

STRATIFIED SAMPLING FOR RISK MANAGEMENT

STEFAN JASCHKE AND PETER MATHÉ

ABSTRACT. The authors discuss the approximation of Value at Risk (VaR) and other quantities relevant to risk management. One of the core problems in this context is the approximation of the distribution of quadratic forms of Gaussian vectors. It appears as an intermediate problem in the variance reduction techniques proposed by Glasserman et. al. as well as in so-called Δ - Γ -normal approaches.

The purpose of this paper is to show that sampling methods are faster than Fourier inversion for a range of practical problems. In fact, the asymptotic cost to achieve a given accuracy is $\mathcal{O}(d)$ for Monte-Carlo and $\mathcal{O}(d^3)$ for Fourier inversion in a specific setting. The theoretical results are supported by a case study based on real-life problems, showing that sampling methods are faster than Fourier inversion for all but the smallest problems. Thus the use of randomized methods is recommended and the issue of variance reduction becomes important. Stratification methods – especially randomized orthogonal arrays – turn out to lead to the most effective methods in terms of accuracy per computational cost.

1. PROBLEM FORMULATION

Given a portfolio, the typical task of a bank's risk controller is to model the change in the portfolio's value over a certain time horizon h in the future. If S_t^i denotes the value of asset i at time t and w_t^i the number of units of assets of type i in the portfolio, then

$$V(w_t, S_t) = \sum_{i=1}^M w_t^i S_t^i$$

is the portfolio's value at time t . The object of interest is the hypothetical profit or loss (P&L) from holding the frozen portfolio (w_t) over the time interval $[t, t + h]$:

$$(1) \quad V(w_t, S_{t+h}) - V(w_t, S_t),$$

Date: October 18, 2004.

2000 Mathematics Subject Classification. 65C05.

Key words and phrases. randomized orthogonal arrays, risk management.

where we assume for simplicity here that the portfolio does not produce or require any cash flows over the forecast horizon.

Specifically, we will consider the problem of computing the 1%-quantile of the distribution of (1), i.e., minus the 99% Value-at-Risk (VaR) of the portfolio, conditional on all the information at time t . This is the mathematical core of “internal models” in the sense of the 1996 Amendment of the Basel accord. One of the requirements for using internal models for the computation of the regulatory capital is that the model is actually used in the risk management of the bank. This means first that the risk manager has to model and approximate not only the 1%-quantile but the whole distribution of (1). Second, the distributions and their 1%-quantiles have to be computed for the whole trading book of a bank as well as for a tree of subportfolios.

For statistical reasons, the asset values S_{t+h}^i are expressed as functions of a vector of risk factors X_{t+h} , as $S_{t+h}^i = S^i(X_{t+h})$, in such a way that the dynamics of the vector of risk factors X_{t+h} can more easily be modeled than the asset prices themselves. Think of the prices of (risk-free) bonds denominated in EUR, each of which can be viewed as the sum of discounted values of the individual payments. The vector X_{t+h} would be a representation of the term structure of (risk-free) interest rates at time $t+h$ and S^i the non-linear function that transforms the interest rates for different maturities into a discounted value for this specific bond. This reduces the dimensionality from potentially several hundred thousand asset types (M) to a few hundred or a few thousand risk factors (d).¹

At the subportfolio (trading desk) level, both the portfolio weights w_t^i and the “exact” functions S^i are in principle available to the risk controller. Due to the diversity of front-office systems in some larger investment banks it is, however, usually not practical to work with the “exact” functions S^i in the centralized IT systems that compute VaR for the whole portfolio. Rather, the functions S^i are approximated by quadratic functions or scenario matrices or a hybrid of both in practice.

For the purpose of this paper we will focus on the problem of VaR-computation for the higher levels of the portfolio tree and assume that only quadratic approximations to the functions S^i are known in the centralized IT systems. Furthermore, the conditional distribution of risk factor increments $X_{t+h} - X_t$ at time t is often modeled as Gaussian $\sim N(0, \Sigma)$. Then, the hypothetical P&L (1) can be written as the

¹In the real-world applications we know of the number of risk factors d ranges from about 100 to 10000.

quadratic form

$$(2) \quad Q(X) = \theta + \langle \Delta, X \rangle + \frac{1}{2} \langle X, \Gamma X \rangle,$$

with the slight abuse of notation $X := X_{t+h} - X_t$. The d -vector Δ denotes the w_t -weighted sum of all linear terms and the $d \times d$ -matrix Γ denotes the weighted sum of all quadratic terms in the approximations of the mapping functions S^i . θ is a constant and represents that part of the change in portfolio value that is predictable at time t . Γ is always sparse, since each S^i depends only on a few risk factors². The number of nonzeros of Γ will be denoted by $c_{nz}d$.

The conditional covariance matrix Σ is usually estimated from daily observations of the risk factor increments $X_{t+1} - X_t$ and thus updated only once per day. The estimate may be a nontrivial function of the observations in GARCH-like models [Gouriéroux, 1997] or just the empirical covariance matrix $\Sigma = x^\top x$, where x denotes an $l \times d$ -matrix of observations of risk factor increments.³ The number of observations l is usually between 250 (1 year) and 1250 (5 years). If Σ is not the empirical covariance or l is larger than d , then a decomposition $\Sigma = BB^\top$ can be computed once overnight. If d is very large ($d > 1000$), then the trailing eigenvalues can be ignored for practical purposes and the $d \times l$ -matrix B can be chosen such that l is in the order of a few hundred. Thus, we can assume w.l.o.g. that $l \leq d$.

Since the portfolio composition changes frequently throughout the day, the computation of the 1%-quantile of the quadratic form (2) has to be computed “online” for frequently changing (θ, Δ, Γ) . Σ and its decomposition $\Sigma = BB^\top$ can be assumed known and fixed for the purpose of the online computations. The required response time ranges from a few seconds at the trading desk level to about one hour at the highest level. The required accuracy is about one decimal digit, i.e., 10% relative error.⁴

The outline of the paper is as follows. Based on the operation counts of *the* standard algorithm for the approximation of the distribution (2), namely Fourier inversion, we show that Monte Carlo methods are

²Some banks use only diagonal matrices Γ .

³Since the requirement of an “effective observation period” of at least one year in the Amendment to the Basel Accord (B.4 (d)) is interpreted in a restrictive way by some regulatory authorities, many banks use the empirical covariance matrix as an estimator for Σ for the computation of the regulatory capital.

⁴Due to the numerous sources of error in the computation, especially noise from the estimation of Σ , a smaller relative error than about 5% cannot be expected in the practical context. See [Jaschke, 2002a, section 6], for example.

preferable for moderate required accuracy and large effective dimension l , since the initial tridiagonalization preceding the Fourier inversion needs $\mathcal{O}(l^2d)$ operations, while Monte Carlo needs $\mathcal{O}(d)$ operations for a given accuracy (section 2).

If sampling is the preferable method for certain classes of problems, then variance reduction becomes an important issue. Most variance reduction techniques for the Value-at-Risk problem as described in the literature – actually *all* VaR-approximation methods described in [Glasserman, 2003, chapter 9] – are based on the assumption that the cdf of the quadratic form (2) can be computed comparatively fast by a deterministic algorithm. First, this assumption is inappropriate, if both the numbers d of risk factors and l of past observations are large. Second, the most effective variance reduction techniques – importance sampling at the saddle point of the exponential family generated by the quadratic form (2) plus importance sampling along that quadratic form as described in [Glasserman, 2003, section 9.2] – reduce the variance primarily near the required quantile, not necessarily for the approximation of the whole distribution.

