Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

EFFICIENT STRATIFIED SAMPLING IMPLEMENTATIONS IN MULTIRESPONSE SIMULATION

İsmail Başoğlu
Department of International Logistics Management
Kemerburgaz University
Bağcılar, 34217
İstanbul, TURKEY

Wolfgang Hörmann
Department of Industrial Engineering
Boğaziçi University
Bebek, 34342
İstanbul, TURKEY

ABSTRACT

Often the accurate estimation of multiple values from a single simulation is of practical importance. Among the many variance reduction methods known in the literature, stratified sampling is especially useful for such a task as the allocation fractions can be used as decision variables to minimize the overall error of all estimates. Two different classes of overall error functions are proposed. The first, including the mean squared absolute and the mean squared relative error, allows for a simple closed-form solution. For the second class of error functions, including the maximal absolute and the maximal relative error, a simple and fast heuristic is proposed. The application of the new method, called "multiresponse stratified sampling", and its performance are demonstrated with numerical examples.

1 INTRODUCTION

We consider the problem of estimating multiple values using Monte Carlo simulation. We define the response function $r(\boldsymbol{\theta})$ as the expectation of a random-valued function under parameter $\boldsymbol{\theta}$ where Θ is the set of possible parameter values. We assume that $r(\boldsymbol{\theta})$ can only be estimated via simulation for each $\boldsymbol{\theta} \in \Theta$. Moreover, we assume that the dimension of the random input is not changed by the parameters. Then, it is possible to simulate r for several $\boldsymbol{\theta}$ values in a single simulation. Such an objective can easily be realized using common random numbers (Law 2014). However, under this simple approach, all $r(\boldsymbol{\theta})$ values are estimated with comparatively large variances. To obtain more accurate estimates for several $r(\boldsymbol{\theta})$ values, one can apply variance reduction methods such as antithetic variates (Myers and Montgomery 2002), control variates (see e.g., Rubinstein and Marcus 1985), and importance sampling (see e.g., Glasserman and Li 2005 and Sak and Hörmann 2012).

A variance reduction method that can be very useful for the above problem is stratified sampling with optimally selected allocation fractions. Interestingly, this seems to be a fact overlooked in the literature and only used in Başoğlu, Hörmann, and Sak (2013). The aim of this paper is to show how stratification can be used to minimize the overall error of all estimates. We, therefore, introduce objective functions that measure the overall error and develop simple methods to minimize (approximately) these objective functions using the allocation fractions as decision variables for the optimization problem. We consider (as a new contribution) the minimization of linear functions of the variance-covariance matrix of the stratified estimates. In addition, the minimization of the maximal absolute and the maximal relative error of the estimates is discussed (see Başoğlu, Hörmann, and Sak 2013 for the application of that idea for risk simulations).

The allocation fractions calculated by the optimization algorithm are then used in the sampling phase. We call the resulting method "multiresponse stratified sampling". The method is applicable to simulation problems where the size of the random input is independent of the parameter space, and it should exhibit a good performance for simulation problems for which stratification reaches a good variance reduction in the single response case. The new method can be useful in stochastic optimization and response surface applications as they require the evaluation of the simulation response for many different parameter values (Box and Wilson 1951).

The paper is organized as follows. In Section 2, we give a brief description of the stratified sampling method. Section 3 presents the nonlinear optimization models and the methodology of multiresponse stratified sampling. In Section 4, we demonstrate the efficiency of multiresponse stratified sampling with several examples and present our experimental results. We give our final comments in Section 5.

Note that, in this paper, vectors and matrices are set in bold to enhance readability.

2 SIMULATION WITH STRATIFIED SAMPLING

Let $\boldsymbol{X} \in \mathbb{R}^D$ be a random vector with density $f_{\boldsymbol{X}}(.)$, and $q : \mathbb{R}^D \to \mathbb{R}$ be a measurable simulation function such that $E\left[q^2(\boldsymbol{X})\right] < \infty$. Suppose that ξ_i, $i = 1,\dots,I$, is a partition of $\mathbb{R}^D$ into I strata and $p_i = \Pr\{\boldsymbol{X} \in \xi_i\}$ is known for $i = 1,\dots,I$. We want to estimate:

$$x = E\left[q(\boldsymbol{X})\right] = \sum_{i=1}^{I} p_i E\left[q(\boldsymbol{X}) | \boldsymbol{X} \in \xi_i\right]$$

based on the latter equality. Henceforth, $\boldsymbol{X}_i$ denotes the random vector that follows the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{X} \in \xi_i$. Let N denote the total number of replications, N_i denote the amount of drawings allocated to stratum i, such that $\sum_{i=1}^{I} N_i = N$, and $\boldsymbol{X}_i^n$, $n = 1,\dots,N_i$, denote the independent drawings of $\boldsymbol{X}_i$. We define the allocation fractions $\pi_i = N_i/N$ for $i = 1,\dots,I$. The stratified Monte Carlo estimator of x is formulated as:

$$\hat{x} = \sum_{i=1}^{I} p_i \hat{x}_i = \sum_{i=1}^{I} p_i N_i^{-1} \sum_{n=1}^{N_i} q(\boldsymbol{X}_i^n) = N^{-1} \sum_{i=1}^{I} p_i \pi_i^{-1} \sum_{n=1}^{\pi_i N} q(\boldsymbol{X}_i^n),$$

where $\hat{x}_i = N_i^{-1} \sum_{n=1}^{N_i} q(\boldsymbol{X}_i^n)$ is the estimated mean conditional on stratum i. Let s_i^2 denote the variance of $q(\boldsymbol{X})$ conditional on the i-th stratum, namely $s_i^2 = V\left[q(\boldsymbol{X}_i)\right] = V\left[q(\boldsymbol{X}) | \boldsymbol{X} \in \xi_i\right]$. Then, the variance of the stratified estimator is:

$$V\left[\hat{x}\right] = \sum_{i=1}^{I} p_i^2 N_i^{-1} s_i^2 = N^{-1} \sum_{i=1}^{I} \pi_i^{-1} p_i^2 s_i^2. \quad (1)$$

The variance of the stratified estimator has the lower bound $V^*\left[\hat{x}\right] = N^{-1}\left(\sum_{i=1}^{I} p_i s_i\right)^2$, which can be attained if we use the optimal allocation fractions (see e.g., Glasserman 2004):

$$\pi_i^* = p_i s_i \bigg/ \sum_{l=1}^{I} p_l s_l, \quad i = 1,\dots,I. \quad (2)$$

Usually, we lack the prior information about the conditional standard deviations, s_i. Therefore, it is not possible to decide about the optimal allocation fractions beforehand. A simple solution is to use the estimates of conditional standard deviations, $\hat{s}_i$, obtained with a pilot sample of total size N_p. To accurately estimate the optimal allocation sizes, we suggest selecting a sufficiently large pilot sample size. The remaining $N - N_p$ drawings can then be used in the main run according to the allocation rule in (2).

3 MULTIRESPONSE STRATIFIED SAMPLING

We define the response function $r(\boldsymbol{\theta}) = E\left[q(\boldsymbol{X}, \boldsymbol{\theta})\right]$, where $r : \Theta \to \mathbb{R}$ and $\boldsymbol{X} \in \mathbb{R}^D$ follows a common distribution that is independent of $\boldsymbol{\theta}$. We assume that for each $\boldsymbol{\theta} \in \Theta$, $r(\boldsymbol{\theta})$ can only be estimated via simulation. If we can find a small number of effective stratification variables in $\boldsymbol{X}$, which have a large contribution to $V\left[q(\boldsymbol{X}, \boldsymbol{\theta})\right]$ and are computationally tractable (i.e., we can generate $\boldsymbol{X}$ conditional to strata defined by these variables), then that stratification will effectively reduce the variance of the estimates.

