Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

RARE-EVENT, HEAVY-TAILED SIMULATIONS USING HAZARD FUNCTION TRANSFORMATIONS, WITH APPLICATIONS TO VALUE-AT-RISK

Zhi Huang
Perwez Shahabuddin

Department of Industrial Engineering and Operations Research
Columbia University
New York, NY 10027, U.S.A

ABSTRACT

We develop an observation that a simulation method introduced recently for heavy-tailed stochastic simulation, namely hazard-rate twisting, is equivalent to doing exponential twisting on a transformed version of the heavy-tailed random-variable; the transforming function is the hazard function. Using this approach, the paper develops efficient methods for computing portfolio value-at-risk (VAR) when changes in the underlying risk factors have the multivariate Laplace distribution.

1 INTRODUCTION

We consider the problem of estimating the *small* probability that a random variable that is an output to a simulation exceeds a large threshold. The output random variable may be a function of several input random variables, and is generated via generating the input random variables. Importance sampling based simulations of such problems have been studied extensively in the context where the input random variables and the output random variables are light-tailed (see, e.g., Bucklew 1990, Heidelberger 1995 for expositions). We consider estimation of such probabilities for the case where some or all of these random variables may have heavy-tailed distribution, i.e., distributions whose tail decay at a subexponential rate.

Rare-event simulation in the heavy-tailed context seems to be a challenging problem. One of the reasons is that "exponential twisting", that is the main importance sampling framework in the light-tailed setting, cannot be used on heavy-tailed random variables. Hence new and innovative methods are needed. Till date, provably efficient simulation techniques and changes of measures exist only for estimating the probability that a sum of a fixed or a geometric number of i.i.d heavy-tailed random variable exceeds a large threshold (Asmussen and Binswanger 1997, Asmussen, Binswanger and Hojgaard 2000, and Juneja and Shahabuddin 2002). This has applications only in some very simple models in queueing and insurance. Some partial success has been obtained for the case of random walks, for some specific heavy-tailed distributions (Boots and Shahabuddin 2000).

This paper attempts to build on a technique introduced in Juneja and Shahabuddin (2002) called hazard rate twisting. Hazard rate twisting of a heavy-tailed random variable involves twisting at a sub-exponential rate, rather than at a exponential rate as is done in exponential twisting. Concavity properties of the hazard function are used in Juneja and Shahabuddin (2002) to prove asymptotic (i.e., as the threshold tends to ∞) efficiency in the simulation of sums and geometric sums of heavy-tailed random variables. We develop the observation that hazard rate twisting of any random variable is equivalent to exponentially twisting the hazard function transformation of the random variable. We then give conditions under which one can use the latter approach to estimate probabilities for more complicated random structures as compared to sums and geometric sums. At the more conceptual level, whereas Juneja and Shahabuddin (2002) thought in terms of hazard rate twisting of the input random variables, we think in terms of hazard rate twisting of the output random variable, and then work backward to find the corresponding changes of measure on the input random variables (that may *not* necessarily be hazard rate twisting). Interpreting hazard rate twisting in terms of hazard function transformations facilitates this approach, as then one can use the experience accumulated in light-tailed, importance sampling simulations.

We then apply this approach to the value-at-risk problem. The value-at-risk is an important concept for quantitifying and managing portfolio risk (see, e.g., Jorion 1997, Wilson 1999). One core problem from the simulation methodology point of view is to estimate the risk of large portfolio losses in given time intervals, where the value of the portfolio depends on several time-dependent and correlated risk factors. Recently new simulation approaches based on importance sampling and stratification have been developed for this problem under different assumptions on the risk factors. Glasserman, Heidelberger and Shahabuddin (2000) developed fast simulation methods when the risk factors have the multivariate normal distribution. They developed a provably efficient, importance sampling change of measure for estimating the risk, when the loss function is replaced by a "quadratic approximation". They then used the same change of measure for estimating the risk associated with the actual loss function, and obtained orders of magnitude of variance reduction in the estimation. Stratification on the quadratic approximation made the technique even more efficient.

However, it has been observed that market returns exhibit systematic deviations from normality in terms of the tail-weight. Two different families of tail behaviors have been advanced in the literature (see, e.g., Heyde and Kou (2002)and references therein). The first is polynomial type tails of which the multivariate t distribution is an example. The second is exponential type tails (which is still an order of magnitude heavier than the Gaussian type tails) of which the multivariate Laplace distribution (see, e.g., Kotz, Kozubowski and Podgorski 2001) is an example.

Glasserman, Heidelberger and Shahabuddin (2002) extend the work in their earlier paper to the case where the risk factors have the multivariate t distribution of the type in Anderson (1984). In this paper we consider the case where the risk factors have the multivariate Laplace distribution. Unlike the case in Glasserman, Heidelberger and Shahabuddin (2000), in both the above cases the quadratic approximation is heavy-tailed and thus, as mentioned above, necessitates the development of new ideas not found in the predominantly light-tailed importance sampling literature. It should be mentioned here that Glasserman, Heidelberger and Shahabuddin (2002) also used a tranformation approach that changes their problem into a light-tailed estimation problem. However the particular transformation they consider is specific to the multivariate t and not easily generalizable to other assumptions on the risk factors like the one we have.

2 A GENERAL PROBLEM

Let $X = (X_1, \ldots, X_m)$ be a vector of independent, non-negative random variables. For simplicity in presentation we will assume that each X_i has a probability density function (pdf) $f_i(x)$ which is positive at all points on $(0, \infty)$. Let the cumulative distribution function (cdf) be $F_i(x)$, and let $\bar{F}_i(x) = 1 - F(x)$. Also let $\lambda_i(x) = f_i(x)/\bar{F}_i(x)$ be the hazard rate function, and $\Lambda_{X_i}(x) = \int_0^x \lambda_i(s)ds$ be the hazard function. Hence $\Lambda_{X_i}(0) = 0$. The assumption on the $f_i(x)$ implies that $\lambda_i(x) > 0$ for all x in $(0, \infty)$, which implies that $\Lambda_{X_i}(x)$ is strictly increasing. It is also well known that $\Lambda_{X_i}(x) = -\ln(\bar{F}_i(x))$. For any generic random variable W we will let $\Lambda_W(x)$ denote its hazard function. For any two functions, say $g_1(x)$ and $g_2(x)$, $g_1(x) \sim g_2(x)$ denotes that $\lim_{x\to\infty} g_1(x)/g_2(x)$ exists and it is equal to 1.

