

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Vihola, Matti

Title: Unbiased Estimators and Multilevel Monte Carlo

Year: 2018

Version: Accepted version (Final draft)

Copyright: © 2017 INFORMS

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Vihola, M. (2018). Unbiased Estimators and Multilevel Monte Carlo. *Operations Research*, 66(2), 448-462. <https://doi.org/10.1287/opre.2017.1670>

UNBIASED ESTIMATORS AND MULTILEVEL MONTE CARLO

MATTI VIHOLA

ABSTRACT. Multilevel Monte Carlo (MLMC) and unbiased estimators recently proposed by McLeish (*Monte Carlo Methods Appl.*, 2011) and Rhee and Glynn (*Oper. Res.*, 2015) are closely related. This connection is elaborated by presenting a new general class of unbiased estimators, which admits previous debiasing schemes as special cases. New lower variance estimators are proposed, which are stratified versions of earlier unbiased schemes. Under general conditions, essentially when MLMC admits the canonical square root Monte Carlo error rate, the proposed new schemes are shown to be asymptotically as efficient as MLMC, both in terms of variance and cost. The experiments demonstrate that the variance reduction provided by the new schemes can be substantial.

1. INTRODUCTION

Multilevel Monte Carlo (MLMC) methods pioneered by Heinrich (2001) and Giles (2008) are now standard for estimation of expectations of functionals of processes defined by stochastic differential equations (SDEs). While the MLMC techniques have origins in the integral operators (Heinrich, 2001) and SDEs (Giles, 2008), they can be applied also in other application domains, where estimates with gradually increasing accuracy are available; see the recent review by Giles (2015) and references therein.

More recently, the so-called debiasing techniques (McLeish, 2011; Rhee and Glynn, 2012, 2015) have attracted a lot of research activity (e.g. Agapiou et al., 2014; Glynn and Rhee, 2014; Jacob and Thiery, 2015; Lyne et al., 2015; Strathmann et al., 2015; Walter, 2017), although similar ideas have been suggested much earlier in more specific contexts (e.g. Glynn, 1983; Rychlik, 1990, 1995). These techniques are based on similar ideas as MLMC, but instead of optimal allocation of computational resources for minimising the error, the primary focus is on providing unbiased estimators. Monte Carlo inference is straightforward with independent unbiased samples, allowing to construct confidence intervals in a reliable way (Glynn and Whitt, 1992b). Debiasing techniques may also be employed within a stochastic approximation algorithm (Dereich and Mueller-Gronbach, 2015). In particular, in a stochastic gradient descent type algorithm (Robbins and Monro, 1951) relevant for instance in maximum likelihood inference (Delyon et al., 1999), unbiased gradient estimate implies pure martingale noise, which is supported by a well-established theory (e.g. Borkar, 2008; Kushner and Yin, 2003).

The debiasing techniques involve balancing with cost and variance, which often boils down to similar methods and conditions as those that are used with MLMC. The connection between MLMC and debiasing techniques has been pointed out earlier at least by Rhee and Glynn (2015), Giles (2015) and Dereich and Mueller-Gronbach (2015), but this connection has not been fully explored yet. The purpose of this paper is to further clarify the connection of MLMC and debiasing techniques, within a general framework for unbiased estimators.

2010 *Mathematics Subject Classification.* Primary 65C05; secondary 65C30.

Key words and phrases. Efficiency, multilevel Monte Carlo, stochastic differential equation, stratification, unbiased.

Many techniques for unbiased estimation other than those considered here have been suggested in the literature. For instance, there is a whole body of literature for ‘perfect sampling’ by Markov chains (Propp and Wilson, 1996) or with certain classes of SDE models (Beskos and Roberts, 2005); see also the recent monograph by Huber (2015) and references therein. Perfect sampling can be used to construct unbiased estimators, but the problem is generally more prestigious and often harder to implement. See also the recent article by Jacob and Thiery (2015) for discussion about other related unbiased estimation techniques.

The rest of the paper is organised as follows. The multilevel Monte Carlo and the previous debiasing methods are presented briefly in Sections 2 and 3, respectively. Section 4 introduces a new general unbiased scheme with an explicit expression for its variance (Theorem 7). The unbiased estimators suggested by McLeish (2011) and Rhee and Glynn (2012, 2015) are reformulated as specific instances of this scheme, as well as an obvious ‘hybrid’ scheme with MLMC and unbiased components (Example 11).

New unbiased estimators are suggested in Section 5. Two of the new schemes, termed stratified and residual sampling estimators, have provably lower variance than simple averages of independent unbiased estimators (Proposition 16). Because stratification is a well-known variance reduction technique, these estimators may be well-known, but they do not seem to be recorded in the literature yet. The first main finding of this paper is Theorem 19 which shows that the variances of two of the new schemes are asymptotically equal to that of MLMC under general conditions. This result suggests that unbiasedness can often be achieved with asymptotically negligible additional variance.

The expected cost of the methods is discussed in Section 6. The new schemes appear even more appealing after seeing that the expected cost of MLMC and the unbiased schemes are also asymptotically equivalent (Proposition 22) and therefore the efficiency of an estimator with new schemes can be made arbitrarily close to MLMC (Corollary 23). The limiting variance formulation in Theorem 19 leads into an easily applicable optimisation criteria for the sampling distribution related to the new estimators.

Section 7 presents a generalisation of the unbiased scheme, which accommodates further conditioning and dependent randomisation schemes based on stopping times. Numerical experiments in Section 8 show how the efficiency bounds predicted by theory are attained in four examples, three of which were also studied by Rhee and Glynn (2015). The paper is concluded by a discussion about the implications of the findings in Section 9. Some practical guidelines and possible topics of further research are discussed as well.

We denote $a \wedge b := \min\{a, b\}$, $a \vee b := \max\{a, b\}$, $(x)_+ := 0 \vee x$ and $\mathbb{I}\{\cdot\}$ stands for the indicator function. The natural numbers are defined strictly positive $\mathbb{N} := \{1, 2, \dots\}$, empty sums $\sum_{j=1}^0(\cdot)$ are taken as zero, and we use the convention $0/0 = 0$.

2. MULTILEVEL MONTE CARLO

The starting point of MLMC is the existence of estimators $(Y_i)_{i \geq 1}$, which are increasingly accurate approximations of a given ‘target’ random variable Y , whose expectation is of interest. That is, $(Y_i)_{i \geq 1}$ are random variables with $\mathbb{E}Y_i \rightarrow \mathbb{E}Y$, and the task is to provide a numerical approximation of $\mathbb{E}Y$. The simulation cost of a single realisation of Y_i increases in i , which calls for optimisation of computational resources.

Such a scenario arises for instance in the context of SDE models, commonly applied in option pricing. In such an application, we might have $Y := f(X_T)$, where $f : \mathbb{R} \rightarrow \mathbb{R}$ is a given ‘payoff’ function, and X_T corresponds to the terminal value of the process $(X_t)_{t \in [0, T]}$

solving the SDE

$$(1) \quad dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \quad t \in [0, T], \quad X_0 \equiv x_0,$$

where $(B_t)_{t \geq 0}$ stands for the standard Brownian motion, the functions $\mu, \sigma : \mathbb{R} \rightarrow \mathbb{R}$ are the drift and diffusion parameters of the model, and $x_0 \in \mathbb{R}$ is the initial value. Unless μ and σ are of certain specific form, the random variable X_T cannot be simulated exactly, but admits easily implementable approximations based on time-discretisation of (1). The commonly applied dyadic uniform meshes are defined as follows

$$\tau_i := (0 = t_0^{(i)} < t_1^{(i)} < \dots < t_{2^i}^{(i)} = T) \quad \text{where} \quad t_j^{(i)} := jT2^{-i} \quad \text{for } i \geq 0.$$

The *Milstein* discretisation (e.g. Kloeden and Platen, 1992) corresponding to τ_i is defined by letting $X_0^{(\tau_i)} = x_0$ and iteratively calculating

$$X_k^{(\tau_i)} := X_{k-1}^{(\tau_i)} + \mu(X_{k-1}^{(\tau_i)})\Delta t_k^{(i)} + \sigma(X_{k-1}^{(\tau_i)})\Delta B_k^{(i)} + \frac{1}{2}\sigma(X_{k-1}^{(\tau_i)})\sigma'(X_{k-1}^{(\tau_i)})[(\Delta B_k^{(i)})^2 - \Delta t_k^{(i)}],$$

for $k = 1, \dots, 2^i$, where $\Delta t_k^{(i)} := t_k^{(i)} - t_{k-1}^{(i)}$ and $\Delta B_k^{(i)} := B_{t_k^{(i)}} - B_{t_{k-1}^{(i)}}$ are independent $N(0, \Delta t_k^{(i)})$ random variables. The final value of such an iteration provides an approximation of X_T , so we may set $Y_{i+1} := f(X_{2^i}^{(\tau_i)})$ for $i \geq 0$. The cost of simulating a single realisation of Y_i is of order 2^i , and under certain, fairly general, conditions on μ, σ and f , the mean square error $\mathbb{E}(Y - Y_{i-1})^2 \leq c2^{-\lambda i}$ with $\lambda > 1$ (e.g. Kloeden and Platen, 1992).

