Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

STRATIFICATION ISSUES IN ESTIMATING VALUE-AT-RISK

Paul Glasserman

Columbia University
New York, NY 10027

Philip Heidelberger

IBM T.J. Watson Research Center
Yorktown Heights, NY 10598

Perwez Shahabuddin

Columbia University
New York, NY 10027

ABSTRACT

This paper considers efficient estimation of *value-at-risk*, which is an important problem in risk management. The value-at-risk is an extreme quantile of the distribution of the loss in portfolio value during a holding period. An effective importance sampling technique is described for this problem. The importance sampling can be further improved by combining it with stratified sampling. In this setting, an effective stratification variable is the likelihood ratio itself. The paper examines issues associated with the allocation of samples to the strata, and compares the effectiveness of the combination of importance sampling and stratified sampling to that of stratified sampling alone.

1 INTRODUCTION

This paper is concerned with efficient simulation techniques for estimating *value-at-risk* (VAR), a problem of importance in risk management (see Jorion (1997) and Wilson (1999)). The VAR is a quantile of the distribution of the loss in portfolio value during a holding period. Let Δt denote the duration of the holding period and let L denote the loss in portfolio value during this holding period. Then, for a given probability p, the VAR, x_p is defined by the relationship $P\{L > x_p\} = p$. Typically p is near zero, e.g., $p = 0.01$, and Δt is either one day or two weeks. Monte Carlo simulation is often used to estimate the VAR. Such a simulation consists of first generating changes in the "risk factors," ΔS, that affect the value of portfolio. Examples of risk factors include interest rates, currency exchange rates, asset prices, etc. The portfolio is re-evaluated, using the new risk factor values at the end of the holding period, and the loss (or gain) in portfolio value is calculated. This may be quite time consuming since the portfolio may consist of a large number of financial instruments. This process is repeated multiple times so that the loss distribution can be estimated. However, for small values of p, a large number of trials may be required to accurately estimate $P\{L > x\}$ for x in the region of interest. Thus the VAR calculation may be computationally intensive.

Using variance reduction techniques to reduce the number of trials required to obtain accurate VAR estimates is therefore an attractive possibility. As VAR estimation involves a "rare event" simulation problem for small p, importance sampling (IS) is a natural candidate for variance reduction. Glasserman, Heidelberger and Shahabuddin (1999b, 1999d) (henceforth GHS) proposed and analyzed an IS procedure for the VAR problem. The approach used there is applicable when the change in risk factors, ΔS, has a multivariate normal distribution; this is often assumed in practice. First, the portfolio loss L is approximated by a quadratic function, Q, of ΔS. Then the IS uses an exponential change of measure for Q, which is asymptotically optimal for estimating $P\{L > x\}$ for large values of x, provided the approximation is exact. In an asymptotically optimal procedure, the second moment of the estimate goes to zero at twice the rate that $P\{L > x\}$ approaches zero, which is the best possible rate. With this IS, large values of Q are much more likely and thus if $L \approx Q$, the event $\{L > x\}$ is no longer a rare event under IS. The IS is further combined with stratified sampling, where the stratification variable is Q; the distribution of Q can be computed numerically. Typically, we allocate an equal number of samples to each equiprobable (under IS) stratum. While large variance reductions were obtained in test examples, this allocation policy is suboptimal. Preliminary experimental results indicated that substantial additional gains were possible with a better allocation policy. A crude form of stratification on Q, without IS and using only two strata and proportional allocation was independently proposed by Cárdenas et al. (1999).

This paper further considers issues associated with IS, stratified sampling, and their combination. In particular, we study the interaction between IS and the allocation of samples to the strata in stratified sampling. Under IS, large values of Q are more likely to be generated. In stratified sampling, with or without IS, any number of samples can be allocated to the strata, and thus the sampling distribution of Q can be arbitrarily modified. For example, without IS, an extreme modification of the allocation policy can lead to a sampling distribution of Q which is approximately equal to the sampling distribution of Q under IS. When viewed this way, it is natural to ask whether the combination of IS and stratification is significantly better than just stratification alone.