We consider the problem of estimating $\alpha(y) = P(Y > y)$ where $Y = h(X)$ (we assume the distribution of Y to have the same regularity properties as those of the X_i's). We will assume $h(x_1, \ldots x_n)$ to be a non-negative-valued function, but the theory presented below also goes through for the general case if $h^+(X) = \max(h(X), 0)$ and $h(X)$ have the same decay behavior in their tail probabilities, i.e., $\Lambda_{h(X)}(y) \sim \Lambda_{h^+(X)}(y)$. There is signicant work in the literature when Y is light-tailed. We concern ourselves with the more general case when Y may be light-tailed or heavy-tailed.

3 PRELIMINARIES

3.1 Importance Sampling and Exponential Twisting

For y large, the event $\{Y > y\}$ may be rare, and we use importance sampling to simulate for $P(Y > y)$ more efficiently. In particular, if $\tilde{f}_i(x)$ is a new probability density function for X_i, with the same support as X_i, then we may express

$$P(Y > y) = E(I(Y > y)) = \tilde{E}(I(Y > y)l(X)) \qquad (1)$$

where

$$l(x_1, \ldots, x_n) = \prod_{i=1}^{m} \frac{f_i(x_i)}{\tilde{f}_i(x_i)},$$

and the $\tilde{E}(\cdot)$ indicates that the X_i's have the new pdf, i.e., the $\tilde{f}_i$'s. The quantity within the expectation on the RHS (right-hand side) of (1) forms an unbiased, "importance sampling" estimator of $P(Y > y)$.

The attempt is to find $\tilde{f}_i$'s so that the variance of this new estimator is as low as possible. More specifically, we want to $\tilde{E}(I(Y > y)l^2(X))$ to be the least possible. The change of measure $(\tilde{f}_1, \ldots, \tilde{f}_n)$ is called "asymptotically logarithmically efficient" (also called asymptotically efficient) if

$$\liminf_{y\to\infty} \frac{\ln \tilde{E}(I(Y > y)l^2(X))}{2 \ln \alpha(y)} \geq 1. \qquad (2)$$

This means that the exponential rate of decrease of the second moment is twice the exponential rate of decrease of the probability one is trying to estimate. Non-negativity of the variance implies that this is the fastest possible rate for any unbiased estimator. This is the reason why asymptotic logarithmic efficiency is also called "asymptotic logarithmic optimality" or simply "asymptotic optimality'. Note that for standard simulation, $(\ln \tilde{E}(I(Y>y)l(X)))/(2\ln\alpha(y)) \sim 1/2$.

Define a random variable, say X_i, to be light-tailed if $\liminf_{x\to\infty} \Lambda_i(x)/x \geq \lambda_i$ for some postive constant λ_1. For light-tailed random variables one may use a special change of measure that is obtained by "exponential twisting". For example, for the case of $f_i(x)$, the exponentially twisted density by amount θ, $\theta > 0$, is given by

$$f_{i,\theta}(x) = \frac{f_i(x)e^{\theta x}dx}{M_{X_i}(\theta)},$$

where $M_W(\theta)$ denotes the moment generating function (mgf) of the random variable W. Note that if $\Lambda_{X_i}(x) \sim \lambda_i x$ for some $\lambda_i > 0$, then exponential twisting makes sense only when $\theta \leq \lambda_i$. In many cases below we will use $\tilde{f}_i(x) = f_{i,\theta}(x)$ for some appropriately chosen θ.

3.2 Light-Tailed Simulations

Consider the case when Y is light-tailed. For simplicity in presentation, we will consider the special case when Y has an "exponential tail", i.e., $\Lambda_Y(y) \sim \lambda y$ for some $\lambda > 0$. In that case the attempt in the literature is to choose $\tilde{f}_1, \ldots, \tilde{f}_m$, so that one "achieves" exponential twisting on the Y by amount θ. If one is able to do that, then by the definition of exponential twisting, $l(X) = M_Y(\theta)e^{-\theta Y}$. Hence

$$\begin{aligned} \tilde{E}(I(Y>y)l^2(X)) &= \tilde{E}(I(Y>y)M_Y^2(\theta)e^{-2\theta Y}) \\ &\leq M_Y^2(\theta)e^{-2\theta y}. \end{aligned} \tag{3}$$

One then chooses θ to minimize $M_Y^2(\theta)e^{-2\theta y}$ or equivalently to minimize $\ln M_Y(\theta) - \theta y$. It can be shown that appropriate convexity properties hold so that the optimal solution, θ_y^*, may be obtained as the solution of the equation (see, e.g., Bucklew 1990 for this and other results mentioned here)

$$\frac{M_Y'(\theta)}{M_Y(\theta)} = y. \tag{4}$$

The θ_y^* is continuous and increasing in y, and $\lim_{y\to\infty}\theta_y^* = \lambda$. It is also known that (except for some pathological examples),

$$-\ln\left\{M_Y(\theta_y^*)e^{-\theta_y^* y}\right\} \sim \lambda y. \tag{5}$$

Using (2),(3) (with θ replaced by θ_y^*) and (5), one can infer that doing exponential twisting on Y by amount θ_y^* is asymptotically logarithmically efficient.

We still need to show as to how to "achieve" the exponential twisting on the Y by any amount θ, $0 < \theta < \lambda$, i.e., what $\tilde{f}_i$ should one choose for the X_i. We will illustrate this for a case where exponential twisting is very useful, i.e., where $Y = \sum_{i=1}^m X_i$, and X_i's are light-tailed random variables. Consider doing exponential twisting by amount θ on X_i. Then one can easily see that

$$l(X) = \prod_{i=1}^m (M_{X_i}(\theta)e^{-\theta X_i}) = M_Y(\theta)e^{-\theta Y},$$

and hence we have achieved exponential twisting on Y by amount θ.