Suppose we are given J points, $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_J$, in Θ. Our goal is to estimate $x_j = r(\boldsymbol{\theta}_j)$ for $j = 1,\ldots,J$ in a single simulation using stratified sampling. Let $\hat{x}_{ij}$ be the mean conditional on stratum i estimated under parameter $\boldsymbol{\theta}_j$. Then, the stratified estimator of x_j is calculated as $\hat{x}_j = \sum_{i=1}^{I} p_i \hat{x}_{ij}$. Each of these estimates is unbiased and asymptotically normal (Etoré and Jourdain 2010). Let $\hat{s}_{ij}^2$ be the variance of the sample in stratum i drawn under parameter $\boldsymbol{\theta}_j$. We also define $\hat{s}_i^{jk}$ as the covariance of the samples in stratum i drawn under parameters $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_k$. Given the total sample size, N, and the vector, $\boldsymbol{\pi} = (\pi_1,\ldots,\pi_I)'$, that holds allocation fractions, the elements of the sample variance-covariance matrix $\boldsymbol{\Sigma}$ of $\hat{\boldsymbol{x}}$ can be calculated as a function of the allocation fractions:

$$\Sigma_{jj}(\boldsymbol{\pi}) = N^{-1} \sum_{i=1}^{I} \pi_i^{-1} p_i^2 \hat{s}_{ij}^2, \quad j = 1,\ldots,J, \tag{3}$$

$$\Sigma_{jk}(\boldsymbol{\pi}) = N^{-1} \sum_{i=1}^{I} \pi_i^{-1} p_i^2 \hat{s}_i^{jk}, \quad j = 1,\ldots,J, \quad k = 1,\ldots,J, \quad j \neq k. \tag{4}$$

We define the objective function $\omega(\boldsymbol{\pi}) = g(\boldsymbol{\Sigma}(\boldsymbol{\pi}))$, where $g : \mathbb{R}^{J\times J} \to \mathbb{R}^+ \cup \{0\}$ is a continuous function of the variance and covariance values. This objective function is used for representing the overall error of all J estimates, for example:

- $\omega_{MSE}(\boldsymbol{\pi}) = \sum_{j=1}^{J} \Sigma_{jj}(\boldsymbol{\pi})$, the mean squared error of all estimates,
- $\omega_{MSR}(\boldsymbol{\pi}) = \sum_{j=1}^{J} \hat{x}_j^{-2} \Sigma_{jj}(\boldsymbol{\pi})$, the mean squared relative error of all estimates,
- $\omega_{SUM}(\boldsymbol{\pi}) = \sum_{j=1}^{J} \sum_{k=1}^{J} \Sigma_{jk}(\boldsymbol{\pi})$, the sum of all elements in the variance covariance matrix,
- $\omega_{MAXE}(\boldsymbol{\pi}) = \max\{j : \Sigma_{jj}(\boldsymbol{\pi})\}$, the maximum of the squared errors of all estimates,
- $\omega_{MAXR}(\boldsymbol{\pi}) = \max\{j : \hat{x}_j^{-2} \Sigma_{jj}(\boldsymbol{\pi})\}$, the maximum of the squared relative errors of all estimates,

all of which are convex in $\boldsymbol{\pi}$. Therefore, we assume ω to be a convex function of $\boldsymbol{\pi}$ for practically relevant examples. Our objective is to solve the following optimization model:

$$\begin{array}{ll} \min & \omega(\boldsymbol{\pi}) \\ \text{s.t.} & \sum_{i=1}^{I} \pi_i = 1 \quad \text{and} \quad \pi_i \geq 0, \quad i = 1,\ldots,I. \end{array} \tag{5}$$

The constraints in (5) form a convex and bounded feasible region. Thus, (5) becomes a convex programming problem and a local optimum found in the feasible solution set will also be a global optimum. If a closed-form optimal solution is unavailable, the optimal solution of problem (5) can be found by using an interior-type method with trust regions (Byrd, Nocedal, and Waltz 2006). However, since we use the estimates of conditional variance and covariance values, $\hat{s}_{ij}^2$ and $\hat{s}_i^{jk}$, the optimal solution of an instance will only be an estimate for the real optimal allocation fractions. Thus, a sub-optimal solution for the model in (5) is enough in practice.

The above examples for the objective function $\omega(\boldsymbol{\pi})$ can be categorized in two general classes. The first three examples are all linear functions of the elements of the variance-covariance matrix $\boldsymbol{\Sigma}$. We were able to find closed-form solutions under this class of objective functions. The second class consists of the last two examples where we consider the maximum of the variances, Σ_{jj}, which are weighted with non-negative coefficients. We provide a heuristic method for the second class.

3.1 Minimizing a Linear Combination of the Elements of the Variance-Covariance Matrix

As a first example, suppose, we want to minimize the mean squared relative error of all estimates. We consider the objective function

$$\omega_{MSR}(\boldsymbol{\pi}) = \sum_{j=1}^{J} \hat{x}_j^{-2} \Sigma_{jj}(\boldsymbol{\pi}) = N^{-1} \sum_{i=1}^{I} \pi_i^{-1} p_i^2 \sum_{j=1}^{J} \hat{x}_j^{-2} \hat{s}_{ij}^2,$$

which has a form that is similar to the variance of the stratified estimator given in (1), only the expression of the conditional variances, s_i^2, is now replaced by $\sum_{j=1}^J \hat{x}_j^{-2}\hat{s}_{ij}^2$. The lower bound for $\omega_{MSR}(\boldsymbol{\pi})$ is:

$$\omega_{MSR}^*(\boldsymbol{\pi}) = N^{-1}\left(\sum_{i=1}^I p_i\left(\sum_{j=1}^J \hat{x}_j^{-2}\hat{s}_{ij}^2\right)^{1/2}\right)^2$$

and, according to (2), it can be attained if we choose the allocation fractions:

$$\pi_i^* = p_i\left(\sum_{j=1}^J \hat{x}_j^{-2}\hat{s}_{ij}^2\right)^{1/2} \Big/ \sum_{l=1}^I p_l\left(\sum_{j=1}^J \hat{x}_j^{-2}\hat{s}_{lj}^2\right)^{1/2}, \quad i=1,\ldots,I. \tag{6}$$

Using the pilot sample, we estimate the response values, $\hat{x}_j,\ j=1,\ldots,J$, and conditional variances, $\hat{s}_{ij}^2$, $i=1,\ldots,I,\ j=1,\ldots,J$. Then, we can use (6) to determine the optimal allocation of the sample in the main simulation, so that the mean squared relative error of all estimates is minimized.

Clearly, the arguments of above remain valid if we replace s_i^2 of Equation (1) by arbitrary non-negative constants. Thus, we can generalize the previous example by assuming g to be a linear function of the elements of $\boldsymbol{\Sigma}(\boldsymbol{\pi})$.