We address this issue from both a theoretical and experimental perspective. First, we prove that the variance of a sample from a small stratum under stratification (on Q) is approximately equal to the variance of a sample from that stratum under IS and stratification. This is true because of the particular way in which the IS is done; the likelihood rato becomes a function of Q and so stratifying on Q is equivalent to stratifying on the likelihood rato. This suggests that, given the same strata definition and allocation policy, the variances of the two methods should be approximately equal. In particular, the variances under optimal allocation should be approximately equal. While it is thus tempting to claim that IS is not needed, we will argue that IS provides, at a minimum, a simple framework for defining strata and an efficient initial allocation policy, one that is asymptotically optimal.

We compare the efficiencies of the two methods empirically and confirm that the best possible variance reductions of the two methods are about the same. In practice, however, estimation of the optimal allocation (using, e.g., pilot studies) may be not be practical since a large fraction of the computer budget might be exhausted simply obtaining estimates of the optimal allocation. Thus we consider several heuristic allocation policies and empirically study how close their performance is to optimality. For several sample portfolios, we show that large variance reductions are achieved using these heuristic allocation policies.

2 THE BASIC METHOD

We summarize the methodology described in GHS (1999b, 1999d). The goal is to estimate $P\{L > x\}$. We assume that the change in risk factors ΔS is an m dimensional column vector having a multivariate normal distribution with mean vector 0 and covariance matrix Σ, and that $\Sigma = \tilde{C}\tilde{C}'$ for some matrix $\tilde{C}$ (such as the Cholesky decomposition of Σ). We assume that a quadratic approximation to L is given by

$$L \approx a_0 + a'\Delta S + \Delta S' A \Delta S \equiv a_0 + Q$$

(e.g., the "delta-gamma" approximation, p. 192 of Jorion (1997)). Express Q in diagonalized form as $Q = b'Z + Z'\Lambda Z$ where Z is a vector of m independent standard normals and

1. Λ is the diagonal matrix with the eigenvalues $\{\lambda_i\}$ of $\tilde{C}'A\tilde{C}$ on the diagonal, and
2. $b' = a'C$ where $C = \tilde{C}U$ and U is the orthogonal matrix whose columns are the eigenvectors of $\tilde{C}'A\tilde{C} (= U\Lambda U')$.

Changes in risk factors can be generated by setting $\Delta S = CZ$.

Let $\lambda_1 \geq \lambda_i$ and assume $\lambda_1 > 0$. For IS, let $0 \leq \theta < \lambda_1/2$ be a "twisting parameter" and let the Z_i's be independent normals where the variance of Z_i is changed from 1 to

$$\sigma_i^2(\theta) = \frac{1}{(1 - 2\theta\lambda_i)}$$

and the mean of Z_i is changed from 0 to

$$\mu_i(\theta) = \theta b_i \sigma_i^2(\theta).$$

Then $P\{L > x\} = E_\theta[\ell I(L > x)]$ where E_θ denotes expectation under IS with twisting parameter θ and ℓ is the likelihood ratio (LR) which in this case simplifies to

$$\ell = \ell(Q) = \exp\{\psi(\theta) - \theta Q\} \tag{1}$$

where

$$\psi(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(\frac{(\theta b_i)^2}{1 - 2\theta\lambda_i} - \log(1 - 2\theta\lambda_i)\right).$$

Because of the form of the LR, this IS is equivalent to exponentially twisting the quadratic form Q. If the portfolio is exactly quadratic, i.e., if $L = a_0 + Q$, then setting $\theta = \theta_x$ where

$$\psi'(\theta_x) = x - a_0$$

results in an asymptotically optimal IS procedure. Under IS with twisting parameter θ_x, the mean of Q is $x - a_0$ and, if the portfolio is quadratic, the mean of L is x. Variations on this basic method are possible; e.g., twisting only the Z_i's associated with positive eigenvalues λ_i.

Under IS,

$$\ell I(L > x) \approx \exp\{\psi(\theta) - \theta Q\} I(Q > x - a_0)$$

which motivates combining IS with stratification on Q. With this combination, most of the variance in both the LR and the indicator is removed. For stratification, define k intervals (strata) $\mathcal{S}_j = (s_{j-1}, s_j]$ and let $p(\theta, j) = P_\theta\{Q \in \mathcal{S}_j\}$. Typically the strata are defined so that $p(\theta, j) = 1/k$; this is what was done in GHS (1999b). Numerical transform inversion techniques are used to compute the distribution of Q and then to find the $\{\mathcal{S}_j\}$ from the $\{p(\theta, j)\}$.