4 THE HAZARD FUNCTION TRANSFORMATION APPROACH

4.1 Heavy-Tailed Random Variables and Hazard Function Transformations

Now consider estimating $\alpha = P(Y > y)$, where Y is a heavy-tailed random variable. For the purposes of this paper we will characterize random variables being heavy or light-tailed based on their hazard function $\Lambda(x)$. In particular, we assume heavy-tailed to mean that $\Lambda(x)/x \to 0$ as $x \to \infty$. Three common examples are $Weibull(\lambda,\alpha)$, $\alpha < 1$, with $\Lambda(x) = \lambda^\alpha x^\alpha$, $Lognormal(\mu,\sigma^2)$ with $\Lambda(x) \sim \ln^2(x)/(2\sigma^2)$, and the $Pareto(\lambda,\alpha)$ with $\Lambda(x) = \alpha\ln(1+\lambda x)$. Note that the above three distributions are ordered with respect to increasing heaviness of their tails.

Exponential twisting is not defined for heavy-tailed Y since the mgf $M_Y(\theta)$ is not defined for $\theta > 0$. Hence the approach described in Section 3.2 cannot be used here. However, it can be trivially shown that $\Lambda_Y(Y)$ is exponentially distributed with rate 1 (see Lemma 4.1). Also, by the monotonically increasing property of $\Lambda_Y(y)$, $P(Y > y) = P(\Lambda_Y(Y) > \Lambda_Y(y))$. Hence one has transformed a heavy-tailed estimation problem into a light-tailed one!

Lemma 4.1 *Let W be a random variable with increasing and continuous $\Lambda_W(y)$. Then $\Lambda_W(W)$ is an exponential random variable with rate 1.*

Proof. Since the hazard function $\Lambda_W(y)$ is strictly increasing and continuous, the inverse $\Lambda_W^{-1}(y)$ is defined, and is also increasing and continuous. Hence

$$\begin{aligned} P(\Lambda_W(W) > y) &= P(\Lambda_W^{-1}(\Lambda_W(W)) > \Lambda_W^{-1}(y)) \\ &= P(W > \Lambda_W^{-1}(y)) \\ &= e^{-\Lambda_W(\Lambda_W^{-1}(y))} = e^{-y}. \end{aligned}$$

■

However it is usually not possible to know $\Lambda_Y(y)$ (otherwise one can trivially compute $\alpha(y)$). In those cases one uses the transformation $\Lambda(Y)$, instead of $\Lambda_Y(Y)$ where $\Lambda(y)$ is an increasing continuous function such that $\Lambda(y) \sim \Lambda_Y(y)$. Then one is still assured that $\Lambda(Y)$ is a random variable with its tail probability decaying exponentially at rate 1.

Lemma 4.2 *Suppose $\Lambda(y) \sim \Lambda_Y(y)$ with $\Lambda(y)$ continuous and increasing for all sufficiently large y. Then*

$$\lim_{y\to\infty} -\frac{\log P(\Lambda(Y) > y)}{y} = 1.$$

Proof.

$$\begin{aligned}
&\lim_{y\to\infty} -\frac{\log P(\Lambda(Y) > y)}{y} \\
&= \lim_{y\to\infty} -\frac{\log P(Y > \Lambda^{-1}(y))}{y} \\
&= \lim_{y\to\infty} \frac{\Lambda_Y(\Lambda^{-1}(y))}{\Lambda(\Lambda^{-1}(y))} \\
&= \lim_{z\to\infty} \frac{\Lambda_Y(z)}{\Lambda(z)} = 1.
\end{aligned}$$

Note that the third equality follows by making the change of variable $z = \Lambda^{-1}(y)$, and then making use of the fact that $\Lambda(y)$ is increasing and continuous for all sufficiently large y. ■

In most cases it is easy to determine such a $\Lambda(y)$. For example, consider the case where $Y = \sum_{i=1}^m X_i$ where the X_i's are i.i.d. and $Weibull(\lambda, \alpha)$, $\alpha < 1$. Hence $\Lambda_{X_1}(x) = \lambda^\alpha x^\alpha$. Now the $Weibull(\lambda, \alpha)$, $\alpha < 1$, belongs to a large class of heavy-tailed distributions called subexponential distribution, for which

$$\begin{aligned}
P(\sum_{i=1}^m X_i > y) &\sim P(\max(X_1, \ldots X_m) > y) \\
&\sim mP(X_1 > y) \qquad (6)
\end{aligned}$$

(see, e.g., Embrechts, Klüppelberg and Mikosch 1997). Hence

$$P(Y > y) = P(\sum_{i=1}^m X_i > y) \sim me^{-\lambda^\alpha y^\alpha},$$

and one may choose $\Lambda(y) = \lambda^\alpha y^\alpha$.

In summary, the hazard function transformation approach is based on first recognizing the fact that with $\Lambda(y)$ satisfying the properties in Lemma 4.2, $\Lambda(Y)$ is a light-tailed random variable, and that due to the monotonicity of $\Lambda(y)$,

$$P(Y > y) = P(\Lambda(Y) > \Lambda(y)).$$

So once again we have a light-tailed problem, and we can use exactly the same procedure as we did to estimate $P(Y > y)$ when Y was light-tailed. In particular, we would now try to find new pdfs for the X_i's, $\tilde{f}_i$'s, that achieve exponential twisting of the $\Lambda(Y)$ by amount θ. The optimal θ, θ_y^*, will now be obtained as the solution of

$$\frac{M'_{\Lambda(Y)}(\theta)}{M_{\Lambda(Y)}(\theta)} = \Lambda(y), \qquad (7)$$

instead of (4). Observing (5), it should be clear that (we set $\lambda = 1$ in (5), since $\Lambda(Y)$ has an exponential tail of rate 1)

$$-\ln\left\{M_{\Lambda(Y)}(\theta_y^*)e^{-\theta_y^*\Lambda(y)}\right\} \sim \Lambda(y), \qquad (8)$$

since the main thing that has changed in (7) is that y has been replaced by a continuous, increasing function $\Lambda(y)$. This immediately yields asymptotic logarithmic efficiency in the simulation.