The focus of this paper is on such variance reduction techniques that reduce the variance at all quantiles and do not require the initial tridiagonalization. Section 3 describes the mathematical framework for such *stratification* techniques. To rigorously compare different sampling methods we introduce quantities, which measure the performance of specific methods against plain Monte Carlo, both locally at a given quantile or globally. Furthermore, section 3 also provides details on the construction and theoretical properties of *randomized orthogonal arrays* of strength 1 or 2 (subsection 3.1). Then we turn to digital nets and indicate why this is also a way of stratifying samples. For *randomized digital nets* the results for randomized orthogonal arrays apply, and a profound theoretical basis for their use is given (subsection 3.2). The rest of section 3 shows how stratification can be combined with other means of variance reduction.

The case study (section 4) compares the effectiveness of a number of variance reduction techniques to both standard Monte Carlo and Fourier inversion on a real-world problem. In a first step, the effectiveness is measured in terms of the Anderson-Darling distance of the approximated cdf from the true cdf (section 4.1). This allows to compare the main methods of interest, randomized orthogonal arrays and randomized digital nets, to Fourier inversion.

In a second step, effectiveness of variance reduction is measured in terms of the relative root mean squared error of the approximation

of the 1%-quantile (section 4.2). This allows to compare the various methods with the method of [Glasserman, 2003, section 9.2].

We shall indicate that these methods are generally applicable in risk management; they may be used in the more general “full valuation Monte-Carlo” context of [Glasserman, 2003, chapter 9] as well.

2. OPERATION COUNTS OF STANDARD ALGORITHMS

What makes quadratic forms of Gaussian vectors especially tractable is that the Fourier transform of their probability densities is known analytically and thus also their cumulants and moments.

Several methods have been proposed to compute a quantile of the distribution defined by the model (2), among them Monte-Carlo simulation [Pritsker, 1996], moment-based methods, and methods based on the evaluation of either the cumulant generating function $\kappa(s) = \log Ee^{sQ}$ or the Fourier transform $\phi(t) = e^{\kappa(it)}$. Among the moment-based methods are Johnson transformations [Zangari, 1996a, Longerstaeey, 1996], Cornish-Fisher expansions [Zangari, 1996b, Fallon, 1996, Jaschke, 2002a], the Solomon-Stephens approximation [Britton-Jones and Schaefer, 1999], and moment-based approximations motivated by the theory of estimating functions [Li, 1999]. Both saddle-point approximations [Kuonen, 1999, Feuerverger and Wong, 2000, Studer, 2001] and Fourier-inversion [Rouvinez, 1997, Albanese et al., 2000, Jaschke, 2002b] require the evaluation of one of the function transforms.

Comparisons of methods include [Pichler and Selitsch, 1999, Mina and Ulmer, 1999, Feuerverger and Wong, 2000, Volmar, 2002]. In this subsection, we will compare the floating point operation counts of the two standard methods that can achieve any desired accuracy: Fourier inversion and Monte Carlo simulation.

The characteristic function of the quadratic form (2) of a Gaussian vector with mean zero and covariance matrix $\Sigma = BB^\top$ is

$$(3) \quad \phi(t) = e^{i\theta t} \det(I - itB^\top \Gamma B)^{-1/2} \\ \times \exp\left\{-\frac{1}{2}t^2 \Delta^\top B(I - itB^\top \Gamma B)^{-1} B^\top \Delta\right\},$$

see [Jaschke, 2002a], for example. In this form, each evaluation of the characteristic function would require $\mathcal{O}(l^3)$ operations, needed for the decomposition of the $l \times l$ -matrix $B^\top \Gamma B$. Thus, a transformation $B^\top \Gamma B = QRQ^\top$ is performed once, where Q is orthogonal and R is either tridiagonal or diagonal. Then, (3) transforms to

$$(4) \quad \phi(t) = e^{i\theta t} \det(I - itR)^{-1/2} \exp\left\{-\frac{1}{2}t^2 \Delta^\top BQ(I - itR)^{-1} Q^\top B^\top \Delta\right\}.$$

TABLE 1. Floating point operation counts. The counts for the LAPACK routines are taken from the LAPACK 3.0 file `dop1a.f`. Omitted are the linear terms, i.e., all terms with only one of the “large” factors l , d , and n .

computational step	LAPACK function	FLOP count
Fourier inversion		
1. compute the matrix $B^\top \Gamma B$	–	$2c_{nz}ld$
2. tridiagonalize the result of 1.	DGEMM(1,d,1)	$2l^2d$
3. compute $B^\top \Delta$	DSYTRD(1)	$\frac{4}{3}l^3 + \frac{7}{2}l^2$
4. multiply the result of 3. with the representation of Q generated by DSYTRD in step 2.	DGEMV(1,d)	$2ld$
	DORMTR(1,1)	$2l^2$
setup costs		$\mathcal{O}(l^2d)$
5. factorize $I - itR$	ZPTTRF(1)	$14nl$
6. backsolve, using 5.	ZPTTRS(1)	$22nl$
7. compute $\det(I - itR)$, using 5.	ZDOT(1)	$8nl$
8. another scalar product	ZDOT(1)	$8nl$
n evaluations of the Fourier transform		$\mathcal{O}(nl)$
9. FFT size n	–	$5n \log_2 n$
FFT		$\mathcal{O}(n \log(n))$
plain Monte-Carlo		
1. $\langle X, \Gamma X \rangle$ for n samples	–	$3c_{nz}nd$
2. $\langle \Delta, X \rangle$ for n samples	DGEMV(n,d)	$2nd$

The computations of $(I - itR)^{-1}Q^\top B^\top \Delta$ and $\det(I - itR)$ are $\mathcal{O}(l)$ -operations for both tridiagonal and diagonal R , see Anderson et al. [1999]. Table 1 shows the floating point operation counts of standard Monte-Carlo simulation and Fourier inversion (using the reduction to tridiagonal).

Given the operation counts, we can distinguish three asymptotics:

- (1) $n \rightarrow \infty$. The number of function evaluations in the Fourier inversion needed to achieve a worst-case error ϵ is

$$n = \mathcal{O}(1/\epsilon \log(1/\epsilon)),$$

see Jaschke [2002b]. The number of MC-samples for a required mean squared error ϵ is

$$n = \mathcal{O}(\epsilon^{-2}).$$

Hence, if high accuracy is needed, then Fourier inversion is faster than Monte-Carlo, since the setup costs vanish in the limit.

- (2) l is a constant fraction of d and $d \rightarrow \infty$. Then the asymptotic cost of MC is $\mathcal{O}(d)$, which compares very favourably to the asymptotic cost of the Fourier inversion, $\mathcal{O}(d^3)$.
- (3) $d \rightarrow \infty$, with everything else constant. Then the asymptotic costs of both Fourier inversion and Monte-Carlo are $\mathcal{O}(d)$.

In the practical context we described in the introduction the required accuracy is moderate and both l and d are relatively large. Thus it is plausible that Monte-Carlo methods may be faster than Fourier inversion for the problem at hand. This will be exhibited in the case study presented in section 4.

3. THE MATHEMATICAL FRAMEWORK FOR SAMPLING ALGORITHMS

Suppose that, given a function $f: [0, 1]^d \rightarrow \mathbb{R}$, we want to approximate the cumulative distribution function (cdf), defined as $F(x) = P(f(u) < x)$, $x \in \mathbb{R}$, by sampling, i.e., we want to find samples (X_1, \dots, X_n) such that the empirical cdf $F_n(x) := F(X_1, \dots, X_n, x) = \frac{\#\{j, f(X_j) < x\}}{n}$, $x \in \mathbb{R}$, is close to F . Because we will discuss different sampling strategies it will be convenient to emphasize the underlying strategy, say ‘‘meth’’ by writing $F_n(\text{meth}, x) := F(X_1, \dots, X_n, x)$.

For any $x \in \mathbb{R}$ the computation of $F(x)$ can be regarded as an integration problem because $F_n(x)$ and $F(x)$ both can be written as the integral of the composite function $f_x(u) := \chi_{(-\infty, x)}(f(u))$, $u \in [0, 1]^d$.