Let n_j be the number of samples (the allocation) that are to be drawn from stratum j, and let L_{ij} and ℓ_{ij} be the loss and LR, respectively, of the i-th sample from stratum j. Then, $P\{L > x\}$ is estimated by

$$\hat{P}_x = \sum_{j=1}^{k} p(\theta, j) \frac{1}{n_j} \sum_{i=1}^{n_j} I(L_{ij} > x)\ell_{ij}. \qquad (2)$$

When $\theta = 0$, $\ell_{ij} \equiv 1$ and (2) defines the stratified estimate without IS.

Because of the form of the LR given in (1), stratifying on Q is equivalent to stratifying on the LR. A combination of IS with stratification on the LR was also used to advantage in GHS (1999a, 1999c) in the context of pricing European-style, path-dependent options. In that setting, only the mean was changed and the stratification was done on a linear combination.

To stratify on Q, we must be able to sample Q and also sample Z given Q. A simple method for doing this is described in GHS (1999b, 1999d) and is referred to as the "bin tossing method." First, generate a vector Z of independent normals with the appropriate means and variances and then compute Q. If $Q \in \mathcal{S}_i$, then this Z has the distribution of Z given $Q \in \mathcal{S}_i$. If there are fewer than n_i samples from stratum i, then use this Z to evaluate the portfolio, otherwise discard it. We continue sampling until there are the required number of samples from each stratum.

3 EFFECTIVE ALLOCATION OF SAMPLES TO STRATA

Most of the experimental results in GHS (1999b, 1999d) allocate samples equally to each stratum. In this section we investigate heuristics for allocating samples more wisely in order to improve the variance reduction obtained.

The variance of $\hat{P}_x$ is given by

$$\text{Var}[\hat{P}_x] = \sum_{j=1}^{k} p(\theta, j)^2 v(\theta, j)^2 / n_j$$

where $v(\theta, j)^2 = \text{Var}_\theta[I(L > x)\ell | Q \in \mathcal{S}_j]$. Suppose we have a fixed budget of n samples that can be drawn, i.e., $n = n_1 + n_2 \dots n_k$. Let $f_i = n_i/n$ be the fraction allocated to stratum i. The allocation that minimizes the above variance expression can easily be derived (see, e.g., p. 300 of Fishman (1996)); the optimal fraction of samples devoted to stratum j, f_j^*, is given by

$$f_j^* = \frac{p(\theta, j)v(\theta, j)}{\sum_{i=1}^{k} p(\theta, i)v(\theta, i)}. \qquad (3)$$

Let $\hat{P}_x^*$ be the stratified sampling estimator using this optimal allocation. Then

$$\text{Var}[\hat{P}_x^*] = \frac{(\sum_{i=1}^{k} p(\theta, i)v(\theta, i))^2}{n} \qquad (4)$$

If instead of using f_i^*'s one uses f_i, then

$$\text{Var}[\hat{P}_x] = \left(\sum_{i=1}^{k} \frac{(f_i^*)^2}{f_i}\right) \text{Var}[\hat{P}_x^*]. \qquad (5)$$

Pilot runs may be done to get estimates of the $v(\theta, i)$'s in order to estimate the optimal allocation using (3). Typically, in practice, the total number of samples that one can use for the pilot run are too few to get even reasonably good estimates of these quantities. With this in mind we devise some heuristics for allocations to strata that try to do the best given this constraint. We will assume that we have a total budget of n_p (small) for the pilot runs. For convenience we will assume that n_p is some multiple of k, so that we have n_p/k samples per stratum. We describe and test three simple heuristics to illustrate the potential benefit of improved allocations. There is ample room for the development of other allocation rules.

Heuristic 1: This heuristic is based on the fact that f_i^*, as a function of i, appears to have a "normal-like" shape, centered near x. Once such f_i^* (in particular, for the second case of Portfolio (c), explained in Section 5) is given by "Optimal" line in Figure 1. This motivates the following algorithm:

1. Do a pilot run with n_p/k samples per stratum and get (very) crude estimates of $v(\theta, i)$, say $\hat{v}(\theta, i)$. Let

   $$f_i = \frac{p(\theta, j)\hat{v}(\theta, j)}{\sum_{i=1}^{k} p(\theta, i)\hat{v}(\theta, i)}.$$

   i.e., the (very) crude estimate of the f_i^*. This allocation is given by "Crude" line in Figure 1
2. Find a normal curve that best fits the f_i's. One way is the following:
   - Compute $\bar{f} = \sum_{i=1}^{k} f_i/k$ and $s(f) = \sum_{i=1}^{k}(f_i - \bar{f})^2/k$.
   - For each i, update the fractions f_i using

     $$f_i \leftarrow \frac{\exp\{-(f_i - \bar{f})^2/(2 \cdot s(f))\}}{\sum_{j=1}^{k} \exp\{-(f_j - \bar{f})^2/(2 \cdot s(f))\}}.$$

     The denominator is simply a normalization constant.