The MLMC is an efficient way to use such approximations in order to approximate $\mathbb{E}Y$. It is based on the following, seemingly trivial observation

$$\mathbb{E}Y_m = \sum_{i=1}^m \mathbb{E}(Y_i - Y_{i-1}), \quad \text{with} \quad Y_0 := 0,$$

which suggests that one may construct an estimate of $\mathbb{E}Y_m$ using a separate Monte Carlo approximation of each term $\mathbb{E}(Y_i - Y_{i-1})$:

$$Z_{\text{ML}} := \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \Delta_i^{(j)},$$

where $n_1, \dots, n_m \in \mathbb{N}$ and $\Delta_i^{(j)}$ are independent random variables with $\mathbb{E}\Delta_i^{(j)} = \mathbb{E}Y_i - \mathbb{E}Y_{i-1}$. It is direct to check that then $\mathbb{E}Z_{\text{ML}} = \mathbb{E}Y_m$. There are two key reasons why MLMC is useful:

- (i) The random variables $\Delta_i^{(j)}$ are usually independent realisations of $\Delta_i = Y_i - Y_{i-1}$, where the approximations Y_i and Y_{i-1} are *dependent*, such that $|\Delta_i|$ is typically small when i is large. In the context of the SDE example discussed above, such a coupling arises naturally when using a common Brownian path in the discretisations leading to Y_i and Y_{i-1} .
- (ii) It allows to *optimise* the computational effort devoted to each ‘level’ to estimate $\mathbb{E}(Y_i - Y_{i-1})$. The key benefit is that fewer samples n_i are often necessary with higher i , which leads to lower overall cost.

Theorem 1 of Giles (2015) quoted below gives the complexity of MLMC under the common exponential framework. It assumes that the expected cost (computational complexity) of each term $\Delta_i^{(j)}$ is κ_i , so that the expected cost of Z_{ML} is $\kappa = \sum_{i=1}^m n_i \kappa_i$.

Theorem 1. *Suppose $(\Delta_i^{(j)})$ are independent, and there exist positive $c_1, c_2, c_3, \alpha, \beta, \gamma$ with $\alpha \geq (\beta \wedge \gamma)/2$, such that for all $i \geq 1$,*

$$|\mathbb{E}Y - \mathbb{E}Y_i| \leq c_1 2^{-\alpha i}, \quad \text{var}(\Delta_i^{(j)}) \leq c_2 2^{-\beta i}, \quad \text{and} \quad \kappa_i \leq c_3 2^{\gamma i}.$$

Then, there exists a positive constant c_4 such that for any $\epsilon < 1/e$ there are values m and $n_1, \dots, n_m \in \mathbb{N}$ such that the MLMC estimator satisfies the following mean square error (MSE) bound

$$\mathbb{E}(Z_{\text{ML}} - \mathbb{E}Y)^2 \leq \epsilon^2,$$

and satisfies the expected cost bound

$$\kappa \leq \begin{cases} c_4 \epsilon^{-2}, & \beta > \gamma, \\ c_4 \epsilon^{-2} (\log \epsilon)^2, & \beta = \gamma, \\ c_4 \epsilon^{-2 - (\gamma - \beta)/\alpha}, & \beta < \gamma. \end{cases}$$

The SDE application discussed above often satisfies the conditions of Theorem 1 with $\beta > \gamma$ (Kloeden and Platen, 1992). In a multidimensional SDE setting, the antithetic truncated Milstein (Giles and Szpruch, 2014) often lead to $\beta > \gamma$ as well. This highlights why MLMC has become so popular—it is often possible to attain a ‘canonical’ rate ϵ^{-2} , equivalent to a square root error rate $\kappa^{-1/2}$, despite the bias, using simple standard discretisation methods. Crude Monte Carlo, that is, taking a fixed level m and averaging independent realisations of Y_m , leads to worse rates. The same canonical square root error rate can be attained with similar assumptions using the debiasing techniques which are discussed next.

3. DEBIASING TECHNIQUES

As with MLMC, assume that $(Y_i)_{i \geq 1}$ and Y are integrable random variables with $\mathbb{E}Y_i \rightarrow \mathbb{E}Y$ and $Y_0 \equiv 0$, then

$$\mathbb{E}Y = \lim_{n \rightarrow \infty} \mathbb{E}Y_n = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E}(Y_i - Y_{i-1}).$$

The fundamental observation behind the unbiased schemes is that one may employ randomisation to pick a finite number of terms of the series to construct estimators with expectation $\mathbb{E}Y$.

The two results below due to Rhee and Glynn (2015) propose two such estimators, along with expressions for their second moments. Proofs of Theorems 3 and 5 are shown to follow as a consequence of Theorem 7 in Section 4 (Examples 9 and 10).

Condition 2. Suppose $(\Delta_i)_{i \geq 1}$ are independent random variables with $\mathbb{E}\Delta_i = \mathbb{E}Y_i - \mathbb{E}Y_{i-1}$, and $(p_i)_{i \geq 1}$ is a probability distribution such that $p_i > 0$ for all $i \in \mathbb{N}$ and

$$\sum_{i \geq 1} \frac{\mathbb{E}\Delta_i^2}{p_i} < \infty.$$

Theorem 3 (Single term estimator). *Suppose $(\Delta_i)_{i \geq 1}$ and $(p_i)_{i \geq 1}$ satisfy Condition 2, and $R \in \mathbb{N}$ is a random variable independent of $(\Delta_i)_{i \geq 1}$ with $\mathbb{P}(R = i) = p_i$, then the single-term estimator*

$$Z^{(1)} := \frac{\Delta_R}{p_R} \quad \text{satisfies} \quad \mathbb{E}Z^{(1)} = \mathbb{E}Y \quad \text{and} \quad \mathbb{E}(Z^{(1)})^2 = \sum_{i \geq 1} \frac{\mathbb{E}\Delta_i^2}{p_i}.$$

McLeish (2011) first suggested the following estimators, but with different conditions and different expression for the variance.

Condition 4. Suppose $(\Delta_i)_{i \geq 1}$ are random variables with $\mathbb{E}\Delta_i = \mathbb{E}Y_i - \mathbb{E}Y_{i-1}$, and $(p_i)_{i \geq 1}$ is a probability distribution such that $\tilde{p}_i := \sum_{j \geq i} p_j > 0$ for all $i \in \mathbb{N}$ and either one of the following hold:

(i) $\Delta_i = Y_i - Y_{i-1}$ and

$$\sum_{i \geq 1} \frac{\mathbb{E}(Y_{i-1} - Y)^2}{\tilde{p}_i} < \infty.$$

(ii) $(\Delta_i)_{i \geq 1}$ are independent and

$$\sum_{i \geq 1} \frac{\text{var}(\Delta_i) + (\mathbb{E}Y - \mathbb{E}Y_{i-1})^2}{\tilde{p}_i} < \infty.$$

Theorem 5 (Sum estimator). *Suppose $(\Delta_i)_{i \geq 1}$ and $(p_i)_{i \geq 1}$ satisfy Condition 4 and $R \in \mathbb{N}$ is a random variable independent of $(\Delta_i)_{i \geq 1}$ with $\mathbb{P}(R = i) = p_i$, then the sum estimator*

$$Z_{\Sigma}^{(1)} := \sum_{i=1}^R \frac{\Delta_i}{\tilde{p}_i}$$

satisfies $\mathbb{E}Z_{\Sigma}^{(1)} = \mathbb{E}Y$. In case of Condition 4 (i),

$$\mathbb{E}(Z_{\Sigma}^{(1)})^2 = \sum_{i \geq 1} \frac{\mathbb{E}(Y_{i-1} - Y)^2 - \mathbb{E}(Y_i - Y)^2}{\tilde{p}_i},$$

and, in case of Condition 4 (ii),

$$\mathbb{E}(Z_{\Sigma}^{(1)})^2 = \sum_{i \geq 1} \frac{\text{var}(\Delta_i) + (\mathbb{E}Y - \mathbb{E}Y_{i-1})^2 - (\mathbb{E}Y - \mathbb{E}Y_i)^2}{\tilde{p}_i}.$$

Remark 6. We follow Rhee and Glynn (2015) and call the estimator of Theorem 5 *coupled sum* in case of Condition 4 (i) and *independent sum* in case of Condition 4 (ii). Note that Condition 4 (ii) is equivalent to the assumption of (Rhee and Glynn, 2015, Theorem 5(a)), because \tilde{p}_i is non-increasing, and so

$$\sum_{i \geq 1} \frac{(\mathbb{E}Y_i - \mathbb{E}Y_{i-1})^2}{\tilde{p}_i} \leq 2 \sum_{i \geq 1} \frac{(\mathbb{E}Y_i - \mathbb{E}Y)^2 + (\mathbb{E}Y_{i-1} - \mathbb{E}Y)^2}{\tilde{p}_i} \leq 4 \sum_{i \geq 1} \frac{(\mathbb{E}Y_{i-1} - \mathbb{E}Y)^2}{\tilde{p}_i}.$$

Let us briefly return to the case where $Y_i = f(X_{2^i}^{(\tau_i)})$ and $X_{2^i}^{(\tau_i)}$ corresponds to the solution of a time-discretised SDE with a mesh of 2^i points as discussed in Section 2. Many approximation schemes admit a weak rate $\alpha > 1/2$, which implies that $|\mathbb{E}Y_i - \mathbb{E}Y_{i-1}| \leq |\mathbb{E}Y - \mathbb{E}Y_{i-1}| + |\mathbb{E}Y - \mathbb{E}Y_i| \leq c2^{-\alpha i}$ (cf. Kloeden and Platen, 1992). If additionally $\text{var}(\Delta_i) \leq c2^{-\beta i}$ with $\beta > 1$, often satisfied for instance by the Milstein scheme (Jentzen et al., 2009; Kloeden and Platen, 1992) or in the multivariate case by the antithetic truncated Milstein scheme (Giles and Szpruch, 2014), then taking $p_i \propto 2^{-\xi i}$, where $\xi \in (1, 2\alpha \wedge \beta)$ satisfies Conditions 2 and 4 (ii). Because the cost of level i is of order 2^i , this sampling scheme admits a finite expected cost (see Section 6 for details). This shows that unbiased estimators with finite variance and finite expected cost can be obtained with the same conditions under which MLMC admits the canonical error rate. In case of correlated $(\Delta_i)_{i \geq 1}$ in Condition 4 (i), a slightly more stringent condition about the strong rate $\mathbb{E}(Y - Y_{i-1})^2 \leq c2^{-\lambda i}$ with $\lambda > 1$ is required.