In general, given a function $f: [0, 1]^d \rightarrow \mathbb{R}$ and a set $\{X_1, \dots, X_n\}$ of sample points, the estimate of the integrand shall be given through the sample mean

$$\vartheta_n(f) := \frac{1}{n} \sum_{j=1}^n f(X_j).$$

The mean squared error of $\vartheta_n(f)$ is given by

$$(5) \quad e^2(f, \vartheta_n(f)) := \mathbf{E} \left| \int f(u) du - \vartheta_n(f) \right|^2.$$

If the sampling is such that each marginal X_j is distributed uniformly on $[0, 1]^d$, then $\vartheta_n(f)$ is unbiased and $e^2(f, \vartheta_n(f)) = \text{Var}(\vartheta_n(f))$. Thus for the empirical distribution function, when $F_n(x) = \vartheta_n(f_x)$ we obtain $e^2(f_x, \vartheta_n(f)) = \text{Var}(F_n(\text{meth}, x))$. This is a family of functions depending on the sample size n , the location x and on the method according to which the sample is generated.

When discussing different sampling methods, it is natural to consider plain Monte Carlo (MC) as a benchmark and it is easy to verify that

$$\text{Var}(F_n(\text{MC}, x)) = \frac{F(x)(1 - F(x))}{n}.$$

The gain of the chosen sampling method “meth” over MC can be measured by the quotient

$$(6) \quad \kappa_n(\text{meth}, x) := \left(\frac{\text{Var}(F_n(\text{meth}, x))}{\text{Var}(F_n(\text{MC}, x))} \right)^{1/2} = \left(n \frac{\text{Var}(F_n(\text{meth}, x))}{F(x)(1 - F(x))} \right)^{1/2}.$$

We aim to show that stratification is a means to obtain $\kappa_n(\text{meth}, x) < 1$.

On the one hand risk management requires the approximation of specific quantiles (1% and 5%), as the risk measure Value at Risk (VaR) has special relevance. Given $0 < p < 1$ the error of approximating a quantile $q := F^{-1}(p)$ by means of the empirical quantile

$$\hat{q}_n := \hat{q}_n(\text{meth}) = \min \{x, F_n(\text{meth}, x) > p\}.$$

will be measured in root mean squared sense

$$e(q, \hat{q}_n(\text{meth})) := (\mathbf{E}(q - \hat{q}_n(\text{meth}))^2)^{1/2}.$$

It can be shown that

$$\lim_{n \rightarrow \infty} n e^2(q, \hat{q}_n(\text{meth})) = \frac{\text{Var}(F_n(\text{meth}, q))}{(F'(q))^2},$$

provided $F'(q)$ exists and is non-zero in a neighborhood of q , cf. [Glasserman, 2003, equation 9.7]. In this case the quantity $\kappa_n(\text{meth}, q)$ controls the quality of the chosen sampling method “meth” against MC. For quantile estimation *importance sampling* can improve upon MC, we refer the reader to [Glasserman, 2003, Chapt. 9]. As can be seen in section 3.4 on the basis of some example, stratification by inducing correlation is useful and leads to improved performance.

On the other hand risk management requires the approximation of the distribution F as a whole, as indicated in the introduction. For this global inference the following approaches are immediate.

(1) *Worst performance:*

$$\kappa_n^{\text{sup}}(\text{meth}) := \sup_{x \in \mathbb{R}} \kappa_n(\text{meth}, x),$$

or

(2) *average performance:*

$$\kappa_n^{\text{rms}}(\text{meth}) := \left(\int \kappa_n^2(\text{meth}, x) dF(x) \right)^{1/2}.$$

It is worthwhile to discuss the latter quantity in more detail. Reversing the order of integration in $\kappa_n^{rms}(\text{meth})^2$ we obtain an explicit description as

$$\kappa_n^{rms}(\text{meth})^2 = \mathbf{E} n \int \frac{|F_n(\text{meth}, x) - F(x)|^2}{F(x)(1 - F(x))} dF(x) = \mathbf{E} n D_{AD}^2(F_n, F),$$

where $nD_{AD}^2(F_n, F)$ is usually called the *Anderson-Darling statistic*, which measures the closeness of the ecdf F_n to the cdf F , weighted in the tails, see [Marsaglia and Marsaglia, 2004]. Analogously, $D_{AD}(F_n, F)$ can be called the *Anderson-Darling distance* between the two distributions F_n and F . Plainly, $\kappa_n^{rms}(\text{MC}) = 1$. Thus, if $\kappa_n^{rms}(\text{meth}) < 1$, then at least asymptotically “meth” is superior to MC in mean square sense. By theorem 1, for the methods under consideration the Anderson-Darling distance decreases to 0 at least at a rate $1/\sqrt{n}$.

Recall that our goal of stratification is to reduce $\text{Var}(F_n(\text{meth}, x))$ by inducing correlation. This can be achieved by *randomized orthogonal arrays* of different strength, see 3.1, and also using randomized (scrambled) digital nets. Therefore we shall outline the corresponding constructions and provide theoretical results, in particular results which have implications towards the quantities $\kappa_n^{sup}(\text{meth})$ and $\kappa_n^{rms}(\text{meth})$.

3.1. Randomized orthogonal arrays. The construction starts with the choice of some (large enough) N and corresponding subdivision of $[0, 1]^d$ into N^d equally sized sub-cubes. We then label each cube with the coordinate of its lower left corner times N , which gives a d -tuple $(n_1, \dots, n_d) \in \{0, \dots, N-1\}^d$. The goal of stratification is to choose samples from within a small portion of the sub-cubes still being representative. This will be done by randomized orthogonal arrays.

First we briefly recall the definition of deterministic orthogonal arrays and how they relate to random samples.

An orthogonal array $\text{OA} = \text{OA}(N^t, d, N, t)$ of strength t is a $d \times N^t$ matrix with entries from $\{0, \dots, N-1\}$ such that the N^t rows of each $t \times N^t$ sub-matrix of OA contains *all* N^t pairs from $\{0, \dots, N-1\}^t$. For orthogonal arrays of strength 1, better known as Latin hypercubes, this is related to permutation matrices. For arrays of higher strength the construction may be difficult. We will provide one construction of such initial arrays of strength 2, below in section 4.1. Notice, that any array of strength 2 has strength 1, also. Notice furthermore that such samples are not available for every sample size, because they are based on the existence of such arrays in the given dimension and with given strength, see Hedayat et al. [1999] for more information on this.

A random sample within the integration domain $[0, 1]^d$ is derived from this initial array OA as follows.

Having an orthogonal array OA of strength t its columns may be identified with the labels of certain sub-cubes, they thus select N^t sub-cubes out of the overall N^d . The array OA being of strength t means, that for each choice of t -dimensional margins (d_1, \dots, d_t) , each t -dimensional sub-cube is chosen exactly once.

Randomization is now carried out in two steps. First we notice that the strength of the orthogonal array retains if we relabel the symbols from within each row, i.e., we may and do apply random permutations of the symbols to each row to randomly select N^t sub-cubes by the above identification.

Let us exemplify this at the orthogonal array

$$(7) \quad \text{OA} := \text{OA}(9, 3, 3, 2) = \begin{pmatrix} 2 & 0 & 1 & 0 & 1 & 2 & 1 & 2 & 0 \\ 0 & 1 & 2 & 2 & 0 & 1 & 1 & 2 & 0 \\ 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 \end{pmatrix},$$

which can easily be seen to be of strength 2. Applying the permutations (01), (12) and the identical one to the three rows yields

$$\text{ROA} = \begin{pmatrix} 2 & 1 & 0 & 1 & 0 & 2 & 0 & 2 & 1 \\ 0 & 2 & 1 & 1 & 0 & 2 & 2 & 1 & 0 \\ 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 \end{pmatrix},$$

resulting in a random collection of 9 sub-cubes of the 27 sub-cubes.