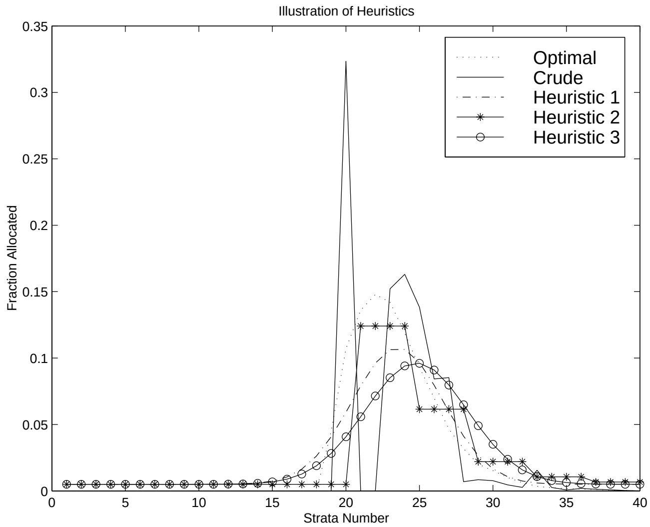

Figure 1: Illustration of sample allocations for Portfolio (c).

3. Finally we allocate 80% of the samples according to this normal curve, and the remaining 20% equally among all k strata, i.e., for each i, we set $f_i \quad 0.8 \cdot f_i + 0.2 \cdot (1/k)$. This ensures that we have at least some samples in each stratum. The final allocation is given in Figure 1.

Heuristic 2: In this case we do a "coarser" stratification (i.e., combine several strata from the original stratification scheme into one) so that we have more samples per stratum in the pilot runs. We then estimate the optimal allocation for this coarser stratification. The fraction thus allocated to the coarser strata is then equally divided among all strata that constituted that coarser stratum. Finally, as in Heuristic 1, we allocate 80% of samples according to the above and 20% equally among all k strata. The final allocation is given in Figure 1.

Heuristic 3: Same as Heuristic 2, except that before the last step, we fit a normal curve as in Heuristic 1. The final allocation is given in Figure 1.

4 STRATIFICATION WITH AND WITHOUT IMPORTANCE SAMPLING

We now compare stratification with importance sampling to plain stratification. In particular, we analyze the contribution of a single sample from a small stratum to the variance of $\hat{P}_x$. By (1), stratifying on Q is equivalent to stratifying on ℓ, so we state the result more generally in terms of using IS and stratifying on the LR to estimate $E[f(Z)]$ for some function f. We assume that the LR can be expressed as $\ell = \ell(Y)$ for some random variable Y. (Identify $Y = Q$ and $f(Z) = I(L > x)$.)

Let the stratum be the interval $(y, y + \Delta y]$ and define

$$\Delta P_\theta(y) = P_\theta\{Y \in (y, y + \Delta y]\},$$

$$m_\theta(s) = E_\theta[f(Z)|Y = s],$$

$$m_\theta^{(2)}(s) = E_\theta[f(Z)^2|Y = s],$$

$$v_\theta^2(s) = \mathrm{Var}_\theta[f(Z)|Y = s] = m_\theta^{(2)}(s) - m_\theta(s)^2,$$

and

$$V_\theta(y) = \Delta P_\theta(y)^2 \mathrm{Var}_\theta[f(Z)\ell|Y \in (y, y + \Delta y]].$$

Let $\Delta P(y)$, $m(s)$, $m^{(2)}(s)$, $v^2(s)$ and $V(y)$ denote the corresponding quantities without IS, i.e, when $\theta = 0$. We assume that $p_\theta(y)$ and $p(y)$, the density of Y with and without IS, exist.