4. GENERAL UNBIASED SCHEME

Giles (2015) pointed out that averaging n independent single term estimators $Z^{(1)}$ introduced in Theorem 3 corresponds to an estimator of the form

$$\sum_{i=1}^{\infty} \frac{1}{np_i} \sum_{j=1}^{N_i} (Y_i^{(j)} - Y_{i-1}^{(j)}),$$

where N_i is the number of samples from level i , and the expectation of N_i is np_i . This shows a close connection with MLMC, which corresponds to taking $N_i \equiv n_i = np_i$ (up to some level m). Inspired by this remark, and the techniques by [Rhee and Glynn \(2015\)](#), consider the following general unbiased scheme, with an expression for its variance.

Theorem 7. *Suppose $(Y_i)_{i \geq 1}$ and Y are integrable random variables such that $\mathbb{E}Y_i \rightarrow \mathbb{E}Y$, and let $(\Delta_i)_{i \geq 1}$ be square integrable random variables such that $\mathbb{E}\Delta_i = \mathbb{E}Y_i - \mathbb{E}Y_{i-1}$ for all $i \geq 1$, with $\mathbb{E}Y_0 := 0$. Assume $(\Delta_i^{(j)})_{i \geq 1}$ for $j \geq 1$ are independent realisations of the process $(\Delta_i)_{i \geq 1}$. Suppose N_i are non-negative square integrable random integers independent of $(\Delta_i^{(j)})_{i,j}$ and with $\mathbb{E}N_i > 0$ for all $i \geq 1$. Define for $0 \leq \ell < m$*

$$v_{\ell,m} := \sum_{i,k=\ell+1}^m \frac{\text{cov}(\Delta_i, \Delta_k) \mathbb{E}(N_i \wedge N_k) + \mathbb{E}\Delta_i \mathbb{E}\Delta_k \text{cov}(N_i, N_k)}{\mathbb{E}N_i \mathbb{E}N_k}.$$

If there exists a strictly increasing sequence of positive integers $(m_k)_{k \geq 1}$ such that $\lim_{k \rightarrow \infty} \sup_{j \geq 1} v_{m_k, m_{k+j}} = 0$ and $\sum_{i \geq 1} N_i < \infty$ almost surely, then the estimator

$$Z := \sum_{i=1}^{\infty} \frac{1}{\mathbb{E}N_i} \sum_{j=1}^{N_i} \Delta_i^{(j)}$$

is unbiased $\mathbb{E}Z = \mathbb{E}Y$ and $\text{var}(Z) = \lim_{k \rightarrow \infty} v_{0, m_k} < \infty$.

The proof of Theorem 7 is given in Appendix B.

Consider first a simple example of Theorem 7, where N_i are taken independent.

Example 8 (Independent levels). Let $(N_i)_{i \geq 1}$ be independent with $\mu_i := \mathbb{E}N_i$ and $\sum_{i \geq 1} \mu_i < \infty$, and suppose $\beta_i := \text{var}(N_i)/\mu_i < \infty$, and let $(\Delta_i)_{i \geq 1}$ be uncorrelated. Then,

$$v_{\ell,m} = \sum_{i=\ell+1}^m \frac{\text{var}(\Delta_i) \mathbb{E}N_i + \text{var}(N_i) (\mathbb{E}\Delta_i)^2}{(\mathbb{E}N_i)^2} = \sum_{i=\ell+1}^m \frac{\text{var}(\Delta_i) + \beta_i (\mathbb{E}\Delta_i)^2}{\mu_i}.$$

If $\sum_i N_i < \infty$ a.s. and $v := \sum_{i \geq 1} [\text{var}(\Delta_i) + \beta_i (\mathbb{E}\Delta_i)^2] / \mu_i < \infty$, then the corresponding estimator is unbiased with variance v . In particular, taking N_i Poisson with intensity np_i where $\sum_{i \geq 1} p_i = 1$, then $\sum_i N_i < \infty$ by the Borel-Cantelli lemma and $v = n^{-1} \sum_{i \geq 1} \mathbb{E}\Delta_i^2 / p_i$.

The averages of single term estimators and sum estimators introduced in Theorems 3 and 5 correspond to certain dependence structures of $(N_i)_{i \geq 1}$, as we shall see next. The following two examples are important, because the new schemes introduced in Section 5 will be based on the same constructions.

Example 9 (Single term estimators). Suppose Condition 2 holds. Define

$$N_i := \sum_{j=1}^n \mathbb{I} \{R^{(j)} = i\}, \quad \text{where } R^{(j)} \text{ are independent with } P(R^{(j)} = i) = p_i,$$

and call the estimator $Z_{\text{id}}^{(n)}$. It is easy to see that this corresponds to an average of n independent single term estimators $Z^{(1)}$. Note that $(N_i)_{i \geq 1}$ follow a multinomial distribution with parameters n and $(p_i)_{i \geq 1}$. We have $\mathbb{E}N_i = np_i$ and

$$\mathbb{E}(N_i N_k) = \mathbb{E} \left[\sum_{j,\ell=1}^n \mathbb{I} \{R^{(j)} = i\} \mathbb{I} \{R^{(\ell)} = k\} \right] = np_i \mathbb{I} \{i = k\} + n(n-1)p_i p_k,$$

so $\text{cov}(N_i, N_k) = n(p_i \mathbb{I}\{i = k\} - p_i p_k)$. By the independence of $(\Delta_i)_{i \geq 1}$,

$$(2) \quad \begin{aligned} v_{\ell, m} &= \sum_{i=\ell+1}^m \frac{\text{var}(\Delta_i)}{\mathbb{E}N_i} + \sum_{i, k=\ell+1}^m \frac{\mathbb{E}\Delta_i \mathbb{E}\Delta_k \text{cov}(N_i, N_k)}{\mathbb{E}N_i \mathbb{E}N_k} \\ &= \frac{1}{n} \left(\sum_{i=\ell+1}^m \frac{\mathbb{E}\Delta_i^2}{p_i} - (\mathbb{E}Y_m - \mathbb{E}Y_\ell)^2 \right). \end{aligned}$$

By assumption $\sum_{i=\ell}^{\infty} \mathbb{E}\Delta_i^2/p_i < \infty$ and $\mathbb{E}Y_i \rightarrow \mathbb{E}Y$, so $\lim_{\ell \rightarrow \infty} \sup_{m > \ell} v_{\ell, m} = 0$. The variance satisfies

$$\text{var}(Z_{\text{iid}}^{(n)}) = \lim_{m \rightarrow \infty} v_{0, m} = \frac{1}{n} \left(\sum_{i=1}^{\infty} \frac{\mathbb{E}\Delta_i^2}{p_i} - (\mathbb{E}Y)^2 \right),$$

so $\text{var}(Z_{\text{iid}}^{(n)}) = n^{-1} \text{var}(Z^{(1)})$. The above also proves Theorem 3 with $n = 1$.

Example 10 (Sum estimators). Suppose $(\Delta_i^{(j)})_{i \geq 1}$ are independent realisations of $(\Delta_i)_{i \geq 1}$ satisfying Condition 4, and define $\tilde{N}_i := \sum_{j=1}^n \mathbb{I}\{R^{(j)} \geq i\}$, where $R^{(j)}$ are independent with $\mathbb{P}(R^{(j)} = i) = p_i$. This estimator, which we denote by $Z_{\Sigma, \text{iid}}^{(n)}$, corresponds to an average of n independent sum estimators $Z_{\Sigma}^{(1)}$:

$$Z_{\Sigma, \text{iid}}^{(n)} = \sum_{i=1}^{\infty} \frac{1}{\mathbb{E}\tilde{N}_i} \sum_{j=1}^{\tilde{N}_i} \Delta_i^{(j)} \stackrel{d}{=} \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^{R^{(j)}} \frac{\Delta_i^{(j)}}{\tilde{p}_i},$$

because $\mathbb{E}\tilde{N}_i = n\tilde{p}_i$, and where the latter equality would hold if $(R^{(j)})$ were non-increasing. Changing the indexing to ensure that $(R^{(j)})$ are non-increasing does not affect the distribution. For all $i, k \geq 1$,

$$\mathbb{E}(\tilde{N}_i \tilde{N}_k) = \sum_{j=1}^n \tilde{p}_{i \vee k} + \sum_{\substack{j, \ell=1 \\ j \neq \ell}}^n \tilde{p}_i \tilde{p}_k = \mathbb{E}\tilde{N}_{i \vee k} + \frac{n-1}{n} \mathbb{E}\tilde{N}_i \mathbb{E}\tilde{N}_k,$$

