value distributions for importance sampling stems from the bound on relative error, but equally from their ease of simulation, the simple closed form for the pdf and cdf and the maximum stability property, which says that the distribution of the maximum of i.i.d. random variables drawn from this distribution is a member of the same family.

We get a better sense of the extent of the variance reduction using IS if we compare sample sizes required to achieve a certain relative error. If we use crude random sampling in order to estimate  $p_t$  using a sample size  $n_{cr}$ , the relative error of the crude estimator is

$$\text{RE}(\text{crude}) = \frac{\sqrt{\frac{p_t(1-p_t)}{n_{cr}}}}{p_t} = \sqrt{\frac{1-p_t}{n_{cr}p_t}}$$

whereas if we use a GEV importance sample of size  $n_{IS}$ , the relative error is  $\text{RE}(IS) \simeq 0.738n_{IS}^{-1/2}$ . Equating these, the ratio of the sample sizes required for a fixed relative error is  $\frac{n_{cr}}{n_{IS}} \simeq \frac{1.84}{p_t}$  for  $p_t$  small. Indeed, if  $p_t = 10^{-4}$ , an importance sample estimator based on a sample size  $5 \times 10^6$  is quite feasible on a small laptop computer, but is roughly equivalent to a crude Monte Carlo estimator of sample size  $9.2 \times 10^{10}$ , possible only on the largest computers.

### 3 Examples

#### 3.1 Example 1: Simulation Estimators of Quantiles and TailVar for the Normal Distribution

Rarely when we wish to simulate an expected value in the region of the space  $[L(Y) > t]$  is this the only quantity of interest. More commonly, we are interested in various functions sensitive to the tail of the distribution. This argues for using an IS estimator with bounded relative error rather than the more common practice of simply conditioning on the region of interest. For a simple example, suppose  $Y$  follows a  $N(0, 1)$  distribution and we wish to estimate a property of the tail defined by  $Y > t$ , where  $t$  is large. Suppose we simulate from the conditional distribution given  $Y > t$ , that is from the pdf

$$\frac{1}{1 - \Phi(t)} \phi(y) I(y > t) \tag{6}$$

where  $\phi$  and  $\Phi$  are the standard normal pdf and cdf respectively. If we wish also to estimate  $P(Y > t + s | Y > t) \sim e^{-st - \frac{s^2}{2}}$  for  $s > 0$  fixed, sampling from this pdf is highly inefficient, since for  $n$  simulations from pdf (6), the RE for estimating  $P(Y > t + s | Y > t)$  is approximately  $n^{-1/2} \sqrt{e^{st + \frac{s^2}{2}} - 1}$  and this grows extremely rapidly in both  $t$  and  $s$ . We would need a sample size of around  $n = 10^4 e^{st + \frac{s^2}{2}}$  (or about 60 trillion if  $s = 3$  and  $t = 6$ ) from the IS density (6) to achieve a RE of 1%.

Crude Monte Carlo fails here but use of IS with the usual standard exponential tilt or Esscher transform with  $T(y) = y$ , though very much better, still fails to deliver bounded relative error. In [11], it is shown that the relative error is  $\sim (\frac{\pi}{2})^{1/4} \sqrt{t/n} \rightarrow \infty$  as  $p_t \rightarrow 0$ . While the IS distribution obtained by the exponential tilt is a very large improvement over crude Monte Carlo and logarithmically efficient, it still results in an unbounded relative error as  $p_t \rightarrow 0$ .

The Normal distribution is in the maximum domain of attraction of the Gumbel distribution ( $\xi = 0$ ) so our arguments suggest that we should use as IS distribution

$$H_0\left(\frac{y-d}{c}\right) = \exp(-e^{-(y-d)/c}) \tag{7}$$

with parameters  $c, d$  selected to match the distribution function  $(\Phi(y))^{k_2/p_t}$  (see Appendix A). Using this Gumbel distribution as an IS distribution permits a very substantial increase in efficiency.

To show how effective this is as an IS distribution, we simulate from the Gumbel distribution with cdf (7). The weights attached to a given IS simulated value of  $Y$  are the ratio of the two pdfs, the standard normal and the Gumbel, or

$$w(Y) = c_\theta \phi(Y) \exp\left(e^{-\frac{Y-d_\theta}{c_\theta}} + \frac{Y-d_\theta}{c_\theta}\right). \tag{8}$$

For example with  $t = 4.7534$ ,  $p_t = 10^{-6}$  and Gumbel parameters  $c = 0.20$  and  $d = 4.85$ , the relative error in  $10^6$  simulations was  $0.729n^{-1/2}$ . We can compare this with the exponential tilt, equivalent to using the normal( $t, 1$ ) distribution as an IS distribution, whose relative error is  $2.32n^{-1/2}$ , or with crude Monte Carlo, with relative error around  $10^3n^{-1/2}$ .