Theorem 1 *Suppose $\ell = \ell(Y)$. If $\ell(s)$, $m_\theta(s)$ and $m_\theta^{(2)}(s)$ are finite and continuous in a neighborhood of y, and if $p(s)$ and $p_\theta(s)$ are positive, finite and continuous in a neighborhood of y, then*

$$\lim_{\Delta y \to 0} \frac{V_\theta(y)}{(\Delta y)^2} = p(y)^2 v^2(y).$$

The theorem states that, under appropriate smoothness conditions, when one combines IS with stratification on the LR, then the variance of a sample from a small stratum is, in the limit, independent of the IS distribution. Furthermore, this limiting variance is identical to that which is obtained under stratified sampling alone. The proof of the theorem, which is given in the Appendix, relies on the fact that since the LR is a function of Y, the distribution of Z given Y under IS is identical to the distribution of Z given Y under the original distribution; thus $m_\theta(s) = m(s)$, $m_\theta^{(2)}(s) = m^{(2)}(s)$, and $v_\theta^2(s) = v^2(s)$. These relationships would not generally hold if the stratification were done on some other variable, rather than on the variable defining the LR. Theorem 1 also holds in the option pricing setting described in GHS (1999a, 1999c). In that setting, the mean of a standard normal Z_i is changed from 0 to some μ_i and $\ell = c \exp\{-\sum \mu_i Z_i\}$ for some constant c. Thus the theorem holds if one stratifies on $Y = \sum \mu_i Z_i$ (but not if the stratification is done on some other linear combination).

5 EXPERIMENTAL RESULTS

We test the performance of some of the methods described above on a variety of portfolios originally used in GHS (1999b). For completeness, we list here all those portfolios.

These were intended to cover a wide range of qualitative features of portfolios. For example, Portfolios (a), (b) and (c), below have increasing quadratic terms relative to the linear terms. Other features like one dominant eigenvalue, linearly increasing eigenvalues, etc., were also incorporated among the choice of portfolios. In all but two cases they were taken to be uncorrelated; case (j), below, used a covariance matrix of 10 international equity indices downloaded from the RiskMetrics™ web site. In the uncorrelated cases, all assets have an annual volatility of 0.30 and an initial value of 100. A 10 day horizon and an interest rate of 5% were assumed in each case.

- **(a)** *0.5yr ATM*: short ten at-the-money calls and five at-the-money puts on each asset, all options having a half-year maturity;
- **(b)** *0.1yr ATM*: same as previous but with maturity of 0.10 years;
- **(c)** *Delta hedged:* same as previous but with number of puts on each asset increased to result in a linear term of zero;
- **(d)** *0.25yr OTM*: short ten calls struck at 110 and ten puts struck at 90, all expiring in 0.25 years;
- **(e)** *0.25yr ITM*: same as previous but with calls struck at 90, puts at 110;
- **(f)** *Large* λ_1: same as "Delta hedged" but with number of calls and puts on first asset increased by a factor of 10;
- **(g)** *Linear* λ: same as "Delta hedged" but with number of calls and puts on ith asset increased by a factor of i, $i = 1, \ldots, 10$;
- **(h)** *100,* $\rho = 0.0$: short ten at-the-money calls and ten at-the-money puts on 100 underlying assets, all options expiring in 0.10 years;
- **(i)** *100,* $\rho = 0.2$: same as previous but with correlations of distinct assets set to 0.20.
- **(j)** *Index*: short fifty at-the-money calls and fifty at-the-money puts on 10 underlying assets, all options expiring in 0.5 years. The covariance matrix for the asset prices is given in GHS (1999b). The initial asset prices are taken as (100, 50, 30, 100, 80, 20, 50, 200, 150, 10).

Table 1 compares the various methods and heuristics described earlier for estimating loss probabilities. Their performance is indicated by the estimated variance ratios in Columns 3 to Columns 8: "IS" is importance sampling, "ISS-Q" is stratification with importance sampling with equal number of samples per stratum, and "H1", "H2" and "H3" are stratifications with importance sampling where Heuristic 1, Heuristic 2 and Heuristic 3, are used, respectively, for allocating samples to strata. The estimated variance ratio is an estimate of the variance using standard simulation divided by the variance using a variance reduction technique. All the stratification methods use 40 equiprobable strata with respect to the importance sampling distribution of the quadratic. Also, to standardize the results, one specifies the loss threshold x as x_{std} standard deviations above the mean loss according to the quadratic approximation and varies x_{std}, i.e.,

$$x = (\sum_i \lambda_i + a_0) + x_{\text{std}} \sqrt{\sum_i b_i^2 + 2\sum_i \lambda_i^2}.$$