so $\text{cov}(\tilde{N}_i, \tilde{N}_k) = \mathbb{E}\tilde{N}_{i \vee k} - n^{-1} \mathbb{E}\tilde{N}_i \mathbb{E}\tilde{N}_k$. We also have $\mathbb{E}(\tilde{N}_i \wedge \tilde{N}_k) = \mathbb{E}\tilde{N}_{i \vee k}$ because $\tilde{N}_i \geq \tilde{N}_k$ for $i \leq k$, so for $1 \leq \ell < m$,

$$\begin{aligned} v_{\ell, m} &= \sum_{i, k=\ell+1}^m \frac{\mathbb{E}\tilde{N}_{i \vee k} (\text{cov}(\Delta_i, \Delta_k) + \mathbb{E}\Delta_i \mathbb{E}\Delta_k)}{\mathbb{E}\tilde{N}_i \mathbb{E}\tilde{N}_k} - \frac{1}{n} \sum_{i, k=\ell+1}^m \mathbb{E}\Delta_i \mathbb{E}\Delta_k \\ &= \sum_{i=\ell+1}^m \left(\frac{\mathbb{E}\Delta_i^2}{\mathbb{E}\tilde{N}_i} + 2 \sum_{k=i+1}^m \frac{\mathbb{E}(\Delta_i \Delta_k)}{\mathbb{E}\tilde{N}_i} \right) - \frac{1}{n} (\mathbb{E}Y_m - \mathbb{E}Y_\ell)^2. \end{aligned}$$

Denote for the rest of the proof $D_{i, m} := Y_i - Y_m$. If Condition 4 (i) holds, we obtain

$$v_{\ell, m} = \frac{1}{n} \sum_{i=\ell+1}^m \frac{\mathbb{E}D_{i-1, m}^2 - \mathbb{E}D_{i, m}^2}{\tilde{p}_i} - \frac{1}{n} (\mathbb{E}D_{m, \ell})^2,$$

because $\Delta_i^2 + 2\Delta_i D_{m, i} = D_{i-1, m}^2 - D_{i, m}^2$. Let $(m_k)_{k \geq 1}$ be from Lemma 29 in Appendix A, then for any $k, j \geq 1$

$$v_{m_k, m_{k+j}} \leq \frac{4}{n} \sum_{i=m_k+1}^{\infty} \frac{\mathbb{E}(Y_{i-1} - Y)^2}{\tilde{p}_i} - \frac{1}{n} (\mathbb{E}D_{m_{k+j}, m_k})^2,$$

which converges to zero as $k \rightarrow \infty$. For any fixed $k \geq 1$, we have $v_{0,m_{k+j}} \rightarrow \text{var}(Z_{\Sigma,\text{iid}}^{(n)})$ as $j \rightarrow \infty$, so

$$\text{var}(Z_{\Sigma,\text{iid}}^{(n)}) = \frac{1}{n} \lim_{j \rightarrow \infty} \left(\sum_{i=1}^{m_k} \frac{\mathbb{E}D_{i-1,m_{k+j}}^2 - \mathbb{E}D_{i,m_{k+j}}^2}{\tilde{p}_i} + (\mathbb{E}Y_{m_{k+j}})^2 + \sum_{i=m_k+1}^{m_{k+j}} \frac{\mathbb{E}D_{i-1,m_{k+j}}^2 - \mathbb{E}D_{i,m_{k+j}}^2}{\tilde{p}_i} \right).$$

The latter sum is upper bounded by $4 \sum_{i \geq m_k+1} \mathbb{E}(Y_{i-1} - Y)^2 / \tilde{p}_i$, and because $\mathbb{E}D_{i,m_{k+j}}^2 \rightarrow \mathbb{E}(Y_i - Y)^2$ and $\mathbb{E}Y_{m_{k+j}} \rightarrow \mathbb{E}Y$ as $j \rightarrow \infty$, we may conclude that

$$\text{var}(Z_{\Sigma,\text{iid}}^{(n)}) = \frac{1}{n} \left(\sum_{i=1}^{m_k} \frac{\mathbb{E}(Y_{i-1} - Y)^2 - \mathbb{E}(Y_i - Y)^2}{\tilde{p}_i} - (\mathbb{E}Y)^2 \right) = \frac{1}{n} \text{var}(Z_{\Sigma}^{(1)}).$$

Suppose then that Condition 4 (ii) holds. It is straightforward to check that

$$v_{\ell,m} = \frac{1}{n} \sum_{i=\ell+1}^m \frac{\text{var}(\Delta_i) + (\mathbb{E}Y_{i-1} - \mathbb{E}Y_m)^2 - (\mathbb{E}Y_i - \mathbb{E}Y_m)^2}{\tilde{p}_i} - \frac{1}{n} (\mathbb{E}Y_m - \mathbb{E}Y_\ell)^2,$$

which satisfies $\lim_{\ell \rightarrow \infty} \sup_{m > \ell} v_{\ell,m} = 0$ by assumption, and similarly as above,

$$\text{var}(Z_{\Sigma,\text{iid}}^{(n)}) = \lim_{k \rightarrow \infty} v_{0,k} = \frac{1}{n} \left(\sum_{i=1}^{\infty} \frac{\text{var}(\Delta_i) + (\mathbb{E}Y_{i-1} - \mathbb{E}Y)^2 - (\mathbb{E}Y_i - \mathbb{E}Y)^2}{\tilde{p}_i} - (\mathbb{E}Y)^2 \right).$$

Taking $n = 1$ concludes also the proof of Theorem 5.

The last example is an obvious ‘hybrid’ scheme involving MLMC and an ‘unbiased tail’ scheme, which also falls into the framework of Theorem 7.

Example 11. Assume that $r, m \in \mathbb{N}$ and that $N_i \equiv n_i \in \mathbb{N}$ for $i = 1, \dots, m$, and $(p_i)_{i > m}$ are positive with $\sum_{i > m} p_i = 1$, let $R^{(j)}$ be independent with $\mathbb{P}(R^{(j)} = i) = p_i$, and define $N_i := \sum_{j=1}^r \mathbb{1}\{R^{(j)} = i\}$ for $i > m$. The estimator can be written as

$$Z = \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \Delta_i^{(j)} + \sum_{i=m+1}^{\infty} \frac{1}{rp_i} \sum_{j=1}^{N_i} \Delta_i^{(j)}.$$

The first term coincides with an MLMC estimator with m levels, with expectation $\mathbb{E}Y_m$, and the second term is an average of single term estimators of $\mathbb{E}Y - \mathbb{E}Y_m$, with r samples. Note that we could have used any unbiased scheme in the latter part, provided it satisfies the conditions in Theorem 7.

The new sampling schemes discussed next in Section 5 provide a different view on the balancing; see in particular Theorem 19.

5. NEW UNBIASED ESTIMATORS

Let us now turn into new practically interesting estimators which correspond to specific choices in Theorem 7. The estimators are based on stratification, a classical variance reduction technique in survey sampling (e.g. Hansen et al., 1953), which have also been widely used in Monte Carlo; see for instance Douc et al. (2005); Glasserman (2003).

The following lemma states classical results on stratification (cf. Vihola, 2015, for proof).

Lemma 12. *Let X be an integrable random variable and let A_1, \dots, A_m be exhaustive disjoint events with $\mathbb{P}(A_i) = q_i > 0$, and let $\ell_1, \dots, \ell_m \in \mathbb{N}$. Assume $X_i^{(j)}$ are random variables with conditional laws $\mathbb{P}(X_i^{(j)} \in B) = \mathbb{P}(\{X \in B\} \cap A_i) / q_i$. The stratified estimator $\tilde{X} := \sum_{i=1}^m (q_i / \ell_i) \sum_{j=1}^{\ell_i} X_i^{(j)}$ satisfies*

- (i) *Unbiasedness* $\mathbb{E}\tilde{X} = \mathbb{E}X$ and $\text{var}(\tilde{X}) \leq \text{var}(X)$.
- (ii) If $\{X_i^{(j)}\}_{i,j}$ are independent and $\ell_i = q_i \ell$ (proportional allocation), then $\tilde{X} = \ell^{-1} \sum_{i=1}^m \sum_{j=1}^{\ell_i} X_i^{(j)}$ and $\text{var}(\tilde{X}) \leq \ell^{-1} \text{var}(X)$.

In what follows, assume $(p_i)_{i \geq 1}$ are positive and such that $\sum_{i \geq 1} p_i = 1$, and denote $\tilde{p}_i := \sum_{j \geq i} p_j$. Let us first consider uniform stratification schemes. For that purpose, recall the definition of the generalised inverse distribution function $F^{-1} : (0, 1) \rightarrow \mathbb{N}$ corresponding to $(p_i)_{i \geq 1}$,

$$F^{-1}(u) := \min\{k \in \mathbb{N} : F(k) \geq u\} \quad \text{with} \quad F(k) := \sum_{i=1}^k p_i.$$

The well-known inverse distribution function method states that a uniform $U \sim U(0, 1)$ is transformed by F^{-1} into $R = F^{-1}(U)$ with $\mathbb{P}(R = i) = p_i$.

Definition 13 (Uniformly stratified estimators). Let $n \in \mathbb{N}$, and assume $U^{(j)}$ are independent $U\left(\frac{j-1}{n}, \frac{j}{n}\right)$ random variables for $j = 1, \dots, n$. Let $R^{(j)} = F^{-1}(U^{(j)})$.