Suppose that our interest is in estimating the conditional tail expectation or  $\text{TVaR}_\alpha$  based on simulations. The  $\text{TVaR}_\alpha$  is defined as  $E(Y | Y > t) = \frac{E[YI(Y>t)]}{P(Y>t)}$ . We designed the GEV parameters for simulating the numerator,  $E[YI(Y > t)]$ . If we are interested in estimating  $\text{TVaR}_{0.0001}$  by simulation,  $t = \text{VaR}_{0.0001} = 3.719$  the true value is

$$\text{TVaR}_{0.0001} = \frac{\int_{3.719}^\infty z e^{-z^2/2} dz}{\int_{3.719}^\infty e^{-z^2/2} dz} \simeq 3.9585.$$

We will generate random variables  $Y_i$  using the Gumbel (0.282, 3.228) distribution and then attach weights (8) to these observations. The estimate of  $\text{TVaR}_\alpha$  is then the

**Table 1** Relative error of estimators of  $p_t$  and of  $E[YI(Y > t)]$

	$n^{1/2} \times \text{RE } \hat{p}_t$	$n^{1/2} \times \text{RE } \hat{E}[YI(Y > t)]$
1. Crude	100	100
2. SN (tilt)	2.05	2.05
3. EV IS	0.73	0.73
4. Cond EV IS	0.47	0.47

average of the values  $w(Y_i) \times Y_i$  averaged only over those values that are greater than  $t$ .

The Gumbel distribution is supported on the whole real line, while the region of interest is only that portion of the space greater than  $t$  so one might generate  $Y_i$  from the Gumbel distribution conditional on the event  $Y_i > t$  rather than unconditionally. The probability  $P(Y_i > t)$  where  $Y_i$  is distributed according to the Gumbel( $c_\theta, d_\theta$ ) distribution is  $\exp(-e^{-(t-d_\theta)/c_\theta})$  and this is typically around 0.80 indicating that about 20 % of the time the Gumbel random variables fall in the “irrelevant” portion of the sample space  $S < t$ . Since it is easy to generate from the conditional Gumbel distribution  $Y|Y > t$  this was also done for a further improvement in efficiency. This conditional distribution converges to the generalized Pareto family of distributions (see Theorem 2 of Appendix B). In this case, since  $\xi = 0$ ,  $P(Y - u \leq z|Y > u) \rightarrow 1 - e^{-z}$  as  $u \rightarrow \infty$ . Therefore, in order to approximately generate from the conditional distribution of the tail of the Gumbel, we generate the excess from an exponential distribution.

Table 1 provides a summary of the results of these simulations. Several simulation methods for estimating  $\text{TVaR}_\alpha = E(Y|Y > t) = \frac{E[YI(Y>t)]}{p_t}$  with  $p_t = P(Y > t)$  as well as estimates of  $p_t$  are compared. Since TVaR is a ratio, we consider estimates of the denominator and numerator, i.e.,  $p_t$  and  $E[YI(Y > t)]$  separately. The underlying distribution of  $Y$  is normal in all cases. The methods investigated are:

- 1. Crude Simulation (Crude)** Generate independently  $Y_i, i = 1, \dots, n$  from original (normal) distribution. Estimate  $p_t$  using  $\frac{1}{n} \sum_{i=1}^n I(Y_i > t)$  and estimate  $E[YI(Y > t)]$  using  $\frac{1}{n} \sum_{i=1}^n Y_i I(Y_i > t)$ .
- 2. Exponential Tilt or Shifted Normal IS (SN)** Generate independently  $Y_i, i = 1, \dots, n$  from  $N(t, 1)$  distribution. Estimate  $p_t$  using  $\frac{1}{n} \sum_{i=1}^n w_i I(Y_i > t)$  and estimate  $E[YI(Y > t)]$  using  $\frac{1}{n} \sum_{i=1}^n w_i Y_i I(Y_i > t)$  where  $w_i$  are the weights, obtained as the likelihood ratio

$$w_i = w(Y_i) = \frac{\phi(Y_i)}{\phi(Y_i - t)}.$$

Since the exponential tilt, applied to a Normal distribution, results in another Normal distribution with a shifted mean and the same variance, this is an application of the exponential tilt.

- 3. Extreme Value IS (EVIS)** Generate independently  $Y_i, i = 1, \dots, n$  from the Gumbel( $c, d$ ) distribution. Estimate  $p_t$  using  $\frac{1}{n} \sum_{i=1}^n w_i I(Y_i > t)$  and estimate

$E [YI(Y > t)]$  using  $\frac{1}{n} \sum_{i=1}^n w_i Y_i I(Y_i > t)$  where  $w_i$  are the weights, obtained as the likelihood ratio

$$w_i = w(Y_i) = \frac{\phi(Y_i)}{\frac{1}{c} h_0(\frac{Y_i-d}{c})}$$

where  $\frac{1}{c} h_0(\frac{Y_i-d}{c})$  is the corresponding Gumbel pdf.

4. **Conditional Extreme Value IS (Cond EVIS)** Generate independently  $Y_i, i = 1, \dots, n$  from the Gumbel( $c, d$ ) distribution conditioned on  $Y > t$ . Estimate  $p_t$  using  $\frac{1}{n} \sum_{i=1}^n w_i I(Y_i > t) = \frac{1}{n} \sum_{i=1}^n w_i$  and estimate  $E [YI(Y > t)]$  using  $\frac{1}{n} \sum_{i=1}^n w_i Y_i I(Y_i > t) = \frac{1}{n} \sum_{i=1}^n w_i Y_i$  where  $w_i$  are the weights, obtained as the likelihood ratio

$$w_i = w(Y_i) = \frac{\phi(Y_i)}{g(Y_i)}$$

and where

$$g(s) = \frac{\frac{1}{c} h_0(\frac{s-d}{c})}{1 - H_0(\frac{t-d}{c})}, \quad \text{for } s > t$$

is the corresponding conditional Gumbel pdf.