However, there are two problems with this approach. The first is that $M_{\Lambda(Y)}(\theta)$ may not be easily computable. Consider for example the case considered previously where $Y = \sum_{i=1}^m X_i$ where X_i's are i.i.d. $Weibull(\lambda, \alpha)$, $\alpha < 1$. If we use $\Lambda(y) = \lambda^\alpha y^\alpha$ as we had mentioned previously, then $\Lambda(Y) = \lambda^\alpha(\sum_{i=1}^m X_i)^\alpha$, for which it is extremely difficult to compute the mgf (since it is not decomposable as a sum of independent random variables). The second is that it may be very difficult to find $\tilde{f}_i$'s that will achieve the exponential twisting of $\Lambda(Y)$ by amount θ. So the above approach needs to be modified.

4.2 Hazard Function Transformations with Upperbound

Let $\Lambda(y)$ satisfy the conditions of Lemma 4.2. Let $V = \tilde{h}(X_1, \ldots, X_m)$ be a random variable such that:

- $\Lambda(Y) \le V$ w.p. 1.
- V is decomposable as a sum of functions of X_i's respectively, so that $M_V(\theta)$ is easy to compute.
- $\Lambda(Y)$ and V have the same asymptotic log-tail behavior, i.e., $\Lambda_V(y) \sim \Lambda_{\Lambda(Y)}(y) \sim y$.
- It is possible to find $\tilde{f}_i$'s that will achieve exponential twisting of V by amount θ.

The modified approach tries to achieve exponential twisting of V instead of $\Lambda(Y)$. In that case

$$\begin{aligned}
P(Y > y) &= E(I(\Lambda(Y) > \Lambda(y)) \\
&= \tilde{E}(I(\Lambda(Y) > \Lambda(y))M_V(\theta)e^{-\theta V}).
\end{aligned}$$

Hence the new unbiased estimator is $I(\Lambda(Y) > \Lambda(y))M_V(\theta)e^{-\theta V}$, where Y and V are obtained from the X_i's, and the X_i's are sampled using the $\tilde{f}_i$'s.

Given the conditions on V, this new estimator can be shown to be asymptotically, logarithmically efficient. In particular, we have the upperbound

$$\begin{aligned}
&\tilde{E}(I(\Lambda(Y) > \Lambda(y))l^2(X)) \\
&= \tilde{E}(I(\Lambda(Y) > \Lambda(y))M_V^2(\theta)e^{-2\theta V}) \\
&\le \tilde{E}(I(V > \Lambda(y))M_V^2(\theta)e^{-2\theta V}) \\
&\le M_V^2(\theta)e^{-2\theta\Lambda(y)}.
\end{aligned}$$

Hence

$$\tilde{E}(I(\Lambda(Y) > \Lambda(y))l^2(X)) \le M_V^2(\theta^*)e^{-2\theta_y^*\Lambda(y)},$$

where θ_y^* is the solution of

$$\frac{M_V'(\theta)}{M_V(\theta)} = \Lambda(y). \tag{9}$$

As in (8), since V also has an exponential tail of rate 1,

$$-\ln\left\{M_V(\theta_y^*)e^{-\theta_y^*}\right\} \sim \Lambda(y).$$

This leads to asymptotic logarithmic efficiency in the estimation of $P(Y > y)$.

Let us now consider the application of this approach to estimating $P(Y > y)$ where $Y = \sum_{i=1}^n X_i$ and the X_i's are i.i.d and heavy-tailed. We will now restrict ourselves to X_i belonging the class of subexponential distributions that we had mentioned earlier. Different asymptotically, logarithmically efficient changes of measure for this problem have been given in Asmussen, Binswanger and Hojgaard (2000), and Juneja and Shahabuddin (2002). We use this example just to illustrate our approach. Most subexponential distributions (i.e., excepting pathologoical cases) satisfy the property that the hazard function is eventually concave (see, e.g., Juneja and Shahabuddin 2002). For purposes of illustration we will restrict ourselves to the case where $\Lambda_{X_1}(x)$ is always concave (one can check this for the Weibull and Pareto). By the property of subexponential distributions given in (6), $\Lambda_Y(x) \sim \Lambda_{X_1}(x)$ and hence we can use $\Lambda(x) = \Lambda_{X_1}(x)$. We can then use $V = \sum_{i=1}^m \Lambda(X_i)$. Let us check whether this satisfies the properties of V that we had stated earlier.

- Due to the concavity of $\Lambda(y)$,

$$\Lambda(Y) = \Lambda(\sum_{i=1}^m X_i) \le \sum_{i=1}^m \Lambda(X_i) = V.$$

- Unlike $\Lambda(Y)$, V is decomposable into a sum of independent random variables. This together with the fact from Lemma 4.1 that $\Lambda(X_i) \equiv \Lambda_{X_i}(X_i)$ is exponentially distributed with rate 1, we get that $M_V(\theta) = 1/(1-\theta)^m$. Hence $M_V(\theta)$ is easily computable.
- Since the $\Lambda(X_i)$'s are exponentially distributed with rate 1, the V is an Erlang with rate parameter 1. Hence $\Lambda_{\Lambda(Y)}(y) \sim \Lambda_V(y)$.
- Since V is sum of light-tailed random variables, exponential twisting of each $\Lambda(X_i)$ by amount θ will yield exponential twisting of V by amount θ (as described at the end of Section 3.2).

4.3 Hazard Rate Twisting and Hazard Function Transformations

As mentioned in the Introduction, Juneja and Shahabuddin (2002) introduced the idea of "hazard rate twisting". Hazard rate twisting may be viewed as a generalization of exponential twisting to non-negative random variables with both light-tails and heavy-tails. For any non-negative random variable, say X_i, with pdf $f_i(x)$, the hazard rate twisted pdf by amount θ, $0 < \theta < 1$, is given by

$$f_{i,\theta}^h(x) = \frac{e^{\theta\Lambda_{X_i}(x)} f_i(x)}{\int_0^\infty e^{\theta\Lambda_{X_i}(s)} f_i(s)ds} = \frac{e^{\theta\Lambda_{X_i}(x)} f_i(x)}{(1-\theta)}.$$

It is easy to verify that doing hazard rate twisting by amount θ on X_i is equivalent to doing exponential twisting by the same amount on $\Lambda_{X_i}(X_i)$.