- (i) Define $N_i := \sum_{j=1}^n \mathbb{I}\{R^{(j)} = i\}$, then the estimator $Z_{\text{str}}^{(n)}$ defined as in Theorem 7 is the *uniformly stratified single term estimator*.
- (ii) Define $\tilde{N}_i := \sum_{j=1}^n \mathbb{I}\{R^{(j)} \geq i\}$, then the estimator $Z_{\Sigma, \text{str}}^{(n)}$ defined as in Theorem 7 is the *uniformly stratified sum estimator*.

Definition 14 (Systematic sampling estimators). Let $u_j := (j-1)/n$ and define $U^{(j)} := u_j + U$ for all $j = 1, \dots, n$, where $U \sim U(0, 1/n)$, and let $R^{(j)} := F^{-1}(U^{(j)})$. Define then N_i and \tilde{N}_i as in Definition 13 (i) and (ii), respectively, and the corresponding *systematic sampling single-term estimator* $Z_{\text{sys}}^{(n)}$ and *systematic sampling sum estimator* $Z_{\Sigma, \text{sys}}^{(n)}$.

The consistency and a variance bound for the uniformly stratified and systematic sampling estimators are stated in Proposition 16, after introducing another slightly different stratification scheme.

Definition 15 (Residual sampling estimators). Let $n \in \mathbb{N}$, define $n_i := \lfloor np_i \rfloor$ and let $r := n - \sum_{i \geq 1} n_i \geq 0$. If $r > 0$, define the ‘residual’ probability distribution $(p_i^*)_{i \geq 1}$ as $p_i^* := (np_i - n_i)/r$, and let $Q^{(j)}$ be independent random variables such that $\mathbb{P}(Q^{(j)} = i) = p_i^*$ for $j = 1, \dots, r$.

- (i) Define $N_i := n_i + N_i^*$ where $N_i^* := \sum_{j=1}^r \mathbb{I}\{Q^{(j)} = i\}$, then the estimator $Z_{\text{res}}^{(n)}$ defined as in Theorem 7 is the *residual sampling single term estimator*.
- (ii) Define $\tilde{N}_i := \tilde{n}_i + \tilde{N}_i^*$, where $\tilde{n}_i := \sum_{k=1}^i n_k$ and $\tilde{N}_i^* := \sum_{j=1}^r \mathbb{I}\{Q^{(j)} \geq i\}$, then the estimator $Z_{\Sigma, \text{res}}^{(n)}$ defined as in Theorem 7 is the *residual sampling sum estimator*.

Proposition 16. *The estimators in Definitions 13, 14 and 15 satisfy:*

- (i) *Assume Condition 2, then*

$$\begin{aligned} \mathbb{E}Z_{\text{str}}^{(n)} &= \mathbb{E}Z_{\text{res}}^{(n)} = \mathbb{E}Y, & \text{var}(Z_{\text{str}}^{(n)}) \vee \text{var}(Z_{\text{res}}^{(n)}) &\leq n^{-1} \text{var}(Z^{(1)}), \\ \mathbb{E}Z_{\text{sys}}^{(n)} &= \mathbb{E}Y, & \text{var}(Z_{\text{sys}}^{(n)}) &\leq \text{var}(Z^{(1)}). \end{aligned}$$

- (ii) *Assume Condition 4, then*

$$\begin{aligned} \mathbb{E}Z_{\Sigma, \text{str}}^{(n)} &= \mathbb{E}Z_{\Sigma, \text{res}}^{(n)} = \mathbb{E}Y, & \text{var}(Z_{\Sigma, \text{str}}^{(n)}) \vee \text{var}(Z_{\Sigma, \text{res}}^{(n)}) &\leq n^{-1} \text{var}(Z_{\Sigma}^{(1)}), \\ \mathbb{E}Z_{\Sigma, \text{sys}}^{(n)} &= \mathbb{E}Y, & \text{var}(Z_{\Sigma, \text{sys}}^{(n)}) &\leq \text{var}(Z_{\Sigma}^{(1)}). \end{aligned}$$

Proposition 16 follows as a consequence of Lemma 12; the detailed proof is given in Appendix C.

Remark 17. Instead of using $u_j = (j-1)/n$ in systematic sampling, we could use instead any sequence $u_1, \dots, u_n \in [0, 1]$, and let $U^{(j)} := u_j + U \pmod{1}$, where $U \sim U(0, 1)$. This would not be stratification, but it is direct to check the estimators still retain the same expectation, and also the same pessimistic variance bound. For instance, using a low-discrepancy sequence (u_j) would correspond to randomised quasi-Monte Carlo (e.g. [Dick et al., 2013](#)).

Proposition 16 states that all the new estimators are unbiased, and that the uniformly stratified and residual sampling estimators cannot be worse than averages of independent single term and sum estimators, in terms of variance. In fact, they often have a strictly lower variance, but it is generally difficult to quantify the benefit. The following Theorem 19 is often more useful, indicating that stratified and residual sampling schemes attain asymptotically the efficiency of MLMC under general conditions.

Let us first formulate ‘idealised’ MLMC strategies based on limiting allocations $(p_i)_{i \geq 1}$.

Definition 18. Assume $(p_i)_{i \geq 1}$ are non-negative with $\sum_{i \geq 1} p_i = 1$ and define $\tilde{p}_i = \sum_{j \geq i} p_j$. For any $n \in \mathbb{N}$, define $n_i := \lfloor np_i \rfloor$, $\hat{n}_i := \lfloor n\tilde{p}_i \rfloor$, $m_n = \max\{i \geq 0 : n_i > 0\}$ and $\tilde{m}_n = \max\{i \geq 0 : \hat{n}_i > 0\}$ and denote

$$Z_{\text{ML}}^{(n)} := \sum_{i=1}^{m_n} \frac{\mathbb{I}\{n_i > 0\}}{n_i} \sum_{j=1}^{n_i} \Delta_i^{(j)}, \quad \text{and} \quad Z_{\Sigma, \text{ML}}^{(n)} := \sum_{i=1}^{\tilde{m}_n} \frac{1}{\hat{n}_i} \sum_{j=1}^{\hat{n}_i} \Delta_i^{(j)}.$$

Practical implementations of MLMC are often based on application of stopping rules, which may determine n_i and m_n during simulation. Definition 18 can be therefore viewed as ‘idealised’ version of MLMC, where (nearly) optimal allocation strategy $(p_i)_{i \geq 1}$ is known beforehand, and n_i and m_n are determined in terms of a single ‘running-time’ parameter n .

Theorem 19. Below, $Z_*^{(n)}$ (resp. $Z_{\Sigma, *}^{(n)}$) stands for either $Z_{\text{res}}^{(n)}$ or $Z_{\text{str}}^{(n)}$ (resp. $Z_{\Sigma, \text{res}}^{(n)}$ or $Z_{\Sigma, \text{str}}^{(n)}$).

(i) Assume Condition 2 then

$$\lim_{n \rightarrow \infty} n \text{var}(Z_{\text{ML}}^{(n)}) = \lim_{n \rightarrow \infty} n \text{var}(Z_*^{(n)}) = \sigma_\infty^2 := \sum_{i=1}^{\infty} \frac{\text{var}(\Delta_i)}{p_i}.$$

(ii) Assume Condition 4 (i), then

$$\lim_{n \rightarrow \infty} n \text{var}(Z_{\Sigma, *}^{(n)}) = \sigma_{\Sigma, \infty}^2 := \sum_{i=1}^{\infty} \frac{\text{var}(Y - Y_{i-1}) - \text{var}(Y - Y_i)}{\tilde{p}_i},$$

and if $\sum_{i \geq 1} \sup_{k \geq i} \text{var}(Y_{k-1} - Y) / \tilde{p}_k < \infty$, then $\lim_{n \rightarrow \infty} n \text{var}(Z_{\Sigma, \text{ML}}^{(n)}) = \sigma_{\Sigma, \infty}^2$.

(iii) Assume Condition 4 (ii), then

$$\lim_{n \rightarrow \infty} n \text{var}(Z_{\Sigma, \text{ML}}^{(n)}) = \lim_{n \rightarrow \infty} n \text{var}(Z_{\Sigma, *}^{(n)}) = \sigma_{\Sigma, \infty}^2 = \sum_{i \geq 1} \frac{\text{var}(\Delta_i)}{\tilde{p}_i}.$$

The proof of Theorem 19 is given in Appendix D.

Remark 20. The limiting variances in Theorem 19 can be significantly lower than the upper bounds given in Proposition 16, which correspond to averaging independent unbiased estimators. Indeed, $\text{var}(\Delta_i) = \mathbb{E}\Delta_i^2 - (\mathbb{E}\Delta_i)^2$, so

$$\sigma_\infty^2 = \text{var}(Z^{(1)}) - \left[\sum_{i \geq 1} \frac{(\mathbb{E}\Delta_i)^2}{p_i} - (\mathbb{E}Y)^2 \right].$$

Likewise, in case of Condition 4 (i), $\text{var}(Y - Y_i) = \mathbb{E}(Y - Y_i)^2 - (\mathbb{E}Y - \mathbb{E}Y_i)^2$, so

$$\begin{aligned}\sigma_{\Sigma, \infty}^2 &= \mathbb{E}(Z_{\Sigma}^{(1)})^2 - \sum_{i \geq 1} \frac{(\mathbb{E}Y - \mathbb{E}Y_{i-1})^2 - (\mathbb{E}Y - \mathbb{E}Y_i)^2}{\tilde{p}_i} \\ &= \text{var}(Z_{\Sigma}^{(1)}) - \sum_{i \geq 1} (\mathbb{E}Y - \mathbb{E}Y_i)^2 \left(\frac{1}{\tilde{p}_{i+1}} - \frac{1}{\tilde{p}_i} \right).\end{aligned}$$

Note that $\tilde{p}_{i+1} \leq \tilde{p}_i$ for all i , so all the terms in the sum are positive.