In a second step we draw uniformly distributed points, one from within each of the chosen sub-cubes. Formally, with $i \in \{1, \dots, d\}$, $j \in \{1, \dots, N^t\}$,

$$X_j^i := \frac{\text{ROA}_{i,j} + u_{ij}}{N}, \quad u_{ij} \text{ uniform on } [0, 1],$$

and obtain a set $\{X_1, \dots, X_n\}$ of $n = N^t$ sample points. Below this method of stratification will be denoted by “t-oa”, where t denotes the chosen strength.

Notice, that N always denotes the level of subdivision in the cube and is related to the sample size n through N varying with the strength t of the array.

Remark 1. Sampling according to randomized orthogonal arrays of strength 1 is called *Latin Hypercube sampling*.

3.2. Digital nets. One may also look at $|F(x) - F_n(x)|$ from a different perspective. Indeed we can rewrite this as

$$\left| \text{Vol}(U_f(x)) - \frac{\#\{j, X_j \in U_f(x)\}}{n} \right|,$$

for level sets $U_f(x) := \{u, f(u) < x\} \subset [0, 1]^d$. Thus, given f let us consider the following set system

$$\mathcal{C} := \mathcal{C}_f := \{U_f(x), \quad x \in \mathbb{R}\}.$$

Then the goal is to find *deterministic* point sets $\{x_1, \dots, x_n\}$ for which

$$(8) \quad D(x_1, \dots, x_n; \mathcal{C}) := \sup_{U \in \mathcal{C}} \left| \text{Vol}(U) - \frac{\#\{j, x_j \in U\}}{n} \right| \rightarrow \text{MIN!}$$

The quantity $D(x_1, \dots, x_n; \mathcal{C})$ is usually called *discrepancy* with respect to the system \mathcal{C} of sets. ‘‘Good’’ point sets $\{x_1, \dots, x_n\}$ are those which realize (up to logarithmic factors) the best possible

$$D_n^{\mathcal{C}} := \min \{D(x_1, \dots, x_n; \mathcal{C}), \quad x_1, \dots, x_n \in [0, 1]^d\}.$$

For specific systems good point sets have been studied extensively. We only mention the system

$$\mathcal{C}^* := \left\{ \prod_{j=1}^d [0, u_j), \quad 0 < u_j < 1, \quad j = 1, \dots, d \right\}$$

of rectangles anchored at 0, which leads to the notion of **-discrepancy*, which for a given (deterministic) point set $\{x_1, \dots, x_n\}$ can explicitly be given through

$$D^*(x_1, \dots, x_n) := \sup_{u \in [0, 1]^d} \left| \prod_{j=1}^d u_j - \frac{\#\{j, x_j^l < u^l, \quad l = 1, \dots, d\}}{n} \right|.$$

Good point sets for this specific system are called *low discrepancy point sets*. For several specific systems \mathcal{C} the asymptotic behavior of the respective minimal discrepancy $D_n^{\mathcal{C}}$ is known, in particular the bound

$$(9) \quad D_n^* \leq C \frac{\log^{d-1} n}{n},$$

for some constant C shows, that deterministic point sets of low discrepancy outperform randomly chosen points.

Remark 2. For more details on discrepancy and low discrepancy point sets we refer the reader to Kuipers and Niederreiter [1974], Niederreiter [1992]. Usually low discrepancy point sets are used as deterministic alternative to Monte Carlo integration, it is therefore called *Quasi-Monte Carlo integration*. In Finance such methods are also well known. We mention [Glasserman, 2003, Sect. 5] for more information in this context.

We explicitly mention the importance of *digital nets*, see [Glasserman, 2003, Sect. 5.1.4], which constitute point sets of low discrepancy and may also be regarded as stratified samples, as shown below in proposition 1. The original construction goes back to [Sobol', 1961] and has been generalized by H. Niederreiter to provide constructions which yield better discrepancy behavior. In recent years such nets are available in high dimensions, which make them suitable in real applications.

The general construction is based on 4 parameters, a base b for which the construction is carried out, d indicating the spatial dimension, m indicating the number of points as power of the base b , and a fourth parameter t , which indicates the quality of the net.

Precisely, having fixed b , we consider b -ary boxes in dimension d , given through

$$(10) \quad \prod_{j=1}^d \left[\frac{a_j}{b^{l_j}}, \frac{a_j + 1}{b^{l_j}} \right),$$

where $l_j \in \{0, 1, 2, \dots\}$ and $a_j \in \{0, 1, \dots, b^{l_j} - 1\}$. The volume of such box is $b^{-l_1 - \dots - l_d}$.

Definition 1. A set of b^m points in $[0, 1]^d$ is said to be a (t, m, d) -net, if each b -ary box of volume b^{t-m} contains exactly b^t points.

Thus, the fraction of points in each such box coincides with the volume of the box.

There is an intimate connection between orthogonal arrays and digital nets. A discussion on the relation between digital nets can be found in [Niederreiter, 1992, § 4.2]. This relation is also the background for scrambling of digital nets, see Owen [1997]. Here we exhibit a direct relation, which probably goes back to Lawrence [1996]. For its formulation we define the mapping Ψ , which assigns each point $x = (x^1, \dots, x^d)$ in $[0, 1]^d$ a d -tuple in $\{0, \dots, N - 1\}^d$ by letting

$$\Psi(x) := (\lfloor x^1 N \rfloor, \dots, \lfloor x^d N \rfloor).$$

Proposition 1. *Let $\{x_1, \dots, x_{N^s}\}$ be a $(0, s, d)$ -net in base N . Then the $d \times N^s$ -matrix with columns $\Psi(x_1), \dots, \Psi(x_{N^s})$ constitutes an orthogonal array $\text{OA} = \text{OA}(N^s, d, N, s)$ of strength s .*

Proof. Let OA be constructed as above and choose any s coordinates, say $S := \{d_1, \dots, d_s\}$. For each $j = 1, \dots, N^s$ let n_S^j denote the restriction of the j -th column n_j to S . We need to show, that the $s \times N^s$ matrix with columns $n_S^1, \dots, n_S^{N^s}$ has full rank. But this follows from

the observation that the set $n_S^1, \dots, n_S^{N^s}$ equals $\{0, \dots, N-1\}^d$. Indeed, let us choose any d -tuple $n = (n_1, \dots, n_d)$ from $\{0, \dots, N-1\}^d$. The N -ary box $\prod_{l \in S} [\frac{n_l}{N}, \frac{n_l+1}{N}) \times \prod_{l \notin S} [0, 1)$ has volume N^{-s} , thus contains exactly one point, say x_j , such that $n = n_S^j$. \square

It is worth mentioning, that digital nets as well as orthogonal arrays of given strength are not available for all sample sizes. Their construction is usually based on results from the theory of finite fields, we refer to Niederreiter [1992] and Glasserman [2003]. In particular, for strength 2, given dimension d the number N of symbols must obey $N \geq d-1$ for orthogonal arrays of strength 2 to exist. This is one of the fundamental results and we refer the reader to [Niederreiter, 1992, Cor. 4.21] or the seminal paper [Bose and Bush, 1952, Thm.], where one can also find a historical account. It is interesting to note, that analogous limitations hold true for orthogonal arrays of higher strength, see [Hedayat et al., 1999]. Scrambling of digital nets has been initiated by [Owen, 1997] and further been substantiated by many authors. Again we refer to [Glasserman, 2003, § 5.4] for more details. The simulations in section 4 will use randomized Sobol points. This sampling method will be denoted by “dnet”, below.

3.3. Theoretical results. The theoretical analysis of the performance of randomized orthogonal arrays started more than ten years ago, see e.g. Owen [1992], Tang [1993], some of the features were mentioned in Patterson [1954] without proof.