**Conditional Normal IS** Since we are interested in the tail behavior of the random variable  $Y$  given  $Y > t$  it would be natural to simulate from the conditional distribution  $Y|Y > t$ . Unfortunately, this is an infeasible method because it requires advance knowledge of  $p_s = P(Y > s)$  for all  $s \geq t$ .

We indicate in Table 1 the relative error of these various methods in the case  $p_t = 0.0001, t = 3.719$ . The corresponding parameters of the Gumbel distribution that we used were  $c = 0.243, d = 3.84$  but the results are quite robust to the values of these parameters. Notice that the efficiency gain of the conditional extreme value simulation, as measured by the ratio of variances, is around  $(\frac{100}{0.47})^2 \simeq 45, 270$  relative to a crude Monte Carlo and around  $(\frac{2.05}{0.47})^2 \simeq 19$  relative to the exponential tilt.

### 3.2 Example 2: Simulating a Portfolio Credit Risk Model

We provide a simulation of a credit risk model using importance sampling. The model, once the industry standard, is the normal copula model for portfolio credit risk, introduced in Morgan’s CreditMetrics system<sup>3</sup> (see [5]). Under this model, the  $k$ ’th firm defaults with probability  $p_k(Z)$ , and this probability depends on  $m$  unobserved factors that comprise the vector  $Z$ . Losses on a portfolio then take the form  $L = \sum_{k=1}^v c_k Y_k$  where  $Y_k$ , the default indicator, is a Bernoulli random variable with  $P(Y_k = 1) = p_k(Z)$ , (denoted  $Y_k \sim \text{Bern}(p_k(Z))$ ),  $p_k$  are functions of

---

<sup>3</sup> Very popular prior to 2008!

common factors  $Z$ , and  $c_k$  is the portfolio exposure to the  $k'$ th default. Suppose we wish to estimate  $P(L > t)$ .

### 3.2.1 One-Factor Case

In the simplest one-factor case,  $p_k$  are functions  $p_k(Z) = \Phi\left(\frac{a_k Z + \Phi^{-1}(\rho_k)}{\sqrt{1-a_k^2}}\right)$  of a common standard normal  $N(0, 1)$  random variable  $Z$ , the scalars  $a_k$  are the factor loadings or weights and  $\rho_k$  represents the marginal probability of default (it is easy to see that  $E[p_k(Z)] = \rho_k$ ).

If we wish to simulate an event  $P(L > t)$  which has small probability, there are two parallel opportunities for importance sampling, both investigated by [4].

For example, for each given  $Z$ , we might replace the Bernoulli distribution of  $Y_k$  with a distribution having higher probabilities of default, i.e., replace  $p_k(Z)$  by  $q_k$  where  $q_k \geq p_k(Z)$ . The choice of  $q_k$  is motivated by an exponential tilt as is argued in [4]. Conditional on the factor  $Z$ , the tilted Bernoulli random variables  $Y_k$  are such that  $E(\sum_{k=1}^v c_k Y_k | Z) = L$ . We do not require the use of importance sampling in this second stage of the simulation so having used an IS distribution for  $Z$ , we generate Bernoulli ( $p_i(Z)$ ) random variables  $Y_i$ . There are two similar alternatives in the first stage, generate  $Z$  from the Gumbel or generate  $\tilde{L} = \sum_{k=1}^v c_k p_k(Z)$ , a proxy for the loss, from the Gumbel distribution. These two alternatives give similar results since the Gumbel is the extreme value distribution corresponding to both  $Z$  and  $\tilde{L}$ , and  $\tilde{L}$  is a nondecreasing function of  $Z$ . In Table 2, we give the results corresponding to the second of these alternatives, simulating  $\tilde{L} = \sum_{k=1}^v c_k p_k(Z)$  and then solving for the factor  $Z$ . Unlike [4], where a shifted normal IS distribution for  $Z$  is used, we use the Gumbel distribution for  $L$  motivated by the arguments of Sect. 2. Extreme value importance sampling provides a very substantial variance reduction over crude simulation of course, but also over importance sampling using the exponential tilt. We determine appropriate parameters for the Gumbel extreme value distribution by quantile matching and then draw  $Z$  from a Gumbel( $c, d$ ) distribution. We use the parameters taken from the numerical example in [4], i.e.,  $v = 1,000$  obligors, exposures  $c_k$  are 1, 4, 9, 16, and 25 with 200 at each level of exposure, and the marginal default probabilities  $\rho_k = 0.01 (1 + \sin 16\frac{\pi k}{v})$  so that they range from 0 to 2%. The factor loadings  $a_k$  were generated as uniform random variables on the interval (0,1). In summary, the main difference with [4] is our use of the Gumbel distribution for simulating  $\tilde{L}$  rather than the shifted normal and the lack of a tilt for  $Y_k$ .