Results for IS and ISS-Q are taken from GHS (1999b). Variance ratios in the last four columns were estimated using the theoretical expressions in (3)-(5). The last column (OPT) gives estimates of the potential variance reduction that can be achieved if samples were allocated optimally to strata, i.e., the best that any of the heuristics could possibly do. These were done by using 1000 samples per stratum to accurately estimate the stratum variances $v(\theta, i)$ and then using (4) to estimate the optimal variance reduction. These long runs were also used to accurately estimate the f_i^*'s using (3). The variance reduction factors for H1, H2 and H3 were obtained by running 100 trials of the heuristic. For each trial, we generate 200 samples, determine the allocation (f_i's) from these samples, and determine the variance reduction achieved using (5). We then average over all 100 trials. Thus, the results in the table are estimates of how much variance reduction could be achieved on the average by applying the heuristics, with a budget of 200 samples for the pilot runs.

The results in Table 1 suggest some consistent patterns. Of the three heuristics, Heuristic 2 seems to dominate when

Table 1: Comparison of variance reduction methods. For portfolios (a), (b) and (c), the smallest values of x_{std} have $P\{L > x\} \approx 0.05$ while the largest values of x_{std} have $P\{L > x\} \approx 0.005$. For all other values of x_{std}, $P\{L > x\} \approx 0.01$. The heuristics use a total of 200 samples to determine an allocation.

		Estimated Variance Ratios					
Port.	x_{std}	IS	ISS-Q	H1	H2	H3	OPT
(a)	1.65	7.8	86.0	221	459	198	1064
	2.5	29.5	271	772	1436	740	2661
	2.8	54.1	454	1431	2602	1454	4986
(b)	1.75	7.3	30.0	103	109	78	171
	2.6	21.9	69.9	299	270	224	469
	3.3	27.1	73.0	732	593	583	1105
(c)	1.9	6.0	13.8	41	23	33	54
	2.8	17.6	30.3	90	47	81	128
	3.2	28.5	48.1	133	71	127	200
(d)	2.7	23.0	60.2	215	138	169	332
(e)	2.7	23.0	60.3	217	140	171	332
(f)	3.5	9.6	22.8	133	88	94	221
(g)	3.0	17.3	29.2	103	78	89	150
(h)	2.5	26.9	45.4	118	94	114	165
(i)	2.5	10.3	23.4	98	77	70	175
(j)	3.2	18.3	119	401	276	337	1148

there is a strong linear component (e.g., Portfolio (a)), and Heuristic 1 seems to dominate when the quadratic component is particularly strong (e.g. Portfolio (c)). The performance of Heuristic 3 seems to be dominated by the better of Heuristic 1 and Heuristic 2 in each case, but seems to be as good as the others when averaged over the wide range of portfolios considered. In general, on the average, the heuristics provide at least a 2 times improvement and up to 10 times improvement over the ISS-Q case, by using just 200 samples for the pilot runs.

Since the heuristics use so few samples to determine allocations, one may expect some degree of variability in the amount of variance reduction a heuristic obtains for a given portfolio. Conceptually, this means that in (5), the f_i's may be thought of as being random variables, and so $\mathrm{Var}[\hat{P}_x]$ is a random variable. As a general rule we found that the more the fraction of the optimal variance reduction a heuristic captures (on the average) for a portfolio, the less is the variability in this fraction; indeed if a heuristic captures 100% of the optimal variance reduction "on the average", then it does so each time, and then this variability will be zero. To illustrate this trend experimentally, define the (estimated) coefficient of variation (CV) to be the ratio of the (estimated) standard deviation of the variance reduction obtained for a given portfolio, to the (estimated) expected value of the variance reduction obtained, expressed as a percentage. Heuristic 1 had a CV of about 0.8% for the first case of Portfolio (c) to about 29% for the last case of Portfolio (a). Heuristic 2 had a CV of 2% for the first case of Portfolio (a), to 46% for the first case of Portfolio (c). The CV of Heuristic 3 seemed to be the most consistent over the wide range of portfolios considered, being less than 20% each time.