If a sample set is obtained from a randomized orthogonal array of strength t or a scrambled digital net, then we shall denote its sample mean at a function f by $\vartheta_n^{t\text{-oa}}(f)$ and $\vartheta_n^{\text{dnet}}(f)$, respectively. The error analysis for orthogonal arrays and for scrambled nets is developed, we have chosen to keep close to [Owen, 1997].

The basic observation for randomized orthogonal arrays is comprised next. Let P_t denotes the orthogonal projection onto the space of superpositions of t -variate functions. For Latin hypercube sampling this is the projection onto the span of all additive functions.

Proposition 2. *Let f be any square-integrable function. The error of sampling using randomized orthogonal arrays of strength t obeys*

$$\sqrt{ne}(f, \vartheta_n^{t\text{-oa}}) \longrightarrow \|f - P_t f\|.$$

As a consequence, for $f_{t\text{-var}}$ which are superpositions of t -variate functions, we have

$$\sqrt{ne}(f_{t\text{-var}}, \vartheta_n^{t\text{-oa}}) \longrightarrow 0.$$

Note, that this provides the typical Monte Carlo rate. But, while the error of plain Monte Carlo is controlled by the distance of f to the constant functions, the error using Latin hypercubes is controlled by the distance to the univariate ones and the error using randomized orthogonal arrays of strength t is controlled by the distance of f to superpositions of t -variate functions.

For the specific class of functions $f_x(u) := \chi_{(-\infty, x)}(f(u))$, $u \in [0, 1]^d$ relevant in the context of approximating cdfs we may rephrase Proposition 2 as

$$(11) \quad n \operatorname{Var}(F_n(t - oa, x)) \longrightarrow \operatorname{dist}^2(f_x, tVAR) \leq F(x)(1 - F(x)).$$

Uniform bounds may be stated as follows, see [Owen, 1997, Thm. 3, Prop. 4].

Theorem 1. (1) *Let $m \geq d$. For scrambled $(0, m, d)$ -nets in base b the following bound holds true.*

$$(12) \quad \kappa_n^{sup}(dnet) \leq (b/(b-1))^{(d-1)/2}.$$

(2) *For randomized orthogonal arrays of strength $t = 1, 2$ we have*

$$(13) \quad \kappa_n^{sup}(t - oa) \leq \frac{N}{N-1}.$$

Remark 3. For $(0, m, d)$ -nets in base b to exist necessarily it holds $b \geq d$. Thus from (12) we deduce $\kappa_n^{sup}(dnet) \leq \sqrt{e}$.

The above bounds are pessimistic and allow to state only that one does not loose using stratification compared to plain MC. In practice one can observe $\kappa_n^{sup}(dnet) \leq \kappa_n^{sup}(t - oa) \leq 1$, see 4.1. This can actually be proven for $\kappa_n^{rms}(t - oa)$.

Theorem 2. *Suppose that there is $C < \infty$ for which*

$$(14) \quad \kappa_n^{sup}(t - oa) \leq C.$$

Then

$$(15) \quad \kappa_n^{rms}(t - oa)^2 \longrightarrow \int \frac{\operatorname{dist}(f_x, tVAR)^2}{F(x)(1 - F(x))} dF(x).$$

Proof. This is an easy consequence of the Lebesgue dominated convergence theorem. Indeed, by (11) we have point-wise convergence

$$\kappa_n(t - oa, x)^2 \rightarrow \operatorname{dist}(f_x, tVAR)^2 / F(x)(1 - F(x)) \leq 1,$$

Because by (14) the sequence $\{\kappa_n(t - oa, x)^2\}$ is uniformly bounded we obtain convergence of the integrals. \square

Remark 4. Notice that by (13) assumption (14) is fulfilled for arrays of strength 1 or 2. We conjecture this to be true in general.

Summarizing our theoretical findings we can state, that stratification using randomized orthogonal arrays, in particular those of strength 1 (Latin hypercubes) or strength 2 or scrambled digital nets provides samples which are not worse than MC, when quality is measured as above. For the expected Anderson–Darling distance this bound can be made precise. The same holds true, given any level, for the distance between the true and the empirical quantile.

3.4. Combining orthogonal arrays with other means of variance reduction. We will indicate how to combine the use of orthogonal arrays or other means of stratification with importance sampling on the basis of the algorithm [Glasserman, 2003, § 9.2, Importance Sampling Algorithms]. There the algorithm is based on changing from the original probability measure under which the vector Z is standard normal to the measure under which $Z \sim N(\mu, \Lambda^2)$, for a suitably chosen mean μ and diagonal matrix Λ . Based on these parameters the description of the importance sampling algorithm is reproduced here (in our notation) as

1. (a) Generate $U \in [0, 1]^d$ uniformly.
- (b) Generate $Z \sim N(\mu, \Lambda^2)$ by $Z := N^{-1}(U)\Lambda + \mu$.
- (c) Evaluate $X := BZ$.
- (d) Evaluate $Q(X)$ from equation (2).
- (e) Calculate

$$(16) \quad L(Q)\mathbf{1}\{Q < x\},$$

where $L(Q)$ is the appropriate likelihood ratio for the measure change.

2. Calculate average over n replications of step 1.

Stratification may be incorporated in this algorithm simply by replacing the MC-sample U with a randomized orthogonal array or a scrambled digital net. The performance of this modification is compared to the original one in 4.2.

4. CASE STUDY

The goal of this case study is to compare the efficiency of the stratification methods described so far with the other established methods, notably Fourier inversion and the importance sampling methods described by [Glasserman, 2003, section 9.2], on a real-world problem.

The sample problem comes from a medium-sized bank that uses the Δ - Γ -normal approach to Value at Risk for regulatory purposes. It uses about 1000 risk factors. On the specific date all sensitivities w.r.t. more than half of the risk factors were zero, such that the problem reduced

to dimension $d = 479$. Σ is the empirical covariance matrix $\Sigma = BB^\top$, where B is the $d \times l$ sample matrix from the last $l = 250$ business days.

4.1. Approximation of the whole distribution. In the first simulation study we compare the behavior of $\kappa_n^{rms}(meth)$ for different sampling schemes “meth”. The approximation errors for the whole distribution are displayed in figure 1 for the following methods:

- (1) Latin hypercube sampling (“lhs”),
- (2) randomized orthogonal arrays of strength 2 (“2-oa”),
- (3) randomized digital nets, Sobol sequences (“dnet”),
- (4) stratified sampling along the linear approximation (“strat”).

Implementation details. The first three methods all use the same function to transform an $n \times d$ -matrix U of n d -dimensional uniform (Q)MC-samples into an n -sample Q of the quadratic form:

$$(17) \quad X := \Phi^{-1}(U)B^\top$$

$$(18) \quad Q_i := \theta + (X\Delta)_i + 0.5 \sum_{(j,k):\Gamma_{jk} \neq 0} X_{ik}\Gamma_{jk}X_{ij}.$$

The sample matrix X can be computed once per day offline, such that the online computations (18) require $(2 + 3c_{nz})nd$ floating point operations.⁵

The generation of a Latin hypercube sample matrix U is standard, see [Mathé, 2001] or [Glasserman, 2003].

The construction of orthogonal arrays, in particular of low strength is interesting and has a long history, we recommend the recent monograph Hedayat et al. [1999]. Our implementation is based on the seminal paper Bose and Bush [1952], that proves that

$$C \quad \times \quad mm \quad \text{mod } N$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \dots & \\ 1 & d-1 \\ 1 & d \end{pmatrix} \times \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 2 & 2 & \dots \\ 1 & 2 & 3 & \dots & N & 1 & 2 & \dots \end{pmatrix} \quad \text{mod } N.$$

produces an orthogonal array of strength two in dimension d , if N is prime and larger than d . The result is a $d \times N^2$ -matrix, providing N^2 d -dimensional vectors. For $d = 3, N = 3$, this provides the array OA from (7).

⁵As in section 2 we count only those terms with at least two of the “large” factors n, d or l and ignore the constant and linear terms.