Juneja and Shahabuddin (2002) applied hazard rate twisting to the problem mentioned in the previous subsection, i.e., of estimating $P(Y > y)$ where $Y = \sum_{i=1}^m X_i$, with X_i's being i.i.d. subexponential random variables. Our contribution is to extend this approach to estimating $P(Y > y)$ for some more general functions $Y = h(X)$. At the more conceptual level, whereas Juneja and Shahabuddin (2002) thought in terms of hazard rate twisting of the X_i's, we think in terms of hazard rate twisting directly of the Y's. We then work backwards to determine the corresponding change of measure on the X_i's (that may *not* necessarily be hazard rate twisting). Viewing hazard rate twisting in terms of hazard function transformations facilitates this approach.

5 APPLICATIONS TO VALUE-AT-RISK

5.1 The Value-at-Risk Problem

We give a brief overview of the standard setting that has also been considered in Glasserman, Heidelberger, Shahabuddin (2000) and Glasserman, Heidelberger, Shahabuddin (2002). Consider a portfolio that is based on m risk factors and let $S(t) = (S_1(t), \ldots, S_m(t))$ denote their values at time t. Let $\Delta S = [S(t+\Delta t) - S(t)]^T$ (the notation A^T stands for the transpose of the matrix A) be the random change in risk factors over the future interval $(t, t+\Delta t)$. The value of the portfolio at current time t is given by $V(S(t), t)$ and the loss over the interval Δt is given by $L = V(S(t), t) - V(S(t) + \Delta S, t + \Delta t)$ (note that the only random quantity in the expression for the loss is ΔS). The risk problem is to estimate $P(L > x)$ for a given x, and the value-at-risk problem is to estimate x such that $P(L > x) = p$ for a given p, $0 < p < 1$. Usually p is of the order 0.01. As mentioned in Glasserman, Heidelberger and Shahabuddin (2000) and Glasserman, Heidelberger and Shahabuddin (2002), techniques that are efficient for estimating $P(L > x)$ for a given x, can be adapted to estimate the value-at-risk. Hence the focus in this paper, as in the previous papers, is efficient estimation of $P(L > x)$.

Usually some probability model is assumed for the ΔS, and parameters of the model are estimated from the data. The usual assumption is that ΔS is multivariate normal with mean 0 and some covariance matrix Σ (see, e.g., Glasserman, Heidelberger and Shahabuddin 2000). The multivariate normal is quite light-tailed and there is evidence from empirical finance that risk factors may have tails that are heavier than normal. This led Glasserman, Heidelberger and Shahabuddin (2002) to consider the case where the ΔS has the multivariate t distribution. The t distribution has a tail that decays polynomially, rather than according to $\frac{1}{x}e^{\frac{-x^2}{2\sigma^2}}$, as in the case of a normal.

As mentioned in the Introduction, we consider the case where ΔS has the multivariate Laplace distribution. In this case, the tails of the marginal distributions decay according to e^{-cx}, for some constant $c > 0$. Also, the multivariate Laplace random-variable may be expressed as $\sqrt{B}W$ where B is an exponentially distributed random variable with rate 1, and $W = (W_1, \ldots, W_n)$ is the multivariate normal random vector with mean 0, and covariance matrix Σ (see, e.g., Kotz, Kozubowski and Podgorski 2001). Hence one can write

$$\Delta S = \sqrt{B}W.$$

Once a probability model is assumed for the ΔS, then one can estimate $P(L > x)$ by simulation. The naive simulation method is to generate ΔS, compute $V(S(t) + \Delta S, t + \Delta t)$ and compute the loss L. Then $I(L > x)$ is an estimator of $P(L > x)$. However x may be large leading to most samples of $I(L > x)$ being 0, i.e., the typical rare-event simulation problem. Also, a portfolio may consist of many different types of instruments based on the m risk factors, making each evaluation of $V(S(t) + \Delta S, t + \Delta t)$ very time consuming. Hence one needs to use variance reduction techniques that reduce the number of samples needed for an accurate estimation.

5.2 A Quadratic Approximation for the Delta-Hedged Case

A quadratic approximation to L is given by

$$L \approx a_0 + a^T \Delta S + (\Delta S)^T A \Delta S \equiv a_0 + Q, \qquad (10)$$

where a_0 is a scalar, a is a vector, and A is a matrix. The importance sampling approach given in Glasserman, Heidelberger, Shahabuddin (2000), involves finding efficient change of measure for estimating $P(Q + a_0 > x)$, and then using the same change of measure for estimating $P(L > x)$; since $L \approx a_0 + Q$, it is likely that such an approach will be efficient for estimating the latter. Since Q is more tractable it is easier to come up with efficient changes of measure for estimating $P(Q > x - a_0)$ and proving their asymptotic logarithmic efficiency.

One quadratic approximation for the L is the delta-gamma approximation. This is simply a Taylor series expansion of the loss L in terms of ΔS. In particular, $a_0 = -\Theta\Delta t$, $a = -(\delta_i)$, and $A = - - \frac{1}{2}(\Gamma_{ij})$ in (10), where $\Theta = \frac{\partial V}{\partial t}$, $\delta_i = \frac{\partial V}{\partial S_i}$, and $\Gamma_{ij} = \frac{\partial^2 V}{\partial S_i \partial S_j}$ (all partial derivatives are evaluated at $(S(t), t)$). Many of the δ_i's and Γ_{ij}'s (especially the Γ_{ii}'s) are routinely computed for other purposes and hence are usually available prior to running the simulation.