6. EXPECTED COST AND ASYMPTOTICALLY OPTIMAL DISTRIBUTION

Let us next consider the cost of simulating estimators Z of the form given in Theorem 7. Assume each $\Delta_i^{(j)}$ has random cost $K_{i,j}$, such that the total cost of Z is $K := \sum_{i \geq 1} \sum_{j=1}^{N_i} K_{i,j}$. Assume also that $\{K_{i,j}\}_{i,j \geq 1}$ are independent of $(N_i)_{i \geq 1}$ and $\{K_{i,j}\}_{j \geq 1}$ have a common mean $\kappa_i = \mathbb{E}K_{i,j} < \infty$. The expected cost can be written down as follows.

$$\mathbb{E}K = \sum_{i \geq 1} \mu_i \kappa_i \quad \text{with} \quad \mu_i := \mathbb{E}N_i.$$

The following records the immediate fact that the estimators introduced in Section 5 have the same expected cost as the simple averages of independent estimators.

Proposition 21. *Denote the cost of the estimator $Z_*^{(n)}$ by $K_*^{(n)}$, where ‘*’ is a place holder. Then,*

$$\begin{aligned}(i) \quad &\mathbb{E}K_{\text{str}}^{(n)} = \mathbb{E}K_{\text{res}}^{(n)} = \mathbb{E}K_{\text{sys}}^{(n)} = \mathbb{E}K_{\text{iid}}^{(n)} = n\mathbb{E}K^{(1)} = n \sum_{i \geq 1} p_i \kappa_i, \\ (ii) \quad &\mathbb{E}K_{\Sigma, \text{str}}^{(n)} = \mathbb{E}K_{\Sigma, \text{res}}^{(n)} = \mathbb{E}K_{\Sigma, \text{sys}}^{(n)} = \mathbb{E}K_{\Sigma, \text{iid}}^{(n)} = n\mathbb{E}K_{\Sigma}^{(1)} = n \sum_{i \geq 1} \tilde{p}_i \kappa_i.\end{aligned}$$

The following proposition records that both of the MLMC estimators introduced in Definition 18 have a cost that is asymptotically equivalent to the corresponding unbiased estimators, which are based on the same limiting allocation strategy.

Proposition 22. *The expected cost of MLMC estimators in Theorem 19 satisfy*

$$\lim_{n \rightarrow \infty} n^{-1} \mathbb{E}K_{\text{ML}}^{(n)} = \sum_{i \geq 1} p_i \kappa_i \quad \text{and} \quad \lim_{n \rightarrow \infty} n^{-1} \mathbb{E}K_{\Sigma, \text{ML}}^{(n)} = \sum_{i \geq 1} \tilde{p}_i \kappa_i.$$

Proof. Proof. The expected cost of $Z_{\text{ML}}^{(n)}$ is

$$\sum_{i \geq 1} \mathbb{I}\{n_i > 0\} n_i \kappa_i, \quad \text{where} \quad n_i := \lfloor np_i \rfloor.$$

Clearly $n_i/n \leq p_i$, $n_i/n \rightarrow p_i$ and $\mathbb{I}\{n_i > 0\} \rightarrow 1$ as $n \rightarrow \infty$, so if $\sum_{i \geq 1} p_i \kappa_i < \infty$, the result follows by dominated convergence; otherwise one can use Fatou’s lemma. The proof with $\tilde{K}_{\text{ML}}^{(n)}$ is identical. \square

Glynn and Whitt (1992a) formulate an asymptotic efficiency principle, which states that if $(Z^{(j)})_{j \geq 1}$ are i.i.d. estimators with finite variance $\sigma^2 < \infty$ and with expected cost $\kappa < \infty$, then the average of such estimators has asymptotic relative efficiency $[\kappa \sigma^2]^{-1}$. Considering this notion of efficiency, let $n \in \mathbb{N}$, and consider the following estimator, which is the average of m independent realisations of the estimators above,

$$(3) \quad \mathcal{Z}_*^{(n,m)} := \frac{1}{m} \sum_{j=1}^m Z_*^{(n,j)},$$

where $\{Z_*^{(n,j)}\}_{j \geq 1}$ are independent realisations of $Z_*^{(n)}$ and ‘*’ is a place holder. By letting $m \rightarrow \infty$, one may consider the asymptotic efficiency of this estimator as in the following

result, stating that stratified and residual sampling procedures always improve upon averages of independent single term and sum estimators, and can be made arbitrarily close to MLMC in asymptotic efficiency.

Corollary 23. *Suppose $n \in \mathbb{N}$ is fixed, then the following asymptotic efficiency holds when $m \rightarrow \infty$.*

- (i) *If Condition 2 holds and $\sum_{i \geq 1} p_i \kappa_i < \infty$, then the asymptotic efficiency of both $Z_{\text{str}}^{(n,m)}$ and $Z_{\text{res}}^{(n,m)}$ is no worse than that of $Z_{\text{id}}^{(mn)}$, and can be made arbitrarily close to that of $Z_{\text{ML}}^{(mn)}$ by choosing n sufficiently large.*
- (ii) *If Condition 4 holds and $\sum_{i \geq 1} \tilde{p}_i \kappa_i < \infty$, then the asymptotic efficiency of both $Z_{\Sigma, \text{str}}^{(n,m)}$ and $Z_{\Sigma, \text{res}}^{(n,m)}$ is no worse than that of $Z_{\Sigma, \text{id}}^{(mn)}$, and can be made arbitrarily close to that of $Z_{\Sigma, \text{ML}}^{(mn)}$ by choosing n sufficiently large.*

Remark 24. Strictly speaking, the use of asymptotic efficiency principle requires that the multilevel estimators $Z_{\Sigma, \text{id}}^{(mn)}$ would be guaranteed to satisfy a functional central limit theorem, which is out of the scope of this work; see, however, the recent works of [Alaya and Kebaier \(2015\)](#) and [Zheng and Glynn \(2017\)](#) in this direction in a different setting.

If we assume n is taken sufficiently large so that the limits in [Theorem 19](#) are approximately attained, the asymptotic efficiency principle suggests a rule for tuning the distribution (p_i) for stratified and residual sampling distributions: the distribution $(p_i)_{i \geq 1}$ which minimises the asymptotic inverse relative efficiency (IRE) $\sigma_{\infty}^2 \mathbb{E}K_{*}^{(n)}$ or $\sigma_{\Sigma, \infty}^2 \mathbb{E}K_{\Sigma, * }^{(n)}$ maximises the efficiency. For the single term estimator, [Condition 2](#), this leads to

$$(4) \quad \min_{(p_i)_{i \geq 1}} \left(\sum_{i=1}^{\infty} \frac{\text{var}(\Delta_i)}{p_i} \right) \left(\sum_{i=1}^{\infty} p_i \kappa_i \right).$$

The solution to (4) is proportional to $\beta_i := \sqrt{\text{var}(\Delta_i)/\kappa_i}$, if $\sum_i \beta_i < \infty$ ([Rhee and Glynn, 2015](#), Proposition 1). This is straightforward to implement in practice, because reliable estimates of $\text{var}(\Delta_i)$ are easily available. A straightforward practical procedure, also implemented in the experiments below, is to use the ‘empirical optimal distribution’ $\beta_i / \sum_i \beta_i$ to define directly the first m probabilities p_1, \dots, p_m and a tail probability $\sum_{j > m} p_j$. The tail probabilities p_j for $j > m$ are then defined to follow a parametric distribution which guarantees a finite variance based on theory. In the experiments, the tail distribution was chosen to be geometric.

In case of the independent sum estimator, [Condition 4 \(ii\)](#), similar minimisation (4) with \tilde{p}_i in place of p_i yields the asymptotically optimal distribution $p_i = \tilde{p}_i - \tilde{p}_{i+1}$. However, contrary to the single-term estimator, $(\tilde{p}_i)_{i \geq 1}$ must additionally be non-increasing. Because (4) is invariant under multiplicative constants on $(p_i)_{i \geq 1}$, the independent sum estimator can never outperform the single term estimator, in terms of asymptotic IRE.

The coupled sum estimator, [Condition 4 \(i\)](#), leads to the optimisation problem

$$\min_{(\tilde{p}_i)_{i \geq 1}} \left(\sum_{i=1}^{\infty} \frac{\text{var}(Y - Y_{i-1}) - \text{var}(Y - Y_i)}{\tilde{p}_i} \right) \left(\sum_{i=1}^{\infty} \tilde{p}_i \kappa_i \right).$$

This is more involved for two reasons. Estimation of the terms $\text{var}(Y - Y_i)$ is not straightforward. In practice, a reasonable (but expensive) approximation may be obtained by approximating Y with Y_m , where m is large. As noted in ([Rhee and Glynn, 2015](#), §3), the coupled sum estimator offers one more degree of freedom. Namely, it is possible to consider a coupled sum estimator for a subsequence of the variables (Y_i) , that is, employing level

differences $\tilde{\Delta}_i^{(j)} := Y_{J_i} - Y_{J_{i-1}}$, where $J = (0 = J_0 < J_1 < \dots)$. Then, the optimisation would require solving

$$\min_J \min_{(\tilde{p}_i)_{i \geq 1}} \left(\sum_{i=1}^{\infty} \frac{\text{var}(Y - Y_{J_{i-1}}) - \text{var}(Y - Y_{J_i})}{\tilde{p}_i} \right) \left(\sum_{i=1}^{\infty} \tilde{p}_i \kappa_{J_i} \right).$$

This leads to a combinatorial problem, which is discussed in depth in (Rhee and Glynn, 2015, §3), who also describe a dynamic programming algorithm which finds such optimal J , up to an index m , in $O(m^3)$ time.