Theorem 1 Assume the objective function $\omega(\boldsymbol{\pi}) = \sum_{j=1}^J\sum_{k=j}^J c_{jk}\Sigma_{jk}(\boldsymbol{\pi})$ with real coefficients. We plug in the variance and the covariance values given in (3) and (4) into this objective function and obtain:

$$\omega(\boldsymbol{\pi}) = N^{-1}\sum_{i=1}^I \pi_i^{-1}p_i^2\left(\sum_{j=1}^J c_{jj}\hat{s}_{ij}^2 + \sum_{j<k} c_{jk}\hat{s}_i^{jk}\right).$$

For $\omega(\boldsymbol{\pi})$ to be convex in the feasible region of (5), a simple necessary condition is the non-negativity of $\sum_{j=1}^J c_{jj}\hat{s}_{ij}^2 + \sum_{j<k} c_{jk}\hat{s}_i^{jk}$ for $i=1,\ldots,I$. Under this assumption, the optimal solution that minimizes $\omega(\boldsymbol{\pi})$ is:

$$\pi_i^* = p_i\left(\sum_{j=1}^J c_{jj}\hat{s}_{ij}^2 + \sum_{j<k} c_{jk}\hat{s}_i^{jk}\right)^{1/2} \Big/ \sum_{l=1}^I p_l\left(\sum_{j=1}^J c_{jj}\hat{s}_{lj}^2 + \sum_{j<k} c_{jk}\hat{s}_l^{jk}\right)^{1/2}, \quad i=1,\ldots,I. \tag{7}$$

As a first application of Theorem 1, we have considered $\omega_{MSR}(\boldsymbol{\pi})$. As a second example, we consider minimizing the variance of a convex combination of all estimates, $\sum_{j=1}^J \lambda_j\hat{x}_j$ where $\lambda_j \geq 0,\ j=1,\ldots,J$, are fixed and $\sum_{j=1}^J \lambda_j = 1$. This is of practical relevance since, in response surface applications, we require linear approximations of the intermediate values using the estimates of $r(\boldsymbol{\theta}_j),\ j=1,\ldots,J$ (Box and Wilson 1951). Our objective function is, then, $\omega(\boldsymbol{\pi}) = \sum_{j=1}^J\sum_{k=1}^J \lambda_j\lambda_k\Sigma_{jk}(\boldsymbol{\pi})$. If we aim to minimize the variance of the average of all estimates, the coefficients λ_j become all equal and the objective function simplifies to the sum of all elements in the variance covariance matrix, $\omega_{SUM}(\boldsymbol{\pi}) = \sum_{j=1}^J\sum_{k=1}^J \Sigma_{jk}(\boldsymbol{\pi})$. For any case, the optimal solution can be found using (7) and the estimates of the conditional variance and covariances.

A third example in this class of objective functions is related to the estimation of the ratio $r(\boldsymbol{\theta}_1)/r(\boldsymbol{\theta}_2)$. We obtain the stratified estimates, $\hat{x}_1$ and $\hat{x}_2$, using a single simulation. Then, we can estimate the ratio with $\hat{x}_1/\hat{x}_2$ which has bias:

$$Bias\left[\hat{x}_1/\hat{x}_2\right] = \hat{x}_1\hat{x}_2^{-3}\Sigma_{22}(\boldsymbol{\pi}) - \hat{x}_2^{-2}\Sigma_{12}(\boldsymbol{\pi}) + O\left(N^{-2}\right),$$

see e.g., Fishman (1996), page 109. Here, the leading term is of order $O\left(N^{-1}\right)$. It is possible to reduce the bias by subtracting the estimate of the leading term from the ratio estimate. However, the squared bias is of order $O\left(N^{-2}\right)$, which is small compared to the variance. Thus, we rather consider reducing the variance that is approximated by:

$$V\left[\hat{x}_1/\hat{x}_2\right] \approx \hat{x}_1^2\hat{x}_2^{-4}\Sigma_{22}(\boldsymbol{\pi}) - 2\hat{x}_1\hat{x}_2^{-3}\Sigma_{12}(\boldsymbol{\pi}) + \hat{x}_2^{-2}\Sigma_{11}(\boldsymbol{\pi}) = \omega(\boldsymbol{\pi}). \tag{8}$$

The approximate variance in (8) is of order $O\left(N^{-1}\right)$ and found by using the multivariate Taylor series expansion of the variance of the ratio (see e.g., Glasserman 2004). The second equation in (8) defines an objective function that is a linear function of the variance-covariance matrix. It is easy to show that it satisfies the sufficient condition for convexity stated in Theorem 1. Therefore, we can find optimal allocation fractions using (7):

$$\pi_i^* = p_i\left(\hat{x}_2^{-4}\hat{x}_1^2\hat{s}_{i2}^2 - 2\hat{x}_1\hat{x}_2^{-3}\hat{s}_i^{12} + \hat{x}_2^{-2}\hat{s}_{i1}^2\right)^{1/2} \Bigg/ \sum_{l=1}^{I} p_l\left(\hat{x}_2^{-4}\hat{x}_1^2\hat{s}_{l2}^2 - 2\hat{x}_1\hat{x}_2^{-3}\hat{s}_l^{12} + \hat{x}_2^{-2}\hat{s}_{l1}^2\right)^{1/2}, \quad i = 1,\ldots,I, \quad (9)$$

where the conditional sample variances, $\hat{s}_{i1}^2$ and $\hat{s}_{i2}^2$, the conditional sample covariances, $\hat{s}_i^{12}$, and the estimates, $\hat{x}_1$ and $\hat{x}_2$, are found using the pilot sample.

Note that the allocation fractions given in (9) minimize the variance of a single ratio estimate. We can also use the multiresponse stratified sampling algorithm for minimizing the overall error of multiple ratio estimates. In fact, the third example given in Section 4 demonstrates the application of this idea.

3.2 Minimizing the Maximum of the Variances Weighted with Non-negative Coefficients

Now, we consider minimizing $\omega(\boldsymbol{\pi}) = \max\left\{j : c_j\Sigma_{jj}(\boldsymbol{\pi})\right\}$ for non-negative c_j values. We use the short notation $\omega_j(\boldsymbol{\pi})$ for functions $c_j\Sigma_{jj}(\boldsymbol{\pi})$, $j = 1,\ldots,J$. The objective function is non-differentiable at some points in the solution set; thus, the model in (5) is replaced by:

$$\begin{array}{ll} \min & z \\ \text{s.t.} & z - \omega_j(\boldsymbol{\pi}) \geq 0, \quad j = 1,\ldots,J, \quad \sum_{i=1}^{I}\pi_i = 1, \quad \text{and} \quad \pi_i > 0, \quad i = 1,\ldots,I. \end{array} \quad (10)$$

For the model in (10), Başoğlu, Hörmann, and Sak (2013) propose a simple allocation heuristic which yields a satisfactory sub-optimal solution in a short period of time. The heuristic method calculates the respective optimum allocation fractions $\boldsymbol{\pi}^j \in \mathbb{R}^I$ which minimize the variances of the stratified estimates, $\boldsymbol{\Sigma}_{jj}(\boldsymbol{\pi})$, $j = 1,\ldots,J$:

$$\pi_i^j = p_i\hat{s}_{ij} \Big/ \sum_{l=1}^{I} p_l\hat{s}_{lj}, \quad i = 1,\ldots,I$$

and searches for the best solution in the convex hull of these points. In other words, the allocation heuristic searches the optimal solution of the following model:

$$\begin{array}{ll} \min & z \\ \text{s.t.} & z - \omega_j(\boldsymbol{\pi}) \geq 0, \quad j = 1,\ldots,J, \quad \pi_i - \sum_{j=1}^{J}\lambda_j\pi_i^j = 0, \quad i = 1,\ldots,I, \\ & \sum_{j=1}^{J}\lambda_j = 1, \quad \text{and} \quad \lambda_j \geq 0, \quad j = 1,\ldots,J. \end{array} \quad (11)$$

Thus, the equality and the non-negativity constraints of the model in (10) are automatically satisfied and the dimension of the problem is clearly reduced for practically relevant settings. The numerical results of Başoğlu, Hörmann, and Sak (2013) show that the optimal solution of the model in (11) is either equal or close to the optimal solution of (10).