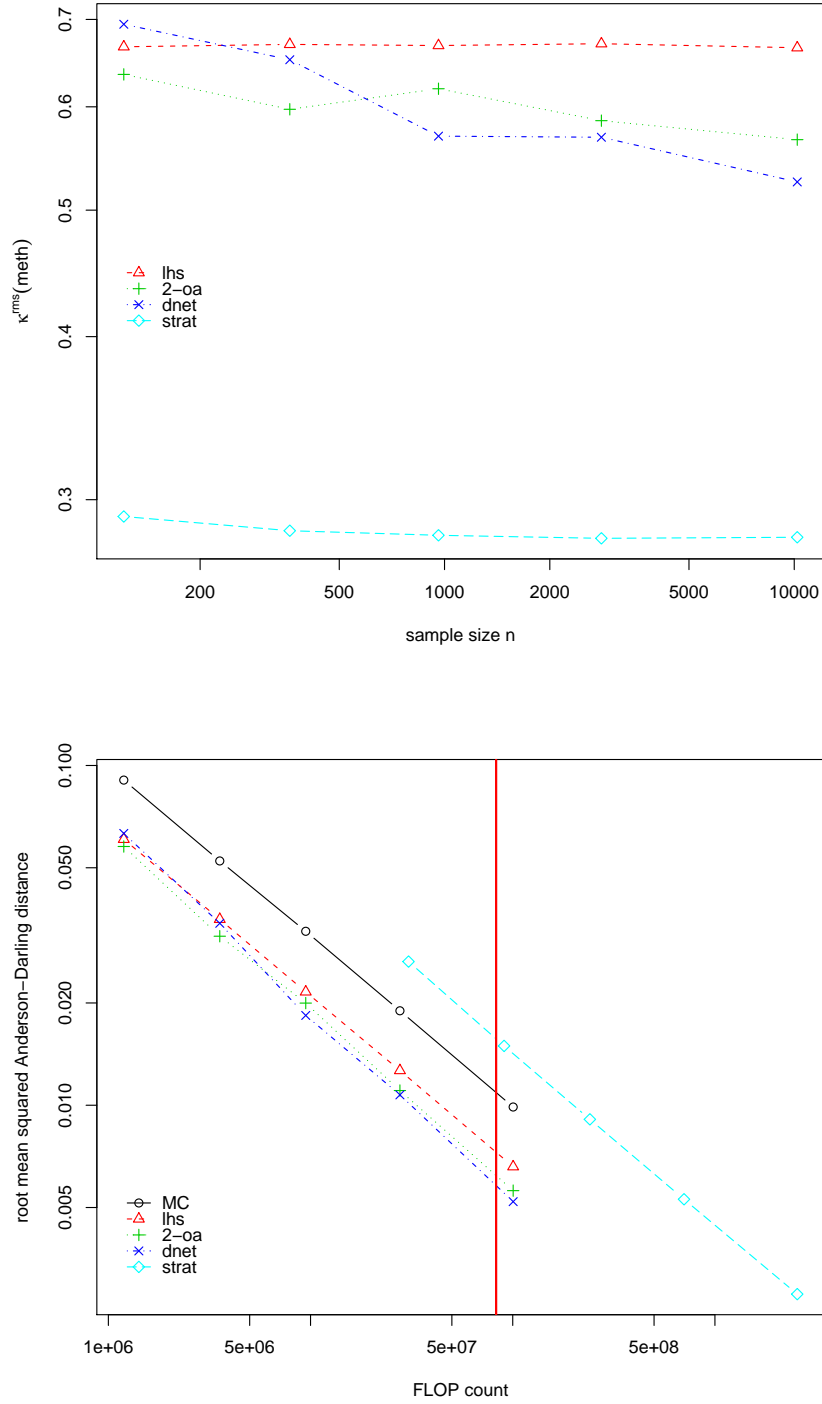


FIGURE 1. Approximation error for the whole distribution: improvement over MC and Anderson-Darling distance

For details of the implementation of digital nets we refer the reader to [Glasserman, 2003, § 5.2], where many details are given concerning Sobol’ nets as well as Faure nets. There are several packages available which provide digital nets. Our computations are based on the package **netgen**, provided by the Hong Kong Baptist University. We stress, that only recently such point sets are available in high dimensions, such that they are now applicable to real-world problems.

Method (4) “strat” stratifies the sample $X = \Phi^{-1}(U)B^\top$ along the direction Δ , as described in [Glasserman, 2003, p.223]. The sample of standard normal vectors $Z_i := \Phi^{-1}(U_{i,\cdot})$ and a stratified sample of standard normal univariates $Z_i^0 := \Phi^{-1}((i-1+U_i^0)/n)$ can be computed offline. Next, we compute the stratification of Z along the appropriate direction w :

$$(19) \quad w := B^\top \Delta / \|B^\top \Delta\|,$$

$$(20) \quad \tilde{Z}_i := Z_i^0 w + (I - ww^\top) Z_i$$

$$(21) \quad X_i := B \tilde{Z}_i.$$

$(I - ww^\top)$ is the projector onto the hyperplane defined by w and $\langle w, \tilde{Z}_i \rangle = Z_i^0$. I.e., (20) stratifies \tilde{Z} along w . This is equivalent to the stratification of X along Δ , since $\langle \Delta, X_i \rangle = \|B^\top \Delta\| \langle w, \tilde{Z}_i \rangle$.

Analysis. As can be drawn from figure 1, the improvement in accuracy is about 60% for “2-oa” and “dnet” at the relevant sample size. This is achieved by stratification, only. Note, that the construction for “2-oa” is guaranteed to provide valid orthogonal arrays of strength 2 for the sample problem with $l = 250$ only if $n \geq 251^2 = 63001$, while Latin Hypercube sampling works for any n . Indeed, the values κ_n^{rms} in the upper part of figure 1 show that the variance reduction of “2-oa” improves with increasing n , while Latin Hypercube sampling has the same variance reduction for all n . A similar argument holds for the randomized digital nets “dnet”.

Method (4) clearly outperforms the other three sampling strategies. One reason for this is that the sample problem has a relatively large $|\Delta|$ compared to $|\Gamma|$. For “delta-hedged” portfolios with $\Delta \approx 0$ the last method would not produce any reduction in the approximation error. On the other hand, this reduction of the approximation error by method (4) comes with the additional FLOP counts $2ld$ (19), $3nl$ (20), and $2nld$ (21). In the case when l is a constant fraction of d and $d \rightarrow \infty$, this method would asymptotically require $\mathcal{O}(d^2)$ FLOPS, compared to $\mathcal{O}(d^3)$ for Fourier inversion and $\mathcal{O}(d)$ for the first three stratification methods.

Therefore the lower part of figure 1 compares the average Anderson-Darling distances against floating point operation counts instead of sample sizes. The vertical line shows the operation count of the tri-diagonalization (“setup costs for Fourier inversion” in table 1). The results show that the stratification methods “2-oa” and “dnet” achieve the best approximation in the area of interest: with fewer operations than the setup costs of Fourier inversion (for about $n \leq 3500$). If the corresponding accuracy is sufficient, these methods are faster than Fourier inversion.

4.2. Approximation of the 1%-Quantile. The approximation of the cdf at specific quantiles is considered for the following methods additionally to the first four methods of the last subsection:

- (5) importance sampling by exponential tilting (“IS:et”), based on the linear term $\langle \Delta, X \rangle$,
- (6) the modification (“IS:et+dnet”) of (“IS:et”), as described in 3.4, using randomized digital nets.
- (7) the modification (“IS:et+strat”) of (“IS:et”) by stratifying along the linear term, cf. [Glasserman, 2003, § 9.2.3].