The resulting relative errors estimated from 30,000 simulations are shown in Table 2, and evidently there is a significant variance reduction achieved by the choice of the Gumbel distribution. For example, when the threshold  $t$  was chosen to be 2,000, there was a decrease in the variance by a factor of approximately  $(\frac{2.33}{0.69})^2$  or about 11

**Table 2** Relative error of estimators: Crude, shifted normal (G & L) and EVIS

$t$	$p_t$	$n$	$n^{1/2}$ RE(Crude)	$n^{1/2}$ RE(EVIS)	$n^{1/2}$ RE(G&L)
1,500	0.0075	30,000	12.01	0.70	2.03
2,000	0.0041	30,000	15.34	0.69	2.33
3,000	0.0022	30,000	22.12	0.70	2.74

and a much more substantial decrease over crude by a factor of around  $\left(\frac{15.34}{0.69}\right)^2$  or about 494.

### 3.2.2 Multifactor Case

In the multifactor case, the event that an obligor  $k$  fails is determined by a Bernoulli random variable  $Y_k \sim \text{Bern}(p_k)$ . The loss function  $L = \sum_{k=1}^v c_k Y_k$  is then a linear function of  $Y_k$  and corresponding exposures  $c_k$ . We wish to estimate the probability of a large loss:  $P(L > t)$ . The values  $p_k$  are functions

$$p_k(\mathbf{Z}) = \Phi\left(\frac{\mathbf{a}_k \mathbf{Z} + \Phi^{-1}(\rho_k)}{\sqrt{1 - \mathbf{a}_k \mathbf{a}_k^T}}\right)$$

of a number of factors  $\mathbf{Z}^T = (Z_1, \dots, Z_m)$  where the individual factors  $Z_i, i = 1, \dots, m$  are independent standard normal random variables. Here  $\rho_k$  is the marginal probability that obligor  $k$  fails, i.e.,  $P(Y_k = 1) = E[p_k(\mathbf{Z})] = \rho_k$  since  $\mathbf{a}_k \mathbf{Z}$  is  $N(0, \mathbf{a}_k \mathbf{a}_k^T)$  (see [4], p. 1644) and the row vectors  $\mathbf{a}_k$  are factor loadings which relate the factors to the specific obligors.

**Simulation Model** We begin with brief description of the model simulated in [4, p. 1650], that is the basis of our comparison. We assume  $v = 1,000$  obligors, the marginal probabilities of default  $\rho_k = 0.01(1 + \sin(16\pi \frac{k}{v}))$ , and the exposures  $c_k = [\frac{5k}{v}]^2, k = 1, \dots, v$ . The components of the factor loading vector  $\mathbf{a}_k$  were generated as independent  $U(0, \frac{1}{\sqrt{m}})$ , where  $m$  is the number of factors. The simulation described in [4] is a two-stage Monte Carlo IS method. The first stage simulates the latent factors in  $\mathbf{Z}$  by IS, where the importance distributions are independent univariate Normal distributions with means obtained by solving equating modes and with variances unchanged equal to 1. Specifically they choose the normal IS distribution having the same mode as

$$P(L > t | \mathbf{Z} = \mathbf{z}) e^{-\mathbf{z}^T \mathbf{z} / 2} \simeq \left(1 - \Phi\left(\frac{t - \mu(\mathbf{z})}{\sigma(\mathbf{z})}\right)\right) e^{-\mathbf{z}^T \mathbf{z} / 2} \tag{9}$$

because this is approximately proportional to  $P(\mathbf{Z} = \mathbf{z} | L > t)$ , the ideal IS distribution. In other words, the IS distribution for  $Z_i$  is  $N(\mu_i, 1)$ ,  $i = 1, \dots, m$  where the vector of values of  $\mu_i$  is given by (see [4], Eq.(20))

$$\mu = \max_{\mathbf{z}} P(L > t | \mathbf{Z} = \mathbf{z}) e^{-\mathbf{z}^T \mathbf{z} / 2} \tag{10}$$

with (see [4], p. 1648)

$$P(L > t | \mathbf{Z} = \mathbf{z}) \simeq 1 - \Phi \left( \frac{t - E[L | \mathbf{Z} = \mathbf{z}]}{\sqrt{\text{Var}[L | \mathbf{Z} = \mathbf{z}]}} \right) \text{ with } E[L | \mathbf{Z} = \mathbf{z}] > t. \tag{11}$$

Conditional on the values of the latent factors  $\mathbf{Z}$ , the second stage of the algorithm in [4] is to twist the Bernoulli random variables  $Y_k$  using modified Bernoulli distributions, i.e., with a suitable change in the values of the probabilities  $P(Y_k = 1)$ ,  $k = 1, \dots, v$ . Our comparisons below are with this two-stage form of the IS algorithm.