The asymptotic optimality results above suggest that when considering an average of estimators of the form (3), it is possible to sequentially refine the randomisation distribution $(p_i)_{i \geq 1}$ used for $Z_*^{(n,j)}$ based on the random variables generated for the previous estimators $Z_*^{(n,1)}, \dots, Z_*^{(n,j-1)}$, as suggested by Rhee and Glynn (2015). In particular, the probabilities $(p_i)_{i \geq 1}$ of single term estimators could be chosen based on empirical variances of level i variables generated for previous estimators. Then, $Z_*^{(n,j)}$ would not be independent, but remain unbiased, and in fact $(Z_*^{(n,j)} - \mathbb{E}Y)_{j \geq 1}$ would be martingale differences.

In some applications it is not possible to choose $(p_i)_{i \geq 1}$ which yield both finite variance and finite expected cost. Then, it is not possible to attain a canonical square root convergence rate, but Rhee and Glynn (2015) suggest another approach: choose $(p_i)_{i \geq 1}$ that ensure finite variance, but which imply infinite expected cost. Using a result due to Feller (1946), they deduce complexity results for unbiased estimators which are close to what are possible with MLMC; see also Zheng et al. (2017). However, quantifying the efficiency in the present setting is not straightforward, so this is left for future work.

7. GENERALISED UNBIASED SCHEME AND DEPENDENT RANDOMISATION

The general unbiased scheme proposed in Theorem 7 is based on independent randomisation, that is, $(N_i)_{i \geq 1}$ are assumed independent of $(\Delta_i^{(j)})$. Such a scheme is often appropriate in practice, but it is also possible to think of cases where $(N_i)_{i \geq 1}$ could depend on $(\Delta_i^{(j)})$, for instance in a stopping time fashion. It is also possible to retain unbiasedness while replacing $\mathbb{E}N_i$ in the estimator by a conditional expectation. We consider below a scheme which accomodates both of these generalisations, while retaining unbiasedness.

Condition 25. Suppose $(\Delta_i^{(j)})_{i,j \geq 1}$, $(Y_i)_{i \geq 0}$ and Y are integrable random variables, $\mathbb{E}Y_0 = 0$ and $\lim_{i \rightarrow \infty} \mathbb{E}Y_i = \mathbb{E}Y \in \mathbb{R}$, $(\mathcal{F}_i)_{i \geq 0}$ are σ -algebras, and $(N_i)_{i \geq 1}$ are non-negative integer-valued random variables, such that $\mathbb{E}(N_i | \mathcal{F}_{i-1}) \in (0, \infty)$ almost surely for all $i \geq 1$.

Condition 26. Suppose Condition 25 holds, and for all $i, j \geq 1$:

- (i) $\mathbb{E}(\Delta_i^{(j)} \mathbb{I}\{j \leq N_i\} | \mathcal{F}_{i-1}) = (\mathbb{E}Y_i - \mathbb{E}Y_{i-1}) \mathbb{P}(j \leq N_i | \mathcal{F}_{i-1})$ almost surely.
- (ii) There exists random variables C_i such that $\mathbb{E}(|\Delta_i^{(j)}| \mathbb{I}\{j \leq N_i\} | \mathcal{F}_{i-1}) \leq C_i \mathbb{P}(j \leq N_i | \mathcal{F}_{i-1})$ and with $c_i := \mathbb{E}C_i < \infty$.
- (iii) $\sum_{i=1}^{\infty} c_i < \infty$ and $\sum_i N_i < \infty$ almost surely.

Remark 27. Condition 26 allows $(N_i)_{i \geq 1}$ to depend on $(\Delta_i^{(j)})_{i,j}$ in a stopping time fashion. Namely, suppose that Condition 25 holds and $(\Delta_i^{(j)})_{i,j}$ are as in Theorem 7.

- (i) If $(\mathcal{G}_i^{(j)})_{i \geq 1, j \geq 0}$ are σ -algebras such that $\mathcal{F}_{i-1} \subset \mathcal{G}_i^{(j)}$, $\{N_i \leq j\} \in \mathcal{G}_i^{(j)}$ and $\Delta_i^{(j)}$ is independent of $\mathcal{G}_i^{(j-1)}$ for all $i, j \geq 1$, then, Condition 26 (i) holds because

$$\mathbb{E}(\Delta_i^{(j)} \mathbb{I}\{j \leq N_i\} | \mathcal{F}_{i-1}) = \mathbb{E}[\mathbb{E}(\Delta_i^{(j)} | \mathcal{G}_i^{(j-1)}) \mathbb{I}\{j \leq N_i\} | \mathcal{F}_{i-1}] = \mathbb{E}\Delta_i \mathbb{P}(j \leq N_i | \mathcal{F}_{i-1}).$$

Condition 26 (ii) holds similarly, with $c_i = C_i = \mathbb{E}|\Delta_i|$.

- (ii) In particular, if $(\Delta_i^{(j)})_{i,j}$ are independent and N_i are stopping times with respect to $(\Delta_i^{(j)})$, in the sense that $\{N_i \leq j\} \in \mathcal{G}_i^{(j)} := \sigma(\Delta_k^{(1)}, \dots, \Delta_k^{(j)} : 1 \leq k \leq i)$ and $\mathcal{F}_{i-1} \subset \mathcal{G}_i^{(j)}$ for all $i, j \geq 1$, then Condition 26 holds as above.
- (iii) If $c_i = \mathbb{E}|\Delta_i|$ and Condition 2 holds for some probability distribution $(p_i)_{i \geq 1}$, then $\sum_{i \geq 1} c_i = \sum_{i \geq 1} c_i (\mathbb{I}\{c_i \leq p_i\} + \mathbb{I}\{c_i > p_i\}) \leq 1 + \sum_{i \geq 1} c_i^2/p_i < \infty$.

Theorem 28. *Assume Condition 25 holds and denote*

$$Z_m := \sum_{i=1}^m \frac{1}{\mathbb{E}(N_i | \mathcal{F}_{i-1})} \sum_{j=1}^{\infty} \Delta_i^{(j)} \mathbb{I}\{j \leq N_i\},$$

and $Z := \lim_{m \rightarrow \infty} Z_m$ whenever well-defined.

- (i) *If Condition 26 (i) and (ii) hold, then $\mathbb{E}Z_m = \mathbb{E}Y_m$. If also Condition 26 (iii) holds, then $\mathbb{E}Z = \mathbb{E}Y$.*
- (ii) *Suppose the assumptions of Theorem 7 hold and $(\Delta_i^{(j)})_{i,j}$ are independent of $(\mathcal{F}_i)_{i \geq 0}$. If*

$$v_{\ell,m} := \sum_{i,k=\ell+1}^m \left[\text{cov}(\Delta_i, \Delta_k) \mathbb{E} \left(\frac{N_i \wedge N_k}{\mathbb{E}(N_i | \mathcal{F}_{i-1}) \mathbb{E}(N_k | \mathcal{F}_{k-1})} \right) + \mathbb{E}\Delta_i \mathbb{E}\Delta_k \mathbb{E} \left(\frac{N_i N_k}{\mathbb{E}(N_i | \mathcal{F}_{i-1}) \mathbb{E}(N_k | \mathcal{F}_{k-1})} - 1 \right) \right] < \infty$$

for all $0 \leq \ell < k < \infty$ and $\lim_{k \rightarrow \infty} \sup_{j \geq 1} v_{m_k, m_k+j} = 0$ for some subsequence $(m_k)_{k \geq 1}$, then $\mathbb{E}Z = \mathbb{E}Y$ and $\text{var}(Z) = \lim_{k \rightarrow \infty} v_{0, m_k}$.

Proof of Theorem 28 is given in Appendix E.

Theorem 28 (ii) is a generalisation of Theorem 7, and leads to new, potentially interesting estimators. For instance, let $N_i := \sum_{j=1}^n \mathbb{I}\{R^{(j)} = i\}$ where $R^{(1)}, \dots, R^{(n-1)}$ are positive random integers independent of $R^{(n)} \sim (p_i)_{i \geq 1}$. If we take $\mathcal{F}_i := \sigma(R^{(1)}, \dots, R^{(n-1)})$ for all $i \geq 0$, then $\mathbb{E}(N_i | \mathcal{F}_{i-1}) = p_i + \sum_{j=1}^{n-1} \mathbb{I}\{R^{(j)} = i\}$. This can be viewed as a residual sampling scheme applied with the (random) probability distribution $\hat{p}_i := \mathbb{E}(N_i | \mathcal{F}_{i-1})/n$. Analogously, the residual sampling scheme may be viewed through such conditioning, where $R^{(1)}, \dots, R^{(n-r)}$ are deterministic. It is unclear whether Theorem 28 (ii) allows estimators that have practical appeal, such as greater efficiency compared with the estimators introduced in Section 5.