For the purposes of this paper we consider the case where the portfolio is "delta-hedged", i.e, the proportion of investments in the various securities are selected such that the δ_i's are zero. Hence, in the delta-hedged case, $a = 0$ in (10), and thus $Q = (\Delta S)^T A(\Delta S)$. For generating ΔS, one can find C such that $CC^T = \Sigma$. Then one can generate a multivariate standard normal Z, and an exponential B with rate 1, and set $\Delta S = \sqrt{B}CZ$. In that case $Q = B(Z^T C^T ACZ)$. In order to develop importance sampling techniques for estimating $P(Q > x - a_0)$ for large x, it is advisable to find a C such that $Z^T C^T ACZ$ is a "diagonalized quadratic form". To find such a C, first find *any* $\tilde{C}$ such that $\tilde{C}\tilde{C}^T = \Sigma$ (say using Cholesky factorization). We then solve a eigenvalue problem, i.e., find an orthonormal matrix U (i.e., U such that $U^T = U^{-1}$) and a diagonal matrix Λ such that $\tilde{C}^T A\tilde{C} = U\Lambda U^T$. Let $C = \tilde{C}U$. Then we have that $CC^T = \tilde{C}UU^T\tilde{C}^T = \tilde{C}\tilde{C}^T = \Sigma$ and

$$Z^T C^T ACZ \quad = \quad Z^T U^T \tilde{C}^T A\tilde{C}UZ = Z^T \Lambda Z.$$

Hence

$$Q = B(Z^T \Lambda Z) = \sum_{i=1}^{m} B\lambda_i Z_i^2$$

where λ_i's are the diagonal elements of Λ. Without loss of generality we will assume that $\lambda_1 \geq \ldots \geq \lambda_m$.

5.3 Asymptotic Logarithmic Efficiency for Estimating $P(Q > y)$

Let $y \equiv x - a_0$. We now show how the hazard transformation approach of Section 4 can be used to determine an asymptotically, logarithmically efficient change of measure for the estimation of $P(Q > y)$. It is easy to check that each component $B\lambda_i Z_i^2$ of Q is heavy-tailed and so Q may be considered a *dependent* sum of heavy-tailed random-variables. Thus this problem is very different in essence from the ones considered in Asmussen and Binswanger (1997), Asmussen, Binswanger and Hojgaard (2000) and Juneja and Shahabuddin (2002), that considered sums of independent heavy-tailed random variables.

To keep things simple, we will also assume that $\lambda_m \geq 0$, though it is not at all necessary for our method or the asymptotic logarithmically efficient proof. It is easy to check that Q is heavy-tailed, and hence as in Section 4.1, the first step is to find a $\Lambda(y)$ such that $\Lambda_Q(y) \sim \Lambda(y)$. The following theorem gives the asymptotic order of $P(Q > y)$ for large y.

Theorem 5.1 *Suppose $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$, and $\lambda_1 > 0$. Then*

$$P(Q > y) \sim \frac{\sqrt{2\pi}}{y^{1/4}} e^{-\sqrt{\frac{2}{\lambda_1} y}}.$$

The proof uses the Laplace method (see, e.g., Bleistein and Handelsman 1975) for which we need a lemma:

Lemma 5.2 *Suppose $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$, and $\lambda_1 > 0$. Then on domain $\mathcal{D} := \{(u, v_1, \ldots, v_m) : \sum_{i=1}^m \lambda_i u v_i^2 \geq 1, u \geq 0\}$ we have*

$$\max_{\mathcal{D}}[-u - \frac{1}{2}\sum_{i=1}^m v_i^2] = -\sqrt{\frac{2}{\lambda_1}}.$$

Proof. (of Lemma 5.2) Since $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$, on domain $\mathcal{D}$, we have $\sum_{i=1}^m \lambda_1 u v_i^2 \geq \sum_{i=1}^m \lambda_i u v_i^2 \geq 1$, i.e. $u \geq \frac{1}{\lambda_1 \sum_{i=1}^m v_i^2} > 0$. Thus,

$$-u - \frac{1}{2}\sum_{i=1}^m v_i^2 \leq -\frac{1}{\lambda_1 \sum_{i=1}^m v_i^2} - \frac{1}{2}\sum_{i=1}^m v_i^2.$$

Note that

$$\begin{aligned} \frac{1}{\lambda_1 \sum_{i=1}^m v_i^2} + \frac{1}{2}\sum_{i=1}^m v_i^2 &\geq 2\sqrt{\frac{1}{\lambda_1 \sum_{i=1}^m v_i^2} \times \frac{1}{2}\sum_{i=1}^m v_i^2} \\ &= \sqrt{\frac{2}{\lambda_1}}. \end{aligned}$$

Hence, we have that $-u - \frac{1}{2}\sum_{i=1}^m v_i^2 \leq -\sqrt{\frac{2}{\lambda_1}}$. By taking $u = \sqrt{\frac{1}{2\lambda_1}}$, $v_1 = \sqrt[4]{\frac{2}{\lambda_1}}$, $v_2 = v_3 = ... = v_m = 0$, we reach the maximum, which is $-\sqrt{\frac{2}{\lambda_1}}$. ■

Proof. (of Theorem 5.1)

$$\begin{aligned} & P(\sum_{i=1}^m \lambda_i B Z_i^2 > y) \\ = \ & \frac{1}{\sqrt{2\pi})^m} \int ... \int_{\{(t,z_1,...z_m):\sum_{i=1}^m \lambda_i t z_i^2 > y, t>0\}} \\ & \qquad e^{-\frac{1}{2}\sum_{i=1}^m z_i^2 - t} dt dz_1 ... dz_m \\ & [\text{change variables: } u = t/\sqrt{y}, v_i = z_i/\sqrt[4]{y}] \\ = \ & \frac{y^{\frac{m+2}{4}}}{(\sqrt{2\pi})^m} I(y), \end{aligned}$$

where $I(y) = \int ... \int_{\mathcal{D}_0} e^{-\sqrt{x}(\frac{1}{2}\sum_{i=1}^m v_i^2 + u)} du dv_1 ... dv_m$, and the domain $\mathcal{D}_0$ is defined as $\mathcal{D}_0 := \{(u, v_1, \ldots, v_m) : \sum_{i=1}^m \lambda_i u v_i^2 > 1, u > 0\}$. $I(y)$ is an $m+1$-dimension Laplace type integral. By a result in Bleistein and Handelsman (1975), we know that

$$I(y) \sim \frac{(2\pi)^{(m+1)/2}}{y^{\frac{m+3}{4}}} e^{\sqrt{y}\phi_{max}}.$$

Here $\phi_{max} = \max_{\mathcal{D}}[-u - \frac{1}{2}\sum_{i=1}^m v_i^2]$ where $\mathcal{D}$ is the closure of $\mathcal{D}_0$. The result then follows from Lemma 5.2. ■