Our simulation for this portfolio credit risk problem is a one-stage IS simulation algorithm. If there are  $m$  factors in the portfolio credit risk model we simulate  $m - 1$  of them  $\tilde{Z}_i$  from univariate normal  $N(\mu_i, 1)$ ,  $i = 1, \dots, m - 1$  with a different mean, as in [4], but then we simulate an approximation to the total loss,  $\tilde{L}$ , from a Gumbel distribution, and finally set  $\tilde{Z}_m$  equal to the value implied by  $\tilde{Z}_1, \dots, \tilde{Z}_{m-1}$  and  $\tilde{L}$ . This requires solving an equation

$$\tilde{L}(\tilde{Z}_1, \dots, \tilde{Z}_{m-1}, \tilde{Z}_m) = \tilde{L} \tag{12}$$

for  $\tilde{Z}_m$ . The parameters  $\mu = (\mu_1, \mu_2, \dots, \mu_{m-1})$  are obtained from the crude simulation. Having solved (12), we attach weight to this IS point  $(\tilde{Z}_1, \dots, \tilde{Z}_m)$  equal to

$$\omega = \frac{c \times \prod_{i=1}^m \phi(\tilde{Z}_i)}{\frac{\partial L}{\partial Z_m} \times \prod_{i=1}^{m-1} \phi(\tilde{Z}_i - \mu_i)} e^{(\tilde{L}-d)/c} \exp \left( e^{-(\tilde{L}-d)/c} \right), \tag{13}$$

and

$$\frac{\partial L}{\partial \mathbf{Z}_m} = \sum_{k=1}^v \frac{a_{k,m}}{\sqrt{1 - \mathbf{a}_k \mathbf{a}_k^T}} \phi \left( \frac{\mathbf{a}_k \tilde{\mathbf{Z}} + \Phi^{-1}(\rho_k)}{\sqrt{1 - \mathbf{a}_k \mathbf{a}_k^T}} \right).$$

We choose the parameters  $\mu_i$ ,  $i = 1, \dots, m - 1$  for the above IS distributions using estimates of the quantity

$$\mu = E(\mathbf{Z} | L > t) = E \left( \mathbf{Z} \left| \sum_{k=1}^v c_k p_k(\mathbf{Z}) > t \right. \right) \tag{14}$$

based on the preliminary simulation, with the parameters  $c, d$  of the Gumbel obtained from (24).

We summarize our algorithm for the portfolio credit risk problem as follows:

1. Conduct a crude MC simulation and estimate the parameter  $\mu$  in (14).
2. Estimate parameters  $c$  and  $d$  of the Gumbel distribution (24) where  $E(\tilde{L})$  is estimated by  $\text{average}(L|L > t)$ .
3. Repeat (a)–(d) for independent simulations as  $j = 1, \dots, n$  where  $n$  is the sample size of the simulation.
  - (a) Generate  $\tilde{L}$  from the Gumbel( $c, d$ ) distribution.
  - (b) Generate  $\tilde{Z}_i, i = 1, \dots, m - 1$  from the univariate normal  $N(\mu_i, 1)$  distributions.
  - (c) Solve  $\tilde{L}(\tilde{Z}_1, \dots, \tilde{Z}_{m-1}, \tilde{Z}_m) = t$  for  $\tilde{Z}_m$  and calculate (13).
  - (d) Simulate a loss  $L_j = \sum_{k=1}^v c_k Y_k$  where  $Y_k \sim \text{Bern}(p_k(\tilde{\mathbf{Z}}))$  with  $p_k(\tilde{\mathbf{Z}}) = \Phi\left(\frac{\mathbf{a}_k \tilde{\mathbf{Z}} + \Phi^{-1}(p_k)}{\sqrt{1 - \mathbf{a}_k \mathbf{a}_k^T}}\right)$ .
4. Estimate  $p_t$  using a weighted average

$$\frac{1}{n} \sum_{j=1}^n \omega_j I(L_j > t).$$

5. Estimate the variance of this estimator using  $n^{-1}$  times the sample variance of the values  $\omega_j I(L_j > t), j = 1, \dots, n$ .

**Simulation Results** The results in Table 3 were obtained by using crude Monte Carlo, importance sampling using the GEV distribution as the IS distribution, and the IS approach proposed in [4]. In the crude simulations, the sample size is 50,000, while in the later two methods, the sample size is 10,000.

Notice that for a modest number of factors there is a very large reduction in variance over the crude (for example the ratio of relative error corresponding to 2 factors,  $t = 2,500$  corresponds to an efficiency gain or variance ratio of nearly 2,400) and a significant improvement over the Glasserman and Li [4] simulation with a variance ratio of approximately 4. This improvement erodes as the number of factors increases, and in fact the method of Glasserman and Li has smaller variance in this case when  $m = 10$ . In general, ratios of multivariate densities of large dimension tend to be quite “noisy”; although the weights have expected value 1, they often have large variance. A subsequent paper will deal with the large dimensional case.