We have also investigated how the effectiveness of the allocation heuristics varies with the number of samples used for pilot runs. We find very little additional (average) variance reduction from using 400 samples in the pilot runs compared with 200 samples. In most cases, the average variance reduction using 80 samples for the pilot runs (2 per stratum) is almost as much as using 200 samples. However, we found that the CV of the variance reduction increases when going from 400 samples to 80 samples. For example, for Heuristic 1 applied to the second case of Portfolio (c), the CV was 0.7%, 1.8%, and 7.7%, for 400, 200 and 80 samples, respectively.

Overall, these numerical examples suggest that even simple heuristics can capture a significant fraction of the additional variance reduction that can be achieved through optimal allocation of samples to strata rather than equal allocation. More refined heuristics may be able to capture even more of this potential variance reduction.

Finally, we conduct some experiments with plain stratification (and no IS). Table 2 gives the variance ratios for plain stratification for Portfolio (c) where

- the stratification uses equal number of samples in each stratum and the intervals are based on the original distribution of the quadratic (S-Q)
- same as above but the intervals are based on the importance sampling distribution of the quadratic (S-Q-I)
- the stratifications use the estimated optimal allocation and the intervals are based on the original distribution (S-Q-OPT)
- same as above but the intervals are based on the importance sampling distribution (S-Q-OPT-I).

We chose Portfolio (c) for this illustration because its linear term is zero and its quadratic term has all eigenvalues equal; these features facilitate a more direct implementation of stratified sampling as compared to bin tossing (since Q then has a chi-square distribution).

Table 2: Variance ratios for plain stratification applied to Portfolio (c). All variance ratios are based on 40 strata and 1000 samples per stratum.

	Estimated Variance Ratios			
x_{std}	S-Q	S-Q-I	S-Q-OPT	S-Q-OPT-I
1.9	3.9	13.9	46	57
2.8	1.7	30.4	61	125
3.2	1.3	47.2	50	206

Note that S-Q and S-Q-OPT give inferior results to their counterparts S-Q-I and S-Q-I-OPT; using equiprobable intervals based on importance sampling creates a better sampling frequency near the region which matters, i.e., close to x. Also note how close S-Q-I and ISS-Q-I-OPT are to ISS-Q and OPT, respectively (the results of which are in Table 1), consistent with Theorem 1.

Even though we experimented with trials of the heuristics for determining effective allocations, we did not actually implement stratifications with the suggested allocations. The latter may pose some difficulties. For example, though the bin tossing method of stratification works well for IS and stratification with equal allocation, one may expect it to have considerably more wastage of the generated normals with the skewed allocation that one gets from the heuristics. This is also the case when doing plain stratification with equal allocation, but with intervals generated using importance sampling. Currently we are investigating this overhead and also developing an acceptance-rejection algorithm for generating the stratified samples while keeping the rejections to a minimum.

6 CONCLUSIONS

We have tested several heuristics for allocating samples to strata when doing importance sampling with stratification. We showed that simple heuristics can capture a significant fraction of the variance reduction that may be achieved if one allocated samples optimally.

We also compared the combination of stratification and importance sampling to plain stratification, where both schemes use the same stratification intervals and the same number of samples for each stratum. We proved a result that in the limit (i.e., as the size of a stratification interval approaches zero) the stratum variances using the two approaches converge to the same value. Experimental results confirm the above by showing that the two methods of stratification give almost the same variance reduction when either using equal number of samples per stratum or when using optimal allocation of samples to the strata. In fact, the optimal allocation of samples in the two cases is about the same. However, in practice, the importance sampling approach gives a natural method to generate effective strata.

ACKNOWLEDGMENTS

This work was partially supported by NSF NYI grant DMI-9457189 and Career Award grant DMI-96-25297.

APPENDIX: PROOF

First we show that $\ell(y) = p(y)/p_\theta(y)$. Write

$$P\{Y \in [y, y+\Delta y]\} = E_\theta[\ell I(Y \in [y, y+\Delta y])\}. \quad \text{(A-1)}$$

Since $\ell = \ell(Y)$, dividing (A-1) by Δy, applying the mean value theorem, and letting $\Delta y \to 0$ establishes this fact. Now write $V_\theta(y) = \Delta P_\theta(y)^2[I_2(y) - I_1(y)]$ where

$$I_2(y) = \int_{s=y}^{y+\Delta y} E_\theta[\ell^2 f(Z)^2 | Y = s] \frac{p_\theta(s)}{\Delta P_\theta(y)} ds$$

and

$$I_1(y) = \left(\int_{s=y}^{y+\Delta y} E_\theta[\ell f(Z) | Y = s] \frac{p_\theta(s)}{\Delta P_\theta(y)} ds \right)^2.$$