The main idea of the heuristic method is as follows: For every point in the convex hull, the objective value $\omega(\boldsymbol{\pi})$ is attained by one of the convex functions, say, $\omega_j(\boldsymbol{\pi})$. Then, we expect the objective value to decrease if we move towards the respective optimum solution $\boldsymbol{\pi}^j$. We stop moving towards $\boldsymbol{\pi}^j$ if we reach a point for which the objective value is attained by another function $\omega_k(\boldsymbol{\pi})$, $k \neq j$; then, we can move towards $\boldsymbol{\pi}^k$. In summary, for every point in the convex hull, the heuristic automatically determines a descent direction at the function evaluation. The size of the move is determined by the iteration number.

Now, we continue with a more-detailed description of the heuristic algorithm. We direct the reader to Başoğlu, Hörmann, and Sak (2013) for the pseudocode and the numerical performance results of the allocation heuristic.

Heuristic Algorithm for Minimizing $\omega(\boldsymbol{\pi}) = \max\left\{j : c_j \Sigma_{jj}(\boldsymbol{\pi})\right\}$: Assume a feasible solution $\boldsymbol{\pi}$ in the convex hull of $\boldsymbol{\pi}^j$, $j = 1,\ldots,J$. The objective value at $\boldsymbol{\pi}$ is $\omega(\boldsymbol{\pi}) = \max\left\{j : \omega_j(\boldsymbol{\pi})\right\}$, and the index of the function which attains this objective value is obtained by $j(\boldsymbol{\pi}) = \arg\max\left\{j : \omega_j(\boldsymbol{\pi})\right\}$. Assume further that ω is differentiable at $\boldsymbol{\pi}$. Then, the two functions, ω and $\omega_{j(\boldsymbol{\pi})}$ are equal in an open neighborhood of $\boldsymbol{\pi}$, and $\omega_{j(\boldsymbol{\pi})}$ decreases as we move towards $\boldsymbol{\pi}^j$. Hence, $\boldsymbol{\pi}^j - \boldsymbol{\pi}$ forms a simple descent direction for ω at $\boldsymbol{\pi}$.

The allocation heuristic algorithm starts with an initial feasible solution in the convex hull of $\boldsymbol{\pi}^j$, $j = 1,\ldots,J$. The weight center of the $\boldsymbol{\pi}^j$ is a simple candidate. The current and the best solution are denoted by $\boldsymbol{\pi}^c$ and $\boldsymbol{\pi}^h$, respectively. Therefore, we initialize $\boldsymbol{\pi}^h = \boldsymbol{\pi}^c = J^{-1}\sum_{i=1}^{J}\boldsymbol{\pi}^j$ and set the best and the current objective value $\omega^h = \omega^c = \omega(\boldsymbol{\pi}^c)$. We also determine the index of the function which attains the current objective value, $j^c = j(\boldsymbol{\pi}^c)$.

We expect an improvement in the objective value if we make a move towards $\boldsymbol{\pi}^{j^c}$. Let η denote the step number of the current iteration as the algorithm is run. When the algorithm is started, η is initialized to 1. We update the current solution according to the recursion $\boldsymbol{\pi}^c = \left(\eta\boldsymbol{\pi}^c + \boldsymbol{\pi}^{j^c}\right)/(\eta+1)$. As η increases, the current solution $\boldsymbol{\pi}^c$ and, since ω is continuous, the current objective value ω^c converge.

After each move, we update the current objective value ω^c and set $j' = j(\boldsymbol{\pi}^c)$ as the index of the function which attains this value. At this point, we sequentially check the following two conditions:

- If $\omega^c \le \omega^h$, we update the best solution $\boldsymbol{\pi}^h = \boldsymbol{\pi}^c$ and the best objective value $\omega^h = \omega^c$.
- If $j' \ne j^c$, then the current solution can be improved by moving in another direction. We set $j^c = j'$ and increase η by one.

Then, we return to the step where we update the current solution. The algorithm terminates when ω^c converges.

4 NUMERICAL EXAMPLES

To demonstrate the efficiency of multiresponse stratified sampling, we now present examples where we estimate multiple values in a single simulation. Multiresponse stratified sampling yields comparatively good results if stratification is also successful in the estimation of a single value, for example, the simulation function is of rare-event type or the random input has a small number of stratification variables that have a large contribution to the output variance. For this reason, we give examples that are of these types. Notice also that the examples here are new and not included in Başoğlu, Hörmann, and Sak (2013).

We remind that in all multiresponse stratified sampling algorithms the pilot samples use the same random sequence. For each example, we also run naive Monte Carlo simulation using common random numbers and find the variance of multiple estimates in a single simulation. The efficiency results are presented as variance reduction factors, $VRF[\hat{x}] = V\left[\hat{x}^{NV}\right]/V[\hat{x}]$ where $\hat{x}^{NV}$ and $\hat{x}$ denote the naive Monte Carlo and stratified estimators, respectively. The execution times of the methods are also reported.

Example 1: We begin with a toy example where the size of the random input is small. With this first example, we demonstrate the practical aspects of the optimization models given in Section 3. Therefore, we consider a simulation function where we can achieve successful variance reduction by applying stratification over a single direction.

$$q(\mathbf{Z},\boldsymbol{\theta}) = \min\left\{\max\left\{(Z_1+Z_2)^2 + \theta_1 Z_1, \theta_2\right\}, \theta_2+\theta_3\right\}. \tag{12}$$

Here, Z_1 and Z_2 are independent standard normal variables. Our aim is to estimate $x_j = r(\boldsymbol{\theta}_j)$, $j = 1,\ldots,6$, in a single simulation for $\boldsymbol{\theta}_j$ vectors:

$$\begin{aligned} &\boldsymbol{\theta}_1 = (.1,1.1,.722)', \quad \boldsymbol{\theta}_2 = (.1,1.2,.688)', \quad \boldsymbol{\theta}_3 = (.2,1.1,.291)', \\ &\boldsymbol{\theta}_4 = (.2,1.2,.342)', \quad \boldsymbol{\theta}_5 = (.3,1.1,.148)', \quad \boldsymbol{\theta}_6 = (.1,1.2,.192)'. \end{aligned} \tag{13}$$

We divide the sampling domain into 100 equiprobable subsets with planes that are orthogonal to the direction $(\sqrt{2}/2, \sqrt{2}/2)'$. In other words, we stratify the random variable $Z_1 + Z_2$ with $I = 100$ equiprobable strata. Increasing the number of strata promises for more variance reduction. However, the marginal contribution of adding a stratum eventually decreases and the variance reduction becomes saturated. Increasing the number of strata also decreases the number of drawings to be allocated in each stratum in the pilot study and may result in poor estimates of the conditional standard deviations. Thus, the choice for the number of strata depends on the simulation function q and the size of the pilot sample N_p. For all examples in this section, we fix $N_p = 10^5$ and we keep the number of strata between 100 and 300.

We run the multiresponse stratified sampling algorithm under the objective functions listed in Section 3 and also, for minimizing the variance of each single estimate, with the objective functions $\Sigma_{jj}(\boldsymbol{\pi})$, $j = 1, \ldots, 6$. This allows us to observe the minimal error reachable for each $\boldsymbol{\theta}_j$.

The variance and the percentage relative error of all estimates obtained under each of these objective functions are listed in Table 1. Variance reduction factors are also given for performance comparison. The total sample size used in all simulations is $N = 10^6$ and $N_p = 10^5$ of the drawings are used in a pilot study to determine the optimal allocation fractions via (7) or the allocation heuristic described in Section 3.2.