From Theorem 5.1 we see that one can choose $\Lambda(y) = \sqrt{2y/\lambda_1}$. Hence, if we let $\tilde{\lambda}_i = \lambda_i/(2\lambda_1)$,

$$\Lambda(Q) = 2\sqrt{B}\sqrt{\sum_{i=1}^m \tilde{\lambda}_i Z_i^2} \leq B + \sum_{i=1}^m \tilde{\lambda}_i Z_i^2,$$

where the last inequality uses the fact that $2x_1x_2 \leq x_1^2 + x_2^2$. Hence we can use $V = B + \sum_{i=1}^m \tilde{\lambda}_i Z_i^2$. Since $\tilde{\lambda}_i Z_i^2$ is are gamma random variables (with $\tilde{\lambda}_1 Z_1^2$ having the heaviest exponential tail of rate 1), it is easy to check that V has an exponential tail of rate 1. The mgf of V is trivial to compute:

$$M_V(\theta) = \frac{1}{(1-\theta)} \prod_{i=1}^m \frac{1}{\sqrt{1 - 2\tilde{\lambda}_i\theta}}. \tag{11}$$

The only thing we need to do now is to find a changes of measure on B and Z_i's that will achieve exponential twisting on the V by amount θ, $0 < \theta < 1$. As shown at the end of Section 3.2, since V is a sum of independent light-tailed random variables, doing an exponential change of measure by amount θ, $0 < \theta < 1$, on the B and each of the $\tilde{\lambda}_i Z_i^2$'s yields and exponential change of measure by amount θ on the V. The exponential change of measure by amount θ on B yields another exponential distribution with rate $(1-\theta)$. One can also easily show that if the new measure on the Z_i is $N(0, 1/\sqrt{1 - 2\tilde{\lambda}_i\theta})$, then the likelihood ratio is $e^{-\theta\tilde{\lambda}_i Z_i^2}/\sqrt{1 - 2\tilde{\lambda}_i\theta}$. Hence with this new measure on the Z_i, we achieve exponential twisting of $\tilde{\lambda}_i Z_i^2$ by amount θ.

5.4 The Importance Sampling Algorithm

To summarize, we give the steps of the importance sampling algorithm to estimate $P(L > x)$ for a given x and Δt. We assume that we are given a_0 and A from the quadratic approximation, and Σ for the W in $\Delta S = \sqrt{B}W$.

Preprocessing:

- Find current portfolio value $V(t, S(t))$.
- Find $\tilde{C}$ such that $\tilde{C}\tilde{C}^T = \Sigma$ (e.g., use Cholesky factorization). Solve the eigenvalue problem, i.e.,

find a orthonormal vector U and a diagonal matrix Λ, such that $\tilde{C}^T A \tilde{C} = U \Lambda U^T$. Let $\lambda_1, \ldots, \lambda_m$ be the diagonal elements of Λ arranged in descending order and let $\tilde{\lambda}_i = \lambda_i/(2\lambda_1)$. Set $C = \tilde{C}U$.

- Set $y = x - a_0$. Compute θ_y^* as the solution of (9) where $M_V(\theta)$ is given by (11) and $\Lambda(y) = \sqrt{2y/\lambda_1}$. Compute $M_V(\theta_y^*)$.

Generating a sample under importance sampling:

1. Generate B that is exponentially distributed with rate $(1-\theta_y^*)$. Generate independent normals $Z_1, Z_2, ..., Z_m$, with $Z_j =_d N(0, \frac{1}{1-\theta_y^* \lambda_j/\lambda_1})$.
2. Compute likelihood ratio
$$l \equiv l(B, Z_1, \ldots Z_m) = M_V(\theta_y^*) e^{-\theta_y^*(B + \sum_{i=1}^m \frac{\lambda_i}{2\lambda_1} Z_i^2)}.$$
3. Set $\Delta S = \sqrt{B} C(Z_1, \ldots, Z_m)^T$. Compute $L = V(t, S(t)) - V(S(t) + \Delta S, t + \Delta t)$.
4. Compute $I(L > x)l$.

By generating n samples of $I(L > x)l$ independently, and taking the sample mean one gets an unbiased estimator of $P(L > x)$.

Table 1: Variance Ratios (VR) of Standard Simulation to Importance Sampling in Estimating $P(L > x)$.

Portfolio 1: Delta-hedged.			
y	400	500	600
$P(Q > y)$	0.01519	0.00693	0.00326
$P(L > x)$	0.01405	0.00592	0.00257
VR	6.24	11.25	20.39
Portfolio 2: Large λ_1.			
y	1000	1200	1400
$P(Q > y)$	0.01265	0.00762	0.00486
$P(L > x)$	0.01212	0.00716	0.00445
VR	8.96	12.69	17.23
Portfolio 3: Linear λ.			
y	2500	2600	2800
$P(Q > y)$	0.01122	0.00977	0.00765
$P(L > x)$	0.01021	0.00885	0.00674
VR	12.76	14.00	17.35
Portfolio 4: Index.			
y	200	300	400
$P(Q > y)$	0.02070	0.00717	0.00282
$P(L > x)$	0.02048	0.00670	0.00255
VR	6.54	13.39	24.83

6 EXPERIMENTAL RESULTS

We test the performance of the method described above on some test portfolios consisting of calls and puts. We assume 250 trading days in a year, and a continuously compounded, risk-free, annual rate of interest of 5%. We investigate losses over 10 days (Δt = 0.04 years). Each option has a maturity of 0.5 years. We use the Black-Scholes formula to price the options. In the first three portfolios, we take the initial price of each asset to be 100; we also assume the asset prices to be uncorrelated, with each having an annual volatility of 0.3.