Implementation Details. Method (5) “IS:et” is the exponential tilting described in [Glasserman, 2003, section 9.2.2, p.495] and reproduced in section 3.4, except that the quadratic term is ignored in the measure change⁶. The tilting parameter is $\rho := \Phi^{-1}(0.01)/\sigma_l$, where $\delta := B^\top \Delta$ and $\sigma_l = |\delta|$ is the standard deviation of the linear part $\langle \Delta, X \rangle = \langle \delta, Z \rangle$. The sampling of Z happens under the new mean $\mu := \rho\delta$ instead of mean zero. This causes the linear part $\langle \Delta, X \rangle = \langle \delta, Z \rangle$ to have mean $\sigma_l \Phi^{-1}(0.01)$; i.e., most samples lie near the 1%-quantile of the affine linear part of the quadratic form. The variance reduction achieved by “IS:et” comes with the floating point operation counts $2ld$ (for the computation of δ), $2ld$ (for the computation of $\mu_X = B\mu$), and nd (for the shifting of X by μ_X), additionally to the $(2 + 3c_{nz})nd$ FLOPS needed by the first 4 methods⁷.

Method (6) “IS:et+dnet” has the same operation count as “IS:et”.

⁶Tilting in the exponential family generated by the quadratic form itself requires the repeated evaluation of the cumulant generating function of the quadratic form and thus has the same setup costs as Fourier inversion.

⁷The computational cost of the likelihood ratio $\exp\{-\rho \langle \Delta, X \rangle + (\rho \sigma_l)^2/2\}$ is proportional to n (given σ_l and the linear part $\langle \Delta, X \rangle$ are computed before). As before, only terms with at least two of the large factors l , n or d are considered in the operation counts.

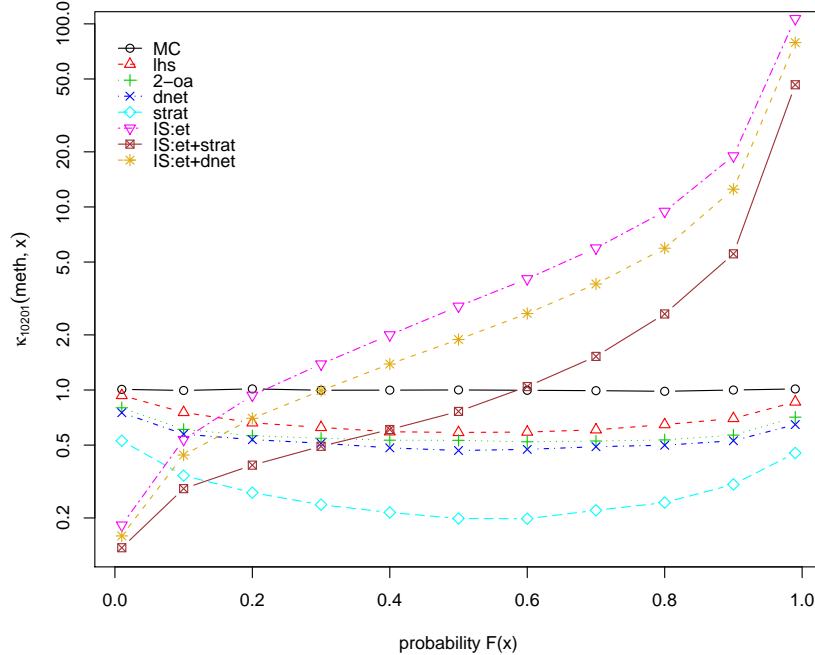


FIGURE 2. The relative improvement over MC at different quantiles: $\kappa_n(\text{meth}, x)$ versus $F(x)$.

Method (7) “IS:et+strat” uses the same measure change as method (7) and additionally stratifies X along Δ like method (5). This method has the same operation count as method (5).

Analysis. Figure 2 illustrates that those methods that employ importance sampling tuned to the 1%-quantile achieve a significant variance reduction at that quantile, but sacrifice accuracy at other quantiles. The stratification methods of the previous subsection improve uniformly on MC, but less in the tails than in the center of the distribution.

The upper part of figure 3 displays the improvement over MC at the 1%-quantile ($\kappa_n(\text{meth}, 0.01)$). It is evident that importance sampling with exponential tilting outperforms those methods which were favorable for approximating the distribution as a whole. Thus importance sampling with exponential tilting is appropriate for variance reduction when approximating VaR. Again this can be improved by additional stratification. While “IS:et+strat” gives a better variance reduction than “IS:et+dnet”, the lower part in figure 3 clearly exhibits that

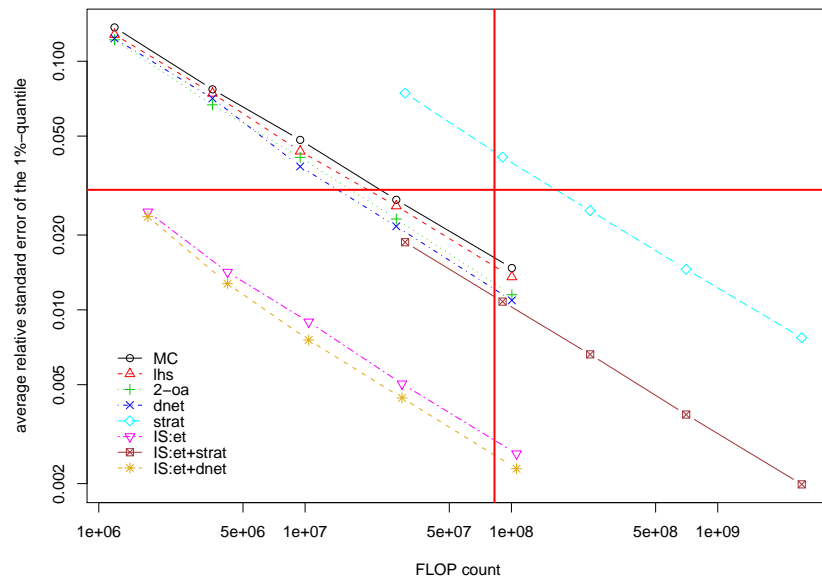
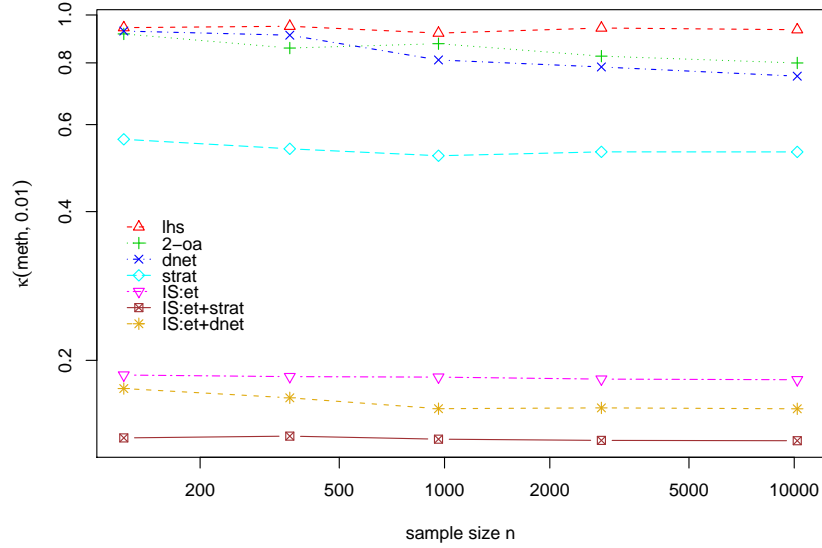


FIGURE 3. Approximation error for the 1%-quantile: improvement over MC and relative standard error at the 1%-quantile ($= e(q, \hat{q}_n(\text{meth}))/q$) for different sampling methods.

the higher accuracy of “IS:et+strat” is a result of additional FLOPS, whereas “IS:et+dnet” has the same FLOP count as the original method “IS:et”. As mentioned in 3.4 this additional accuracy is achieved by just replacing independent samples by (correlated) scrambled Sobol points.