Table 1: The variance and the percentage relative error of all estimates obtained under different objective functions, for the simulation function in (12) and the parameter vectors given in (13).

		$\boldsymbol{\theta}_j$	$\boldsymbol{\theta}_1$	$\boldsymbol{\theta}_2$	$\boldsymbol{\theta}_3$	$\boldsymbol{\theta}_4$	$\boldsymbol{\theta}_5$	$\boldsymbol{\theta}_6$
Method	$\omega(\boldsymbol{\pi})$	$\sim x_j$	1.385	1.462	1.226	1.340	1.166	1.281
Naive	-	$V[\hat{x}_j]$	1.14E-07	1.03E-07	2.00E-08	2.71E-08	5.30E-09	8.78E-09
1.80 sec.	-	$\%RE[\hat{x}_j]$	±0.04785	±0.04307	±0.02262	±0.02410	±0.01224	±0.01434
Multiresponse	$\min \Sigma_{11}(\boldsymbol{\pi})$	$V[\hat{x}_j]$	1.08E-10	1.04E-10	2.26E-10	1.46E-10	5.90E-10	3.52E-10
Stratified	2.52 sec.	$\%RE[\hat{x}_j]$	±0.00147	±0.00137	±0.00240	±0.00177	±0.00409	±0.00287
Sampling		$VRF[\hat{x}_j]$	1059	989	89	186	9	25
	$\min \Sigma_{22}(\boldsymbol{\pi})$	$V[\hat{x}_j]$	1.35E-10	9.09E-11	5.44E-10	2.36E-10	1.01E-09	6.64E-10
	2.54 sec.	$\%RE[\hat{x}_j]$	±0.00164	±0.00128	±0.00373	±0.00225	±0.00534	±0.00394
		$VRF[\hat{x}_j]$	849	1136	37	115	5	13
	$\min \Sigma_{33}(\boldsymbol{\pi})$	$V[\hat{x}_j]$	7.39E-10	1.14E-09	1.07E-10	1.99E-10	1.50E-10	2.01E-10
	2.46 sec.	$\%RE[\hat{x}_j]$	±0.00385	±0.00453	±0.00166	±0.00207	±0.00206	±0.00217
		$VRF[\hat{x}_j]$	155	90	187	136	35	44
	$\min \Sigma_{44}(\boldsymbol{\pi})$	$V[\hat{x}_j]$	1.82E-10	2.47E-10	1.41E-10	1.20E-10	3.33E-10	2.11E-10
	2.52 sec.	$\%RE[\hat{x}_j]$	±0.00191	±0.00211	±0.00190	±0.00160	±0.00307	±0.00222
		$VRF[\hat{x}_j]$	629	418	142	226	16	42
	$\min \Sigma_{55}(\boldsymbol{\pi})$	$V[\hat{x}_j]$	8.10E-10	1.12E-09	1.31E-10	2.29E-10	8.78E-11	1.50E-10
	2.52 sec.	$\%RE[\hat{x}_j]$	±0.00403	±0.00449	±0.00183	±0.00222	±0.00158	±0.00187
		$VRF[\hat{x}_j]$	141	92	153	119	60	59
	$\min \Sigma_{66}(\boldsymbol{\pi})$	$V[\hat{x}_j]$	2.62E-10	3.50E-10	1.23E-10	1.49E-10	1.03E-10	1.20E-10
	2.51 sec.	$\%RE[\hat{x}_j]$	±0.00229	±0.00251	±0.00178	±0.00178	±0.00171	±0.00168
		$VRF[\hat{x}_j]$	437	295	162	182	52	73
Multiresponse	$\min \omega_{SUM}(\boldsymbol{\pi}) = \sum_{j=1}^{J}\sum_{k=1}^{J}\Sigma_{jk}(\boldsymbol{\pi})$	$V[\hat{x}_j]$	1.37E-10	1.49E-10	1.28E-10	1.34E-10	1.34E-10	1.33E-10
Stratified	2.50 sec.	$\%RE[\hat{x}_j]$	±0.00166	±0.00164	±0.00181	±0.00169	±0.00194	±0.00176
Sampling		$VRF[\hat{x}_j]$	835	692	156	203	40	66
	$\min \omega_{MSE}(\boldsymbol{\pi}) = \sum_{j=1}^{J}\Sigma_{jj}(\boldsymbol{\pi})$	$V[\hat{x}_j]$	1.32E-10	1.35E-10	1.35E-10	1.43E-10	1.22E-10	1.32E-10
	2.47 sec.	$\%RE[\hat{x}_j]$	±0.00163	±0.00156	±0.00186	±0.00175	±0.00186	±0.00176
		$VRF[\hat{x}_j]$	864	763	148	190	43	67
	$\min \omega_{MSR}(\boldsymbol{\pi}) = \sum_{j=1}^{J}\hat{x}_j^{-2}\Sigma_{jj}(\boldsymbol{\pi})$	$V[\hat{x}_j]$	1.37E-10	1.42E-10	1.32E-10	1.44E-10	1.18E-10	1.30E-10
	2.53 sec.	$\%RE[\hat{x}_j]$	±0.00165	±0.00160	±0.00184	±0.00176	±0.00182	±0.00174
		$VRF[\hat{x}_j]$	837	726	151	188	45	68
	$\min \omega_{MAXE}(\boldsymbol{\pi}) = \max\{j : \Sigma_{jj}(\boldsymbol{\pi})\}$	$V[\hat{x}_j]$	1.31E-10	1.33E-10	1.36E-10	1.38E-10	1.31E-10	1.35E-10
	2.12 sec.	$\%RE[\hat{x}_j]$	±0.00162	±0.00155	±0.00187	±0.00172	±0.00192	±0.00178
		$VRF[\hat{x}_j]$	874	776	147	197	41	65
	$\min \omega_{MAXR}(\boldsymbol{\pi}) = \max\{j : \hat{x}_j^{-2}\Sigma_{jj}(\boldsymbol{\pi})\}$	$V[\hat{x}_j]$	1.54E-10	1.69E-10	1.23E-10	1.46E-10	1.11E-10	1.26E-10
	2.14 sec.	$\%RE[\hat{x}_j]$	±0.00176	±0.00174	±0.00177	±0.00177	±0.00177	±0.00172
		$VRF[\hat{x}_j]$	741	611	162	185	48	70

If we minimize the variance of a single estimate rather than focusing on the overall error, we observe in Table 1 that the variance for that estimate is reduced, however, we also see severe losses in the variance reduction factors for most of the other estimates. The last five objective functions aim to minimize the overall error of the simulation and, in those cases, we observe reasonably good variance reduction with only moderate losses compared to the maximum variance reduction factors. For the last two objective functions, we utilize the allocation heuristic to determine sub-optimal allocation fractions. The objective function ω_{MAXE} leads to very close variances. A similar result is also valid for relative error values under the objective function ω_{MAXR}.

The calculated variance and the relative error values seem to be very close under the respective objective functions. Therefore, we conclude that the allocation heuristic successfully determines the allocation fractions which minimize the maximum variance and the relative error, respectively.