Since if $Y = s$, $\ell = p(s)/p_\theta(s)$, by the mean value theorem there exists a $y' \in (y, y+\Delta y]$ such that

$$I_2(y) = \left(\frac{p(y')}{p_\theta(y')} \right)^2 m_\theta^{(2)}(y') p_\theta(y') \frac{\Delta y}{\Delta P_\theta(y)}.$$

Thus

$$\lim_{\Delta y \to 0} I_2(y) = \left(\frac{p(y)}{p_\theta(y)} \right)^2 m_\theta^{(2)}(y)$$

and

$$\lim_{\Delta y \to 0} \frac{\Delta P_\theta(y)^2}{(\Delta y)^2} I_2(y) = p(y)^2 m_\theta^{(2)}(y).$$

A similar argument shows that

$$\lim_{\Delta y \to 0} \frac{\Delta P_\theta(y)^2}{(\Delta y)^2} I_1(y) = p(y)^2 m_\theta(y)^2$$

and thus

$$\lim_{\Delta y \to 0} \frac{V_\theta(y)}{(\Delta y)^2} = p(y)^2 v_\theta^2(y).$$

The result follows if $v_\theta^2(y) = v^2(y)$. For a set A,

$$\begin{aligned} P\{Z \in A | Y = y\} &= \frac{E[I(Z \in A); Y = y]}{p(y)} \\ &= \frac{E_\theta[\ell(Y) I(Z \in A); Y = y)]}{p(y)} \\ &= \frac{E_\theta[\frac{p(y)}{p_\theta(y)} I(Z \in A); Y = y]}{p(y)} \\ &= \frac{E_\theta[I(Z \in A); Y = y]}{p_\theta(y)} \\ &= P_\theta\{Z \in A | Y = y\}. \end{aligned}$$

Thus the conditional distribution of Z given Y is the same under IS as it is under the original distribution and therefore $v_\theta^2(y) = v^2(y)$.

REFERENCES

Cárdenas, J., E. Fruchard, J.-F. Picron, C. Reyes, K. Walters, and W. Yang. 1999. Monte Carlo within a day. *Risk* 12 (February): 55 – 59.

Fishman, G.S. 1996. *Monte Carlo Concepts, Algorithms, and Applications*. New York: Springer.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 1999a. Asymptotically optimal importance sampling and stratification for pricing path- dependent options. *Mathematical Finance* 9: 117–152.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 1999b. 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, Cambridge, MA: MIT Press, (to appear).

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 1999c. Importance sampling in the Heath-Jarrow-Morton framework. *Journal of Derivatives* (to appear).

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 1999d. Variance reduction techniques for estimating Value-at-risk. IBM Research Report (to appear), Yorktown Heights, NY.

Jorion, P. 1997. *Value at Risk: The New Benchmark for Controlling Derivatives Risk*. New York: McGraw-Hill.

Wilson. T. 1999. Value at risk. In *Risk Management and Analysis, Vol 1*, ed. C. Alexander, 61 – 124. Wiley, Chichester, England.

AUTHOR BIOGRAPHIES

PAUL GLASSERMAN is a Professor in the Management Science division of the Columbia University Graduate School of Business. Prior to joining the Columbia faculty he was a Member of Technical Staff in the Operations Research department of AT&T Bell Laboratories in Holmdel, NJ. He holds a Ph.D. from Harvard University and an A.B. from Princeton University.

PHILIP HEIDELBERGER received a B.A. in mathematics from Oberlin College in 1974 and a Ph.D. in Operations Research from Stanford University in 1978. He has been a Research Staff Member at the IBM T.J. Watson Research Center since 1978.

PERWEZ SHAHABUDDIN is an Associate Professor in the Industrial Eng. and Operations Research Dept. at Columbia University. From 1990 to 1995, he was a Research Staff Member at the IBM T.J. Watson Research Center. He received his B.Tech in Mechanical Eng. from I.I.T./Delhi, in 1984, and then a M.S. in Statistics and a Ph.D. in Operations Research from Stanford University in 1987 and 1990, respectively.