**Table 3** Comparison between crude simulation, EVIS and Glasserman and Li (2005) for the credit risk model

$t$	$p_t$	$n$	$n^{1/2}$ RE (crude)	$n^{1/2}$ RE (EVIS)	$n^{1/2}$ RE (G&L)
2 factors					
1,500	0.0034	50,000	17.1	0.99	1.73
2,000	0.0015	10,000	26.2	0.96	1.82
2,500	0.00038	10,000	51.3	1.05	1.93
3 factors					
1,500	0.00305	50,000	18.94	1.24	1.72
2,000	0.00111	10,000	31.61	1.15	1.82
2,500	0.00042	10,000	49.99	1.35	1.99
5 factors					
1,500	0.00289	50,000	18.87	1.39	1.71
2,000	0.00099	10,000	39.52	1.55	1.81
2,500	0.00035	10,000	55.89	1.57	1.88
10 factors					
1,500	0.00246	50,000	20.83	1.84	1.79
2,000	0.00081	10,000	33.70	2.15	1.89
2,500	0.00029	10,000	57.73	3.06	1.98

## 4 Conclusion

The family of extreme value distributions are ideally suited to rare event simulation. They provide a very tractable family of distributions and have tails which provide bounded relative error regardless of how rare the event is. Examples of simulating values of risk measures demonstrate a very substantial improvement over crude Monte Carlo and a smaller improvement over competitors such as the exponential tilt. This advantage is considerable for relatively low-dimensional problems, but there may be little or no advantage over an exponential tilt when the dimensionality of the problem increases.

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Appendix A: Assumptions and Results

We suppose without loss of generality that the argument to the loss function is a multivariate normal  $MNV(0, I_m)$  random vector  $\mathbf{Z}$ , since any (possibly dependent) random vector  $\mathbf{Y}$  can be generated from such a  $\mathbf{Z}$ . We begin by assuming that “large”

values of  $L(\mathbf{Z})$  are determined by the distance of  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_m)$  from the origin in a specific direction, i.e.,

**Assumption 1** There exists a direction vector  $\mathbf{v} \in \mathfrak{R}^m$  such that, for all fixed vectors  $\mathbf{w} \in \mathfrak{R}^m$ ,

$$\frac{P(L(\mathbf{Z}_0\mathbf{v}) > t)}{P(L(\mathbf{Z}_0\mathbf{v} + \mathbf{w}) > t)} \rightarrow 1 \text{ as } t \rightarrow \infty \tag{15}$$

where  $Z_0$  is  $N(0, 1)$ .

We propose an importance sampling distribution generated as follows:

$$\mathbf{Z} = Y\mathbf{v} + (I_m - \mathbf{v}\mathbf{v}')\boldsymbol{\varepsilon}, \quad \text{where } \boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \mathbf{I}_m), \tag{16}$$

where  $Y$  has the extreme value distribution  $H_0(\frac{y-d}{c})$ . If we replace the distribution of  $Y$  by the standard normal, it is easy to see that (16) gives  $\mathbf{Z} \sim \text{MVN}(\mathbf{0}, \mathbf{I}_m)$  so the IS weight function in this case is simply the ratio of the two univariate distributions for  $Y$ .

**Assumption 2** Suppose that for any fixed  $\mathbf{w} \in \mathfrak{R}^m$ , there exists  $y_0$  such that  $L(y\mathbf{v} + \mathbf{w})$  is an increasing function of  $y$  for  $y > y_0$ .

**Proposition 1** Under assumptions 1 and 2, there is a sequence of importance sampling distributions of the form (16) which provides bounded relative error asymptotic to  $cn^{-1/2}$  as  $p_t \rightarrow 0$  where  $c \simeq 0.738$ .

In order to prove this result, we will use the following lemma, a special case of Corollary 1 of [11]:

**Lemma 1** Suppose the random variable  $Y$  has a continuous distribution with cdf  $F_Y$ . Suppose that  $T(y)$  is nondecreasing and for some real number  $a$  we have  $a + T(y) \sim -\bar{F}_Y(y)$  as  $y \rightarrow y_F^-$  with  $y_F = \sup\{y; F_Y(y) < 1\} \leq \infty$ . Then the IS estimator for sample size  $n$  obtained from density (3) with  $\theta = \theta_t = \frac{k_2}{p_t}$  has bounded RE asymptotic to  $cn^{-1/2}$  as  $p_t \rightarrow 0$  where  $c = \frac{1}{k_2} \sqrt{e^{k_2} - 1 - k_2^2} \simeq 0.738$ .<sup>4</sup>

*Proof of Proposition 1.* The condition (15) allows us to solve an asymptotically equivalent univariate problem, i.e., estimate  $P(L_1(Y) > t)$  where  $L_1(Y) = L(Y\mathbf{v})$ ,  $Y \sim N(0, 1)$ . Clearly, the Normal distribution for  $Y$  satisfies  $F_Y \in \text{MDA}(H_0(x))$  so that there exist sequences  $c_n, d_n$  such that  $F_Y^n(d_n + c_n x) \rightarrow H_0(x)$  as  $n \rightarrow \infty$  for the GEV  $H_0$ . Lemma 1 shows that the importance sampling distribution

$$f_\theta(y) = \text{constant} \times e^{-\frac{k_2}{p_t} \bar{F}_Y(y)} f(y) \tag{17}$$

---

<sup>4</sup>  $k_2 \simeq 1.5936$  and  $c \simeq 0.738$  are the unique positive solutions to the equations  $e^k = \frac{1}{1-\frac{k}{2}} = 1 + k^2(1 + c^2)$ .