Example 2: We consider a practically more relevant response function; the risk-neutral price of an Asian option with a maturity of one year. The underlying stock price follows a geometric Brownian motion in a risk-neutral environment. We monitor the stock prices at discrete time points and the average stock price is calculated using:

$$\bar{S} = S_0 \boldsymbol{w}' \exp\left\{ \left(\rho - 0.5\sigma^2\right) \boldsymbol{t} + \sigma \boldsymbol{L}\boldsymbol{Z} \right\},$$

where $\boldsymbol{Z} \in \mathbb{R}^D$ denotes a standard multinormal vector with independent elements, $\boldsymbol{t} \in \mathbb{R}^D$ is the vector that holds monitored control points in year units, $\boldsymbol{L}$ is the lower-triangular Cholesky factorization of matrix $\boldsymbol{A} \in \mathbb{R}^{D\times D}$ with elements $A_{ij} = \min\left\{t_i, t_j\right\}$, $\boldsymbol{w} \in \mathbb{R}^D$ holds the weights of the control points, σ denotes the yearly volatility, ρ the risk-free interest rate, and S_0 the initial stock price.

In our experiments, we consider call options with discounted payoff function $q(\boldsymbol{Z}) = e^{-\rho t_D}\left(\bar{S} - K\right)^+$ where K denotes the strike price. The control points are monthly and equally weighted, thus, $D = 12$ and $w_d = 1/12$, $d = 1, \ldots, 12$. We assume $S_0 = 100$, $K = 110$ and $J = 4$ distinct $\boldsymbol{\theta} = (\rho, \sigma)'$ parameter vectors:

$$\boldsymbol{\theta}_1 = (.05, .1)', \quad \boldsymbol{\theta}_2 = (.05, .2)', \quad \boldsymbol{\theta}_3 = (.02, .1)', \quad \boldsymbol{\theta}_4 = (.02, .2)'. \tag{14}$$

Our aim is to estimate the Asian call prices for each $\boldsymbol{\theta}_j$ in a single simulation. We stratify the linear projection $\boldsymbol{v}'\boldsymbol{Z}$ where $\boldsymbol{v} \in \mathbb{R}^D$ is called the stratification direction with the property $\|\boldsymbol{v}\| = 1$ (see, e.g., Glasserman (2004) for stratification of linear projections). We follow the idea of Jourdain, Lapeyre, and Sabino (2011) and choose this direction proportional to the gradient of the average stock price (see also Imai and Tan 2006):

$$v_d \propto \left. \frac{\partial \bar{S}}{\partial Z_d} \right|_{\boldsymbol{Z}=0}, \quad d = 1, \ldots, D.$$

However, since $\bar{S}$ is a function of $\boldsymbol{\theta}$ and we consider many $\boldsymbol{\theta}$ values, we calculate the gradient at $\boldsymbol{\theta}^*$, an arbitrary parameter vector chosen in the convex hull of the continuous parameters. In our experiment, we have chosen $\boldsymbol{\theta}^* = (.035, .15)'$.

The number of equiprobable strata used in this experiment is $I = 200$. Similar to our previous example, we run the multiresponse stratified sampling algorithm under the objective functions listed in Section 3. For each objective function, the percentage relative error of all estimates are listed in Table 2. Variance reduction factors are also given for performance comparison. The total sample size used in all simulations is $N = 10^6$ and $N_p = 10^5$ of the drawings are used in a pilot study to determine the optimal allocation fractions via (7) or the allocation heuristic described in Section 3.2.

In Table 2, we observe that the variance reduction factors obtained under the objective functions ω_{SUM}, ω_{MSE}, and ω_{MSR} are acceptable with a moderate loss compared to the maximum variance reduction factors. We also observe that the variance reduction factors obtained under the objective function that minimizes the maximum variance, ω_{MAXE}, is the same as the variance reduction factors obtained under allocation fractions that minimize the variance of the 2nd estimate, $\Sigma_{22}(\boldsymbol{\pi})$. In other words, the function $\Sigma_{22}(\boldsymbol{\pi})$ dominates the other variance functions in the neighborhood of the heuristic solution. This is expected, since the largest variance for the discounted payoffs in each stratum is obtained under the largest ρ and σ values. On the other hand, if our objective function is ω_{MAXR}, we get almost the same relative error values for all estimates.

Table 2: The variance and the percentage relative error of all estimates obtained under different objective functions for the discounted payoff of Asian call option and the parameter vectors given in (14).

	$\omega(\boldsymbol{\pi})$	$\boldsymbol{\theta}_j$ $\sim x_j$	$\boldsymbol{\theta}_1$ 0.43	$\boldsymbol{\theta}_2$ 2.29	$\boldsymbol{\theta}_3$ 0.25	$\boldsymbol{\theta}_4$ 1.88
Naive	-	$V[\hat{x}_j]$	2.38E-06	2.92E-05	1.37E-06	2.42E-05
9.75 sec.	-	$\%RE[\hat{x}_j]$	±0.70365	±0.46319	±0.90778	±0.51443
Multiresponse	$\min \Sigma_{11}(\boldsymbol{\pi})$	$\%RE[\hat{x}_j]$	±0.01485	±0.02650	±0.02261	±0.02609
Stratified	12.16 sec.	$VRF[\hat{x}_j]$	2236	305	1604	388
Sampling	$\min \Sigma_{22}(\boldsymbol{\pi})$	$\%RE[\hat{x}_j]$	±0.01767	±0.01153	±0.02756	±0.01285
	12.03 sec.	$VRF[\hat{x}_j]$	1579	1609	1080	1599
	$\min \Sigma_{33}(\boldsymbol{\pi})$	$\%RE[\hat{x}_j]$	±0.03114	±0.03324	±0.02087	±0.03450
	12.06 sec.	$VRF[\hat{x}_j]$	508	194	1883	222
	$\min \Sigma_{44}(\boldsymbol{\pi})$	$\%RE[\hat{x}_j]$	±0.01683	±0.01395	±0.02611	±0.01239
	11.95 sec.	$VRF[\hat{x}_j]$	1741	1099	1203	1721
Multiresponse	$\min \omega_{SUM}(\boldsymbol{\pi}) = \sum_{j=1}^{J}\sum_{k=1}^{J}\Sigma_{jk}(\boldsymbol{\pi})$	$\%RE[\hat{x}_j]$	±0.01663	±0.01184	±0.02553	±0.01263
Stratified	12.07 sec.	$VRF[\hat{x}_j]$	1783	1527	1259	1656
Sampling	$\min \omega_{MSE}(\boldsymbol{\pi}) = \sum_{j=1}^{J}\Sigma_{jj}(\boldsymbol{\pi})$	$\%RE[\hat{x}_j]$	±0.01723	±0.01157	±0.02672	±0.01269
	11.95 sec.	$VRF[\hat{x}_j]$	1660	1597	1149	1638
	$\min \omega_{MSR}(\boldsymbol{\pi}) = \sum_{j=1}^{J}\hat{x}_j^{-2}\Sigma_{jj}(\boldsymbol{\pi})$	$\%RE[\hat{x}_j]$	±0.01603	±0.01280	±0.02359	±0.01339
	11.95 sec.	$VRF[\hat{x}_j]$	1918	1307	1474	1473
	$\min \omega_{MAXE}(\boldsymbol{\pi}) = \max\{j : \Sigma_{jj}(\boldsymbol{\pi})\}$	$\%RE[\hat{x}_j]$	±0.01767	±0.01153	±0.02756	±0.01285
	11.58 sec.	$VRF[\hat{x}_j]$	1579	1609	1080	1599
	$\min \omega_{MAXR}(\boldsymbol{\pi}) = \max\{j : \hat{x}_j^{-2}\Sigma_{jj}(\boldsymbol{\pi})\}$	$\%RE[\hat{x}_j]$	±0.01936	±0.01888	±0.02144	±0.01935
	11.66 sec.	$VRF[\hat{x}_j]$	1316	601	1784	705