The vertical line shows the operation count of the tri-diagonalization (“setup costs for Fourier inversion” in table 1). Assume that the 99%-Value at Risk needs to be known with 95% confidence with a relative error of 5%. This translates to a standard relative error of about 0.03, which is the horizontal line in figure 3. On this sample problem, this accuracy is achieved even by plain vanilla Monte Carlo with fewer operations than the setup costs for Fourier inversion. Further reductions in computational effort can be achieved by randomized orthogonal arrays and digital nets in general; and more drastic reductions in computational effort with the other stratification methods if the Γ -term is so small compared to the Δ -term as in this real-world problem.

5. CONCLUSION

The authors analyzed stratified sampling in risk management, both theoretically and empirically on the basis of real data. A comparison of operation counts showed that Monte Carlo methods tend to be more efficient than Fourier inversion for larger dimensions d . On a real-world problem plain vanilla Monte-Carlo was shown to be faster than Fourier inversion for the accuracy that is required in this context. The break-even is even reached on a relatively small problem of dimension $d = 113$ (not shown).

The performance of Monte-Carlo can be improved by variance reduction techniques, for example randomized orthogonal arrays, which stratify along lower dimensional margins. These are easy to implement and improve upon plain Monte Carlo in terms of the approximation of the whole distribution (risk management in a broader sense).

In terms of the approximation of specific quantiles (Value at Risk), importance sampling leads to the most significant variance reductions. Again, this can be combined with stratification to further improve the accuracy.

Hence, the stratification methods described in this paper beat Fourier inversion on a large range of practical problems.

ACKNOWLEDGMENT

The first author points out that the views expressed herein should not be construed as being endorsed by the BaFin. The work on the

paper was initiated at the Weierstrass Institute and the first author likes to thank his former colleagues for fruitful discussions.

Part of the work was done during the second author's visit to the Hong Kong Baptist University, Hong Kong. He wants to thank Fred Hickernell and Gang Wei for their hospitality and interest in this work.

REFERENCES

- Claudio Albanese, Ken Jackson, and Petter Wiberg. Fast convolution method for VaR and VaR gradients. <http://www.math-point.com/fconv.ps>, Aug 2000.
- E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. SIAM, third edition, Aug 1999. <http://www.netlib.org/lapack/lug/>.
- R. C. Bose and K. A. Bush. Orthogonal arrays of strength two and three. *Ann. Math. Statistics*, 23:508–524, 1952.
- Mark Britton-Jones and Stephen Schaefer. Non-linear Value-at-Risk. *European Finance Review*, 2:161–187, 1999.
- William Fallon. Calculating Value at Risk. <http://wrdsenet.wharton.upenn.edu/fic/wfic/papers/96/9649.pdf>, Jan 1996. Wharton Financial Institutions Center Working Paper 96-49.
- Andrey Feuerverger and Augustine C.M. Wong. Computation of Value at Risk for nonlinear portfolios. *Journal of Risk*, 3(1):37–55, Fall 2000.
- Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Appl. of Math. Springer, 2003.
- Christian Gouriéroux. *ARCH models and financial applications*. Springer-Verlag, New York, 1997. ISBN 0-387-94876-7.
- A. S. Hedayat, N. J. A. Sloane, and John Stufken. *Orthogonal arrays*. Springer Series in Statistics. Springer-Verlag, New York, 1999. ISBN 0-387-98766-5.
- Stefan Jaschke. The Cornish-Fisher-expansion in the context of delta-gamma-normal approximations. *Journal of Risk*, 4(4):33–52, 2002a.
- Stefan Jaschke. Error analysis of quantile-approximations using Fourier inversion in the context of delta-gamma-normal models (maths of all angles). *GARP Risk Review*, 06:28–33, May/June 2002b.
- L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences*. Wiley & Sons, New York, London, Sydney, Toronto, 1974.

- Diego Kuonen. Saddlepoint approximations for distributions of quadratic forms in normal variables. *Biometrika*, 86(4):929–935, 1999. ISSN 0006-3444.
- K. Mark Lawrence. A combinatorial characterization of (t, m, s) -nets in base b . *J. Combin. Des.*, 4(4):275–293, 1996. ISSN 1063-8539.
- David Li. Value at Risk based on the volatility, skewness and kurtosis. <http://www.riskmetrics.com/research/working/var4mm.pdf>, Mar 1999. RiskMetrics Group.
- Jacques Longerstae. RiskMetrics technical document. Technical Report fourth edition, J.P.Morgan, Dec 1996. originally from <http://www.jpmorgan.com/RiskManagement/RiskMetrics/>, now <http://www.riskmetrics.com>.
- George Marsaglia and John C. W. Marsaglia. Evaluating the anderson-darling distribution. *Journal of Statistical Software*, 9(2), 2004. <http://www.jstatsoft.org/index.php?vol=9>.
- Peter Mathé. Hilbert space analysis of Latin hypercube sampling. *Proc. Amer. Math. Soc.*, 129(5):1477–1492, 2001. ISSN 0002-9939.
- Jorge Mina and Andrew Ulmer. Delta-gamma four ways. <http://www.riskmetrics.com>, Aug 1999.
- Harald Niederreiter. *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. ISBN 0-89871-295-5.
- Art B. Owen. Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica*, 2(2):439–452, 1992. ISSN 1017-0405.
- Art B. Owen. Monte Carlo variance of scrambled net quadrature. *SIAM J. Numer. Anal.*, 34(5):1884–1910, 1997. ISSN 0036-1429.
- H. D. Patterson. The errors of lattice sampling. *J. Roy. Statist. Soc. Ser. B.*, 16:140–149, 1954.
- Stefan Pichler and Karl Selitsch. A comparison of analytical VaR methodologies for portfolios that include options. http://www.tuwien.ac.at/E330/Research/paper_var.pdf, Dec 1999. Working Paper TU Wien.
- Matthew Pritsker. Evaluating Value at Risk methodologies: Accuracy versus computational time. <http://wrdsenet.wharton.upenn.edu/fic/wfic/papers/96/9648.pdf>, Nov 1996. Wharton Financial Institutions Center Working Paper 96-48.
- C. Rouvinez. Going greek with VaR. *Risk*, 10(2):57–65, Feb 1997.
- I. M. Sobol'. An exact bound for the error of multivariate integration formulas for functions of classes \tilde{W}_1 and \tilde{H}_1 . *Ž. Vyčisl. Mat. i Mat. Fiz.*, 1:208–216, 1961.

- Michael Studer. *Stochastic Taylor Expansions and Saddlepoint Approximations for Risk Management*. PhD thesis, ETH, Zürich, November 2001. <http://www.math.ethz.ch/~studerm/diss/arbeit.pdf>.
- Boxin Tang. Orthogonal array-based Latin hypercubes. *J. Amer. Statist. Assoc.*, 88(424):1392–1397, 1993. ISSN 0162-1459.
- Ulrike Volmar. Comparison of different methods for calculation of delta-gamma value at risk. http://www.d-fine.de/pool/bibliothek/ox_uvo_02.pdf, Nov 2002. Diploma Thesis, University of Oxford.
- P. Zangari. How accurate is the delta-gamma methodology? *RiskMetrics Monitor*, 1996(third quarter):12–29, 1996a.
- P. Zangari. A VaR methodology for portfolios that include options. *RiskMetrics Monitor*, 1996(first quarter):4–12, 1996b.

BUNDESANSTALT FÜR FINANZDIENSTLEISTUNGS-AUFSICHT, GRAURHEINDORFER STRASSE 108, 53117 BONN

E-mail address: stefan@jaschke-net.de

WEIERSTRASS INSTITUTE FOR APPLIED ANALYSIS AND STOCHASTICS, MOHRENSTRASSE 39, D-10117 BERLIN, GERMANY

E-mail address: mathe@wias-berlin.de