provides bounded relative error for the estimation of  $p_t$  as  $t \rightarrow \infty$  and  $p_t \rightarrow 0$ . Note that the probability density function of the maximum  $n = \frac{k_2}{p_t} + 1$  random variables drawn from the distribution  $F_Y(y)$  is given by

$$n (F_Y(y))^{n-1} f(y) = \text{constant} \times (1 - \overline{F}_Y(y))^{\frac{k_2}{p_t}} f(y) \sim \text{constant} \times e^{-\frac{k_2}{p_t} \overline{F}_Y(y)} f(y). \tag{18}$$

Furthermore, by the local limit or density version of convergence to the extreme value distributions, (see Theorem 2 (b) [2] or [14]), with  $y = d_n + c_n x$ , and  $x = (y - d_n)/c_n$ ,

$$n F_Y^{n-1}(d_n + c_n x) f(d_n + c_n x) \rightarrow H_0'(x) = \frac{1}{c_n} \exp(-x - e^{-x}) \text{ as } n \rightarrow \infty$$

which implies, combining (17) and (18), that

$$n c_n (F_Y(y))^n f(y) \sim \exp\left(- (y - d_n)/c_n - e^{-(y-d_n)/c_n}\right). \tag{19}$$

Therefore, the extreme value distribution provides a bounded relative error importance sampling distribution, equivalent to (17).

## Appendix B: Maximum Domain of Attraction and Properties of The Generalized Extreme Value Distributions

### Maximum domain of attraction

If there are sequences of real constants  $c_n$  and  $d_n$ ,  $n = 1, 2, \dots$  where  $c_n > 0$  for all  $n$ , such that

$$F^n(d_n + c_n x) \rightarrow H(x) \text{ as } n \rightarrow \infty, \tag{20}$$

for some nondegenerate cdf  $H(x)$ , then we say that  $F$  is in the maximum domain of attraction (MDA) of the cdf  $H$  and write  $F \in \text{MDA}(H)$ . The Fisher–Tippett theorem (see Theorem 7.3 of [13]) characterizes the possible limiting distributions  $H$  as members of the generalized extreme value distribution (GEV). A cdf is a member of this family if it has cumulative distribution function of the form  $H_\xi(\frac{x-d}{c})$  where  $c > 0$  and

$$H_0(x) = \exp(-e^{-x}), H_\xi(x) = e^{-(1+\xi x)^{-1/\xi}} \text{ for } \xi \neq 0 \text{ and } \xi x > -1. \tag{21}$$

**Theorem 1** (Fisher–Tippet, Gnedenko) *If  $F \in \text{MDA}(H)$  for some nondegenerate cdf  $H$ , then  $H$  must take the form (21).*

The properties of the GEV distributions listed in Table 4 are obtained from routine calculations and properties in [13] or [10].

**Table 4** Some properties of the generalized extreme value distributions

Property	$\xi = 0$	$\xi \neq 0$ and $\xi x > -1$
cdf = $H_\xi(x)$	$\exp(-e^{-x})$	$\exp(-(1 + \xi x)^{-1/\xi})$
pdf = $h_\xi(x)$	$e^{-x} \exp(-e^{-x})$	$(1 + \xi x)^{-\frac{1}{\xi}-1} \exp(-(1 + \xi x)^{-1/\xi})$
Mode: satisfies $H_\xi(x) = \exp(-1 - \xi)$	0	$\frac{(1 + \xi)^{-\xi} - 1}{\xi}$
$H_\xi(x x > x_{1/2})$	$\frac{\exp(1 - e^{-x})}{e - 1}$ , for $x > x_{1/2}$	$\frac{\exp(-(1 + \xi x)^{-1/\xi})}{1 - \exp(-1 - \xi)}$ , for $x > x_{1/2}$
Median	$x_{1/2} = -\ln(\ln 2)$	$\frac{(\ln 2)^{-\xi} - 1}{\xi}$
Inverse $H_\xi^{-1}(p)$	$-\ln(-\ln p)$	$\frac{(-\ln p)^{-\xi} - 1}{\xi}$
Mean	$\gamma = (\text{Euler's constant}) \simeq 0.577216$	$\begin{cases} \frac{\Gamma(1 - \xi) - 1}{\xi} & \text{if } \xi < 1 \\ \infty & \text{if } \xi \geq 1 \end{cases}$
Variance	$\frac{\pi^2}{6} \simeq 1.645$	$\begin{cases} \frac{\Gamma(1 - 2\xi) - 2\Gamma(1 - \xi) + 1}{\xi} & \text{if } \xi < \frac{1}{2} \\ \infty & \text{if } \xi \geq \frac{1}{2} \end{cases}$
Random number generator $U \sim U(0, 1)$	$-\ln(-\ln U)$	$\frac{(-\ln U)^{-\xi} - 1}{\xi}$