Example 3: We aim to estimate the risk of a linear portfolio of $D = 5$ stocks for a fixed time horizon. We assume that the stock log-returns follow a multinormal distribution with mean vector, $\boldsymbol{\mu} = (0,0,0,0,0)'$, volatility vector, $\boldsymbol{\sigma} = (.15,.175,.2,.225,.25)'$, and correlation matrix $\boldsymbol{R}$ with all correlations equal to 0.3. Then, the portfolio return is:

$$Return = \sum_{d=1}^{D} w_d \exp\{\mu_d + \sigma_d \sum_{k=1}^{D} L_{dk} Z_k\},$$

where $\boldsymbol{w} = (w_1,\ldots,w_D)'$ holds the fraction of investments in each stock and $\boldsymbol{L}$ is the lower-triangular Cholesky factorization of the correlation matrix $\boldsymbol{R}$. The loss of the portfolio is calculated as $Loss = S_0(1 - Return)$ where S_0 is the initial amount of investment. For given threshold values τ_j, $j = 1,\ldots,J$, our aim is to estimate the conditional excess values $x_j = E[Loss|Loss > \tau_j]$. For this purpose, we define the indicator function:

$$\mathbb{1}\{Loss > \tau_j\} = \begin{cases} 1, & Loss > \tau_j, \\ 0, & Loss \leq \tau_j, \end{cases}$$

and use the ratio

$$x_j = \frac{x'_j}{x''_j} = \frac{E[Loss\,\mathbb{1}\{Loss > \tau_j\}]}{E[\mathbb{1}\{Loss > \tau_j\}]}, \quad j = 1,\ldots,J,$$

where $x''_j = E[\mathbb{1}\{Loss > \tau_j\}] = \Pr\{Loss > \tau_j\}$ is called tail loss probability, another risk measure.

In our experiments, we take $S_0 = 1$ and consider $J = 10$ equidistant threshold values. Our aim is to estimate x_j for $j = 1,\ldots,J$ in a single simulation. The algorithm mainly estimates x'_j and x''_j for $j = 1,\ldots,J$, and the variance of the j-th ratio estimator $\hat{x}_j = \hat{x}'_j/\hat{x}''_j$ is calculated using (8), by plugging in the variances and the covariance of the stratified estimators $\hat{x}'_j$ and $\hat{x}''_j$.

Similar to our second example, we stratify the linear projection $\boldsymbol{v}'\boldsymbol{Z}$ where the stratification direction $\boldsymbol{v} \in \mathbb{R}^D$ is proportional to the gradient of the loss function at $\boldsymbol{Z} = 0$. The number of equiprobable strata used in this experiment is $I = 300$. We run the multiresponse stratified sampling algorithm under two objective functions. In the first one we aim to minimize the mean squared relative error and, in the second one, we aim to minimize the maximum relative error of all ratio estimates. Thus, the variances, $\Sigma_{jj}(\boldsymbol{\pi})$, used in these objective functions should correspond to the ratio estimates, $\hat{x}_j,\ j = 1,\ldots,J$.

The variance and the percentage relative error of all estimates obtained under these two objective functions are listed in Table 3. Variance reduction factors are also given for performance comparison. The total sample size used in all simulations is $N = 10^6$ and $N_p = 10^5$ of the drawings are used in a pilot study to determine the optimal allocation fractions via (7) or the allocation heuristic described in Section 3.2. Here, the allocation heuristic searches the optimal allocation fractions in the convex hull of $\boldsymbol{\pi}^j,\ j = 1,\ldots,J$, the respective allocation fractions minimizing the variance of each ratio estimate, $\Sigma_{jj}(\boldsymbol{\pi})$ that can be found by using (9).

Table 3: The variance and the percentage relative error of conditional excess estimates obtained under different objective functions for different threshold values. The execution times for the three methods are 17.6, 24.1, and 24.5 seconds respectively.

			Naive		Str. Sampl.: $\min \omega_{MSR}(\boldsymbol{\pi})$			Str. Sampl.: $\min \omega_{MAXR}(\boldsymbol{\pi})$		
τ_j	$\sim x''_j$	$\sim x_j$	$V[\hat{x}_j]$	$\%RE[\hat{x}_j]$	$V[\hat{x}_j]$	$\%RE[\hat{x}_j]$	$VRF[\hat{x}_j]$	$V[\hat{x}_j]$	$\%RE[\hat{x}_j]$	$VRF[\hat{x}_j]$
0.147	0.100	0.198	1.75E-08	±0.131	3.99E-10	±0.020	44	4.11E-10	±0.020	43
0.159	0.083	0.207	1.96E-08	±0.132	3.70E-10	±0.018	53	4.24E-10	±0.019	46
0.171	0.067	0.217	2.22E-08	±0.134	3.85E-10	±0.018	58	4.74E-10	±0.020	47
0.184	0.054	0.227	2.53E-08	±0.137	4.06E-10	±0.017	62	4.99E-10	±0.019	51
0.196	0.042	0.237	2.94E-08	±0.142	4.34E-10	±0.017	68	5.62E-10	±0.020	52
0.208	0.033	0.247	3.45E-08	±0.147	4.67E-10	±0.017	74	5.83E-10	±0.019	59
0.220	0.025	0.257	4.12E-08	±0.154	5.30E-10	±0.018	78	6.40E-10	±0.019	64
0.233	0.019	0.268	4.97E-08	±0.163	6.36E-10	±0.018	78	7.00E-10	±0.019	71
0.245	0.014	0.278	6.15E-08	±0.175	7.90E-10	±0.020	78	7.85E-10	±0.020	78
0.257	0.010	0.289	7.78E-08	±0.189	1.15E-09	±0.023	68	8.15E-10	±0.019	96

According to the results in Table 3, the variance of all conditional excess estimates are successfully reduced under both objectives. If we choose to minimize the sum of squared relative errors, we can easily determine optimal allocation fractions using (7). However, we can further decrease the maximum relative error using the allocation heuristic and obtain similar relative errors, as can be seen in the last two columns of Table 3.

For the estimation of multiple conditional excess values, we have repeated the above experiment also with the objective functions ω_{MSE} and ω_{MAXE}, which focus on the absolute errors of all estimates. In Figure 1 (a), we give the logarithms of absolute errors of all conditional excess estimates obtained with naive simulation and with multiresponse stratified sampling under different objective functions.

According to Figure 1 (a), if we use optimal allocation fractions which minimize the variance of the conditional excess estimate for the 5th threshold value, we are not at all able to reduce the error of estimates for smaller threshold levels. However, by using the optimal allocation fractions which minimize the variance of the estimate that corresponds to the smallest threshold, we observe a fair decrease in the error of all estimates. This is similar to the method suggested by Glasserman and Li (2005), where they utilize importance sampling to estimate tail loss probabilities of a credit portfolio in a single simulation for several thresholds and use the optimal importance sampling parameter that corresponds to the smallest of the thresholds. We also observe that we can obtain better results with multiresponse stratified sampling under more global objective functions. The objective function ω_{MAXE} results in absolute errors almost at a constant level. Compared to that behavior, the objective function ω_{MSE} reduces the relative errors for smaller thresholds and increases the relative error for the largest threshold.