1. Delta-hedged: short ten at-the-money (i.e., strike price is the same as the initial price) calls and certain fixed number of puts on each of 10 underlying assets, such that the portfolio is delta-hedged.
2. Large λ_1: same as ‘Delta-hedged’ but with number of calls and puts on first asset increased by a factor of 10.
3. Linear λ: same as ‘Delta-hedged’ but with number of calls and puts on ith asset increased by a factor of i, $i = 1, ..., 10$.
4. Index: short fifty at-the-money calls and a certain fixed number of at-the-money puts on each of 10 assets, such that the port-folio is delta-hedged. The asset prices are correlated; the covariance matrix was is from the RiskMetrics website, and is given in Glasserman, Heidelberger and Shahabuddin (2000a). The initial asset prices are taken as (100, 50, 30, 100, 80, 20, 50, 200, 150, 10).

Table 1 gives importance sampling estimates of $P(Q > y)$, $P(L > x)$ (recall that $y \equiv x - a_0$) and the variance reduction factor achieved by importance sampling in the estimation of $P(L > x)$. We estimate each of these from 100,000 samples. Results from more detailed experimentation may be found in Huang and Shahabuddin (2003).

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation (U.S.A.) Grant DMI 03-00044. The observation that hazard rate twisting of a heavy-tailed random variable is equivalent to exponentially twisting the hazard function transformation of the random variable, came up in discussions with P. Glasserman (it has been mentioned in our joint proposal for the NSF Grant DMI 03-00044 submitted in October 2002) and S. Juneja.

REFERENCES

Anderson, T.W. 1984. *An Introduction to Multivariate Statistical Analysis*, Second Edition. New York, U.S.A.: Wiley.

Asmussen, S., and K. Binswanger. 1997. Simulation of ruin probabilities for subexponential claims. *ASTIN Bulletin* 27 (2): 297-318.

Asmussen, S., K. Binswanger and B. Hojgaard. 2000. Rare event simulation for heavy-tailed distributions. *Bernoulli* 6 (2): 303-322.

Bleistein, N., and R.A. Handelsman. 1975. *Asymptotic Expansion of Intergrals.* New York, U.S.A.: Holt, Rinehart and Winston.

Boots, N.K., and P. Shahabuddin. 2000. Simulating GI/GI/1 queues and insurance risk processes with subexponential distributions. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J.A. Joines, R.R. Barton, K. Kang, and P.A. Fishwick, 656-665. Piscataway, New Jersey, U.S.A.: Institute of Electrical and Electronic Engineers.

Bucklew, J.A. 1990. *Large Deviations Techniques in Decision, Simulation, and Estimation*. New York, U.S.A.: John Wiley & Sons, Inc.

Embrechts, P., C. Klüppelberg and T. Mikosch. 1997. *Modelling Extremal Events*. Heidelberg, Germany: Springer-Verlag.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000. Variance reduction techniques for estimating value-at-risk. *Management Science* 46(10): 1349-1364.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000a. Importance sampling and stratification for value-at-risk. In *Computational Finance 1999 (Proceedings of the Sixth International Conference on Computational Finance)*, ed. Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo, and A. S. Weigend, 7–24. Cambridge, Massachusetts, U.S.A.: MIT Press.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2002. Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance* 12(3): 239-269.

Heidelberger, P. 1995. Fast simulation of rare events in queueing and reliability models. *ACM Transactions on Modeling and Computer Simulation* 6: 43-85.

Heyde, C.C., and S. Kou. 2002. On the controversy over tailweight of distributions. Research Report, Dept. of IEOR, Columbia University, NY 10027, U.S.A.

Huang, Z., and P. Shahabuddin. 2003. Rare-event, heavy-tailed simulations using hazard function transformations with applications to value-at-risk. Research Report, Dept. of IEOR, Columbia University, New York, NY 10027, U.S.A.

Jorion, P. 1997. *Value at Risk*. New York, U.S.A.: McGraw Hill.

Juneja, S., and P. Shahabuddin. 2002. Simulating heavy-tailed processes using delayed hazard rate twisting. *ACM Transactions on Modeling an Computer Simulation* 12: 94-118.

Kotz, S., T. J. Kozubowski, and K. Podgorski. 2001. *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance.* Boston, U.S.A.: Birkhauser.

Wilson, T. 1999. Value at risk. In *Risk Management and Analysis, Vol. 1*, C. Alexander (ed.), 61-124. Chichester, England, U.K.: Wiley.

AUTHOR BIOGRAPHIES

ZHI HUANG is a Ph.D. candidate at the Department of Industrial Engineering and Operations Research at Columbia University. She received her B.S. in Computer Science (2000) from University of Science and Technology in China, and a M.S in Operations Research (2001) from Columbia University. Her research interests include financial modeling using analytical techniques and simulation. Her e-mail address is `<zh52@columbia.edu>`.

PERWEZ SHAHABUDDIN is an Associate Professor in the Industrial Engineering and Operations Research Department at Columbia University. From 1990 to 1995, he was a Research Staff Member at IBM T.J. Watson Research Center. He received a B.Tech. in Mechanical Eng. (1984) from the Indian Institute of Technology, Delhi, India, and a M.S. in Statistics (1987) and a Ph.D. in Operations Research (1990) from Stanford University. He is a recipient of a 1996 National Science Foundation (U.S.A) Career Award, a 1998 IBM University Partnership Award and the 1996 Oustanding Simulation Publication Award given by INFORMS College on Simulation. He serves or has served on the editorial boards of *Management Science*, *Stochastic Models*, *IIE Transactions on Operations Engineering*, *IEEE Transactions on Reliability*, and *ACM Transactions on Modeling and Computer Simulation* (Guest Editor). His e-mail address is `<perwez@ieor.columbia.edu>`.