**Choosing the parameters  $c$  and  $d$**

The GEV has  $H_\xi(\frac{y-d}{c})$  and probability density function  $c^{-1}h_\xi(\frac{y-d}{c})$ . Other parameters can be easily found in the above table. We wish to choose an extreme value distribution with parameters corresponding to the maximum of a sample of  $\theta_t = k_2/p_t$  random variables from the original density  $f(y)$ . In other words, we wish to find values of  $d_{\theta_t}$  and  $c_{\theta_t}$  so that

$$(F(y))^{\theta_t} \simeq H_\xi\left(\frac{y - d_{\theta_t}}{c_{\theta_t}}\right) \tag{22}$$

and this leads to matching  $t$  with the quantile corresponding to  $e^{-k_2} \simeq 0.203$ . In other words, one parameter is determined by the equation

$$\frac{t - d_{\theta_t}}{c_{\theta_t}} = H_\xi^{-1}(e^{-k_2}) = \begin{cases} -\ln(k_2) & \xi = 0 \\ \frac{k_2^{-\xi} - 1}{\xi} & \xi \neq 0 \end{cases} \tag{23}$$

Another parameter can be determined using the crude simulation and the values of  $Y$  for which  $L(Y) > t$ . We can match another quantile, for example, the median, the mode, or the sample mean which estimates  $E[Y | L(Y) > t]$ . In the case of standard normally distributed inputs and the Gumbel distribution, matching the conditional expected value  $E(L|L > t)$  and (23) results approximately in:

$$c = \frac{E(\tilde{L}) - t}{1.0438}, \quad \text{and } d = t + 0.46659c. \tag{24}$$

Here  $E(\tilde{L}) = \text{average}(L|L > t)$  based on a preliminary crude simulation of values of  $L$  simulated under the original distribution. Of course, one could also use maximum likelihood estimation to determine appropriate parameters for the ID distribution (see [10]) but the specific choice of estimator seemed to have little impact on the quality of the importance sampling provided that the estimated GEV density was sufficiently dispersed.

As an alternative to simulating  $\tilde{L}$  from the GEV, we may simulate instead from  $\tilde{L}|\tilde{L} > t$ , resulting in the generalized Pareto distribution. For a given c.d.f.  $F$ , the **conditional excess distribution** is

$$F_u(y) = P(X - u > y|X > u) = \frac{F(u + y) - F(u)}{1 - F(u)}, \quad x \geq 0.$$

Then the conditional excess distribution can be approximated by the so-called generalized Pareto distribution for large values of  $u$  (see [13], Theorem 7.20):

**Theorem 2** (Pickands, Balkema, de Haan)  $F \in \text{MDA}(H_\xi)$  for some  $\xi$  if and only if

$$\lim_{u \rightarrow \infty} \sup_x |F_u(x) - G_{\xi, \beta(u)}(x)| \rightarrow 0$$

for some positive measurable function  $\beta(u)$  where  $G_{\xi, \beta}$  is the c.d.f of the **Generalized Pareto (GP) distribution**:

$$G_{\xi, \beta}(y) = \begin{cases} 1 - (1 + \frac{\xi y}{\beta})^{-1/\xi} & \text{for } \xi > 0, \quad y > 0 \text{ or} \\ & \xi < 0, \quad 0 < y < \frac{\beta}{-\xi} \\ 1 - e^{-y/\beta} & \text{for } \xi = 0, \quad y > 0 \end{cases} \quad (25)$$

## References

1. Asmussen, S., Glynn, P.W.: Stochastic Simulation: Algorithms and Analysis. Springer, New York (2007)
2. De Haan, L., Resnick, S.I.: Local limit theorems for sample extremes. Ann. Probab. **10**, 396–413 (1982)
3. Glasserman, P.: Monte Carlo Methods in Financial Engineering. Springer, New York (2004)
4. Glasserman, P., Li, J.: Importance sampling for portfolio credit risk. Manag. Sci. **51**(11), 1643–1656 (2005)
5. Gupton, G., Finger, C.C., Bhatia, M.: CreditMetrics, Technical Document of the Morgan Guaranty Trust Company [http://www.defaultrisk.com/pp\\_model\\_20.htm](http://www.defaultrisk.com/pp_model_20.htm) (1997)
6. Hall, P.: On the rate of convergence of normal extremes. J. Appl. Probab. **16**, 433–439 (1979)
7. Homem-de-Mello, T., Rubinstein, R.Y.: Rare event probability estimation using cross-entropy. In: Proceedings of the 2002 Winter Simulation Conference, pp. 310–319 (2002)
8. Kroese, D., Rubinstein, R.Y.: Simulation and the Monte Carlo Method, 2nd edn. Wiley, New York (2008)
9. Kroese, D., Taimre, T., Botev, Z.I.: Handbook of Monte Carlo Methods. Wiley, New York (2011)

10. Kotz, S., Nadarajah, S.: *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London (2000)
11. McLeish, D.L.: Bounded relative error importance sampling and rare event simulation. *ASTIN Bullet.* **40**, 377–398 (2010)
12. McLeish, D.L.: *Monte Carlo Simulation and Finance*. Wiley, New York (2005)
13. McNeil, A.J., Frey, R., Embrechts, P.: *Quantitative Risk Management*. Princeton University Press, Princeton (2005)
14. Pickands, J.: Sample sequences of maxima. *Ann. Math. Stat.* **38**, 1570–1574 (1967)
15. Ridder, A., Rubinstein, R.: Minimum cross entropy method for rare event simulation. *Simulation* **83**, 769–784 (2007)