For the same problem, we also consider the estimation of tail loss probabilities. Since the loss probability converges to zero as the threshold increases, we aim to reduce the relative error rather than the absolute error. Thus, we run the multiresponse stratified sampling algorithm under objective functions ω_{MSR} and

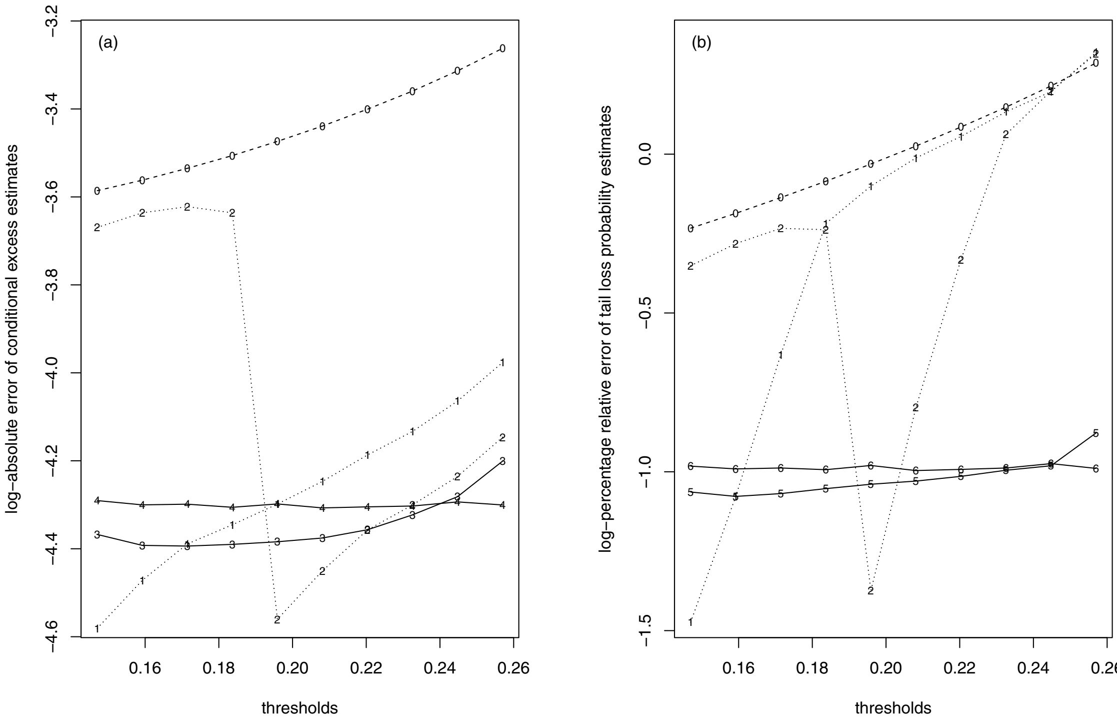

Figure 1: (a) The log-absolute error of all conditional excess estimates obtained with a single simulation. The dashed line shows naive simulation (0) results whereas the dotted lines correspond to stratified sampling minimizing the variances of the first (1) and the 5th (2) estimates. The remaining plots correspond to multiresponse stratified sampling with the objective functions (3) ω_{MSE} and (4) ω_{MAXE}. In (b), the log-percentage relative error of tail loss probability estimates is shown. To reduce the overall relative error in tail loss probability estimates, the objective functions (5) ω_{MSR}, and (6) ω_{MAXR} are considered.

ω_{MAXR}. Figure 1 (b) shows the logarithms of percentage relative errors of all tail loss probability estimates obtained with naive simulation and with multiresponse stratified sampling under these objective functions.

According to Figure 1 (b), for reducing the overall error, it is again not a good idea to use the optimal allocations which minimize the variance of a specific tail loss probability estimate. In this case, optimizing for the smallest threshold leads to worse results than in Figure 1 (a). However, we can obtain very good results with multiresponse stratified sampling. The objective function ω_{MAXR} results in relative errors almost at a constant level and the objective function ω_{MSE} again reduces the errors for small thresholds but increases the error for the largest threshold.

With these examples, we have shown that multiresponse stratified sampling is an efficient method for simulation problems for which we can find efficient stratification variables. We remind that in our examples, the size of the random input is independent of the parameter space. Whether comparable results would be obtainable for all type of discrete event simulations is an area in need of further research.

5 CONCLUSIONS

We have utilized multiresponse stratified sampling for the estimation of multiple values in a single Monte Carlo simulation. In order to increase the efficiency of all estimators, we have proposed two general objective functions. In the first, we consider minimizing linear functions of the variance-covariance matrix of the stratified estimates. In the second class, we minimize the maximum of the variances which are weighted with non-negative coefficients. For these objective functions, we have introduced nonlinear optimization models with allocation fractions as decision variables. We have developed the closed-form solution for the first class of objective functions. For the second class, we have used the allocation heuristic of Başoğlu, Hörmann, and Sak (2013) to develop a sub-optimal solution. These solutions are used in the sampling phase. Variance reduction results indicate that multiresponse stratified sampling is an efficient and flexible method for estimating multiple values in a single simulation. We are convinced that this new methodology can also be useful for stochastic optimization and response surface estimation problems, as they require the evaluation of the simulation function for many different parameter values.

The R-package `riskSimul` available on `http://cran.r-project.org/` applies our new methods to the risk quantification of linear stock portfolios under the t-copula model.

ACKNOWLEDGMENTS

This work was partially supported by Boğaziçi Scientific Research Fund Project 6923.

REFERENCES

Başoğlu, İ., W. Hörmann, and H. Sak. 2013. "Optimally Stratified Importance Sampling for Portfolio Risk with Multiple Loss Thresholds". *Optimization* 62 (11): 1451–1471.

Box, G., and K. Wilson. 1951. "On the Experimental Attainment of Optimum Conditions". *Journal of the Royal Statistical Society* Series B13 (1): 1–38.

Byrd, R. H., J. Nocedal, and R. A. Waltz. 2006. "Knitro: An Integrated Package for Nonlinear Optimization". In *Large-Scale Nonlinear Optimization*, edited by G. Pillo and M. Roma, Volume 83 of *Nonconvex Optimization and Its Applications*, 35–59. Springer US.

Etoré, P., and B. Jourdain. 2010. "Adaptive Optimal Allocation in Stratified Sampling Methods". *Methodology and Computing in Applied Probability* 12 (3): 335–360.

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

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. New York: Springer-Verlag.

Glasserman, P., and J. Li. 2005. "Importance Sampling for Portfolio Credit Risk". *Management Science* 51 (11): 1643–1656.

Imai, J., and K. S. Tan. 2006. "A General Dimension Reduction Technique for Derivative Pricing". *Journal of Computational Finance* 10 (2): 129–155.

Jourdain, B., B. Lapeyre, and P. Sabino. 2011. "Convenient Multiple Directions of Stratification". *International Journal of Theoretical and Applied Finance* 14 (6): 867–897.

Law, A. M. 2014. *Simulation Modeling & Analysis*. New York: McGraw-Hill.

Myers, R., and D. Montgomery. 2002. *Response Surface Methodology*. New York: Wiley.

Rubinstein, R., and R. Marcus. 1985. "Efficiency of Multivariate Control Variates in Monte Carlo Simulation". *Operations Research* 33 (3): 661–677.

Sak, H., and W. Hörmann. 2012. "Fast Simulations in Credit Risk". *Quantitative Finance* 12 (10): 1557–1569.

AUTHOR BIOGRAPHIES

İSMAİL BAŞOĞLU holds a Ph.D. degree in Industrial Engineering from Boğaziçi University. His research deals with Monte Carlo simulation and variance reduction techniques. His email address is ismailbsgl@gmail.com.

WOLFGANG HÖRMANN is an associate professor in the Industrial Engineering Department at Boğaziçi University. His research centers on non-uniform random variate generation, variance reduction techniques, and Monte Carlo methods in finance. His email address is hormannw@boun.edu.tr.