Theorem 28 (i) with Remark 27 is similar to Wald's identity. The stopping time formulation may prove theoretically useful, and suggests a possibility to be used in conjunction with stopping rules developed in the MLMC context (e.g. Collier et al., 2015). However, the practical relevance of such an approach may be limited, because the expectation of a stopping time is often unavailable. It appears also difficult to derive useful explicit variance expressions when (N_i) depend on $(\Delta_i^{(j)})$. The sequential refinement of $(p_i)_{i \geq 1}$ during repeated simulation of unbiased estimators, as discussed in Section 6, could be used instead.

8. NUMERICAL EXPERIMENTS

The performance of the proposed estimators was studied with four SDE examples, where expectations of final-value functionals are estimated. Three of the models were the same as those used by Rhee and Glynn (2015). The first is a *geometric Brownian motion* (gBM)

$$dX_t = \mu X_t dt + \sigma X_t dB_t, \quad t \in [0, 1], \quad X_0 \equiv 1,$$

with $\mu = 0.05$ and $\sigma = 0.2$, where $(B_t)_{t \geq 0}$ is the standard Brownian motion. The target functional is the final value European option $f(x) = e^{-\mu}(x-1)_+$ with approximate expected value $\mathbb{E}f(X_1) = 0.104505836$ (Rhee and Glynn, 2015).

The second model was a *Cox-Ingersoll-Ross* (CIR) model

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dB_t, \quad t \in [0, 1], \quad X_0 \equiv 0.04,$$

with $\kappa = 5$, $\theta = 0.04$ and $\sigma = 0.25$, and with final value European option $f(x) = e^{-0.05}(x-0.03)_+$. According to extensive simulations, the expected value was found to be approximately $\mathbb{E}f(X_1) = 0.01142686$.

The third model is a bivariate *Heston* model

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{V_t} S_t dB_t^{(1)} \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dB_t^{(2)}, \end{aligned} \quad t \in [0, 1], \quad \begin{aligned} S_0 &\equiv 1 \\ V_0 &\equiv 0, \end{aligned}$$

with $\mu = 0.05$, $\kappa = 5$, $\theta = 0.04$, $\sigma = 0.25$, and where $(B_t^{(1)}, B_t^{(2)})$ are coordinates of a correlated Brownian motion with coefficient $\rho = -0.5$. The functional $f(s, v) = f(s) = e^{-\mu}(s-1)_+$, with expected value $\mathbb{E}f(S_1) = 0.10459672$ (Kahl and Jäckel, 2006; Rhee and Glynn, 2015).

The fourth model is an artificial model termed *modified gBM*, which has the same volatility term as gBM but a time-dependent drift:

$$dX_t = t^2 X_t dt + \sigma X_t dB_t, \quad t \in [0, 1], \quad X_0 \equiv 1,$$

and $\sigma = 0.1$. The target functional is the mean, which was found to have an approximate expected value $\mathbb{E}X_1 = 1.395612139$. This last model is intended to have bigger $|\mathbb{E}Y_i - \mathbb{E}Y|$ than the previous examples, highlighting the differences between the new estimators and averages of independent estimators.

Algorithm 1 summarises an implementation of the single-term and coupled-sum estimators $Z_{\text{str}}^{(n)}$, $Z_{\text{sys}}^{(n)}$, $Z_{\Sigma, \text{str}}^{(n)}$ and $Z_{\Sigma, \text{sys}}^{(n)}$ in the SDE context (see Section 2). The random variables $(R^{(1)}, \dots, R^{(n)})$ may be constructed as in Definition 13 or Definition 14, for $Z_{\cdot, \text{str}}^{(n)}$ or $Z_{\cdot, \text{sys}}^{(n)}$, respectively. For $Z_{\cdot, \text{res}}^{(n)}$, one may construct $R^{(1)}, \dots, R^{(n)}$ as in the proof of Proposition 16 in Appendix C. The independent sum estimators may be implemented similarly as the coupled sum, by interchanging the lines 8 and 9. The C++ source code of the implementation developed for the tests is available at <https://bitbucket.org/mvihola/unbiased-mlmc>.

In all but the Heston model, the standard Milstein scheme described in Section 2 was employed. The antithetic truncated Milstein scheme proposed by Giles and Szpruch (2014) was applied with the Heston model. We considered single term, independent sum and coupled sum estimators. With the first two, the distributions $(p_i)_{i \geq 1}$ were set to approximately optimal in each model as discussed in Section 6: p_1, \dots, p_m were set empirically to minimise variance, and the tail probability P_{tail} was calculated from prior simulated data. The tail distribution p_i for $i > m$ was geometric, $p_i = P_{\text{tail}}(1 - 2^{-\gamma})2^{-\gamma(i-m)}$, with the parameter γ set in all cases to 1.5. This provided a very good fit with the empirical optimal distribution $\beta_i / \sum_{j \geq 1} \beta_j$ with $m = 1$ in case of gBM and $m = 6$ with all other examples; see also theoretical results on the Milstein scheme (Jentzen et al., 2009; Kloeden and Platen, 1992) and the antithetic truncated Milstein (Giles and Szpruch, 2014). In the coupled sum estimator, the differences $(\Delta_i)_{i \geq 1}$ were all based on a single Brownian trajectory. The optimisation was based on estimators of $\text{var}(Y - Y_i) \approx \text{var}(Y_m - Y_i)$, where Y_m corresponded to discretisation with a mesh of 2^{13} . An optimal subsequence was found using the dynamic programming algorithm of Rhee and Glynn (2015).

Figure 1 shows results based on averages of 10^5 independent runs of each algorithm in each model. The graphs in Figure 1 show the estimated inverse relative efficiency (IRE) of the

Algorithm 1 Unbiased estimators in the SDE context

```

1: function UNBIASED( $(p_i)_{i \geq 1}, R^{(1)}, \dots, R^{(n)}$ )
2:    $Z \leftarrow 0$ 
3:   for  $i = 1, \dots, \max_j R^{(j)}$  do
4:      $w_i \leftarrow \begin{cases} (np_i)^{-1} & \text{if single term} \\ [n(1 - \sum_{j=1}^{i-1} p_j)]^{-1} & \text{if coupled sum} \end{cases}$ 
5:   end for
6:   for  $j = 1, \dots, n$  do
7:      $L \leftarrow \begin{cases} \{R^{(j)} - 1\} & \text{if single term} \\ \{0, \dots, R^{(j)} - 1\} & \text{if coupled sum} \end{cases}$ 
8:     Simulate Brownian path  $B$  at mesh  $\tau_{R^{(j)}-1}$ .
9:     for  $\ell \in L$  do
10:       $Y_+ \leftarrow f(X_{2^\ell}^{(\tau_\ell)})$ 
11:       $Y_- \leftarrow \begin{cases} 0, & \ell = 0 \\ f(X_{2^{\ell-1}}^{(\tau_{\ell-1})}), & \ell > 0 \end{cases}$ 
12:      // where  $X_{2^\ell}^{(\tau_\ell)}$  and  $X_{2^{\ell-1}}^{(\tau_{\ell-1})}$  are based on  $B$ 
13:       $Z \leftarrow Z + w_{\ell+1}(Y_+ - Y_-)$ 
14:     end for
15:   end for
16:   return  $Z$ 
17: end function

```

TABLE 1. Estimated IREs in the experiments with $n = 10^6$.

	gBM			CIR			Heston			Modified gBM		
	Single	I.sum	C.sum	Single	I.sum	C.sum	Single	I.sum	C.sum	Single	I.sum	C.sum
$Z_{*,\text{ML}}^{(n)}$	0.0271	0.0269	0.0346	0.0769	0.0775	0.0145	1.0331	1.0401	0.3555	0.1149	0.1182	0.3024
$Z_{*,\text{iid}}^{(n)}$	0.0277	0.0308	0.0372	0.0843	0.0816	0.0149	1.1074	1.0579	0.3693	1.7757	14.489	11.482
$Z_{*,\text{str}}^{(n)}$	0.0271	0.0274	0.0356	0.0782	0.0780	0.0147	1.0479	1.0658	0.3691	0.1061	0.2400	0.3819
$Z_{*,\text{res}}^{(n)}$	0.0274	0.0278	0.0358	0.0780	0.0776	0.0147	1.0503	1.0589	0.3730	0.1368	0.5168	0.6809
$Z_{*,\text{sys}}^{(n)}$	0.0271	0.0274	0.0355	0.0778	0.0783	0.0147	1.0543	1.0652	0.3735	0.1119	0.2878	0.4476

methods: the average cost multiplied with average square deviation from the ground truth values as given above. According to the theory, this quantity is constant with independent averages, and with residual and stratified sampling, it converges to the same limit as the corresponding MLMC estimator. Table 1 shows the corresponding numerical values with $n = 10^6$.

The experiments appear to align well with the theoretical findings. In all but the last example, the MLMC estimator admitted the best IRE with small n , and the new estimators appear to admit performance in between the independent and the MLMC case. The performance of the new schemes come close to MLMC performance as n increases, as verified by the differences with $n = 10^6$ reported in Table 1, which are all relatively small, except for the sum estimators in the modified gBM example indicating still some discrepancy.

Note that the MLMC implemented in the experiments is the ‘idealised’ version involving same pre-determined allocation strategy as unbiased estimators (Description 18). This explains the discrepancy between the MLMC IRE reported here and by Rhee and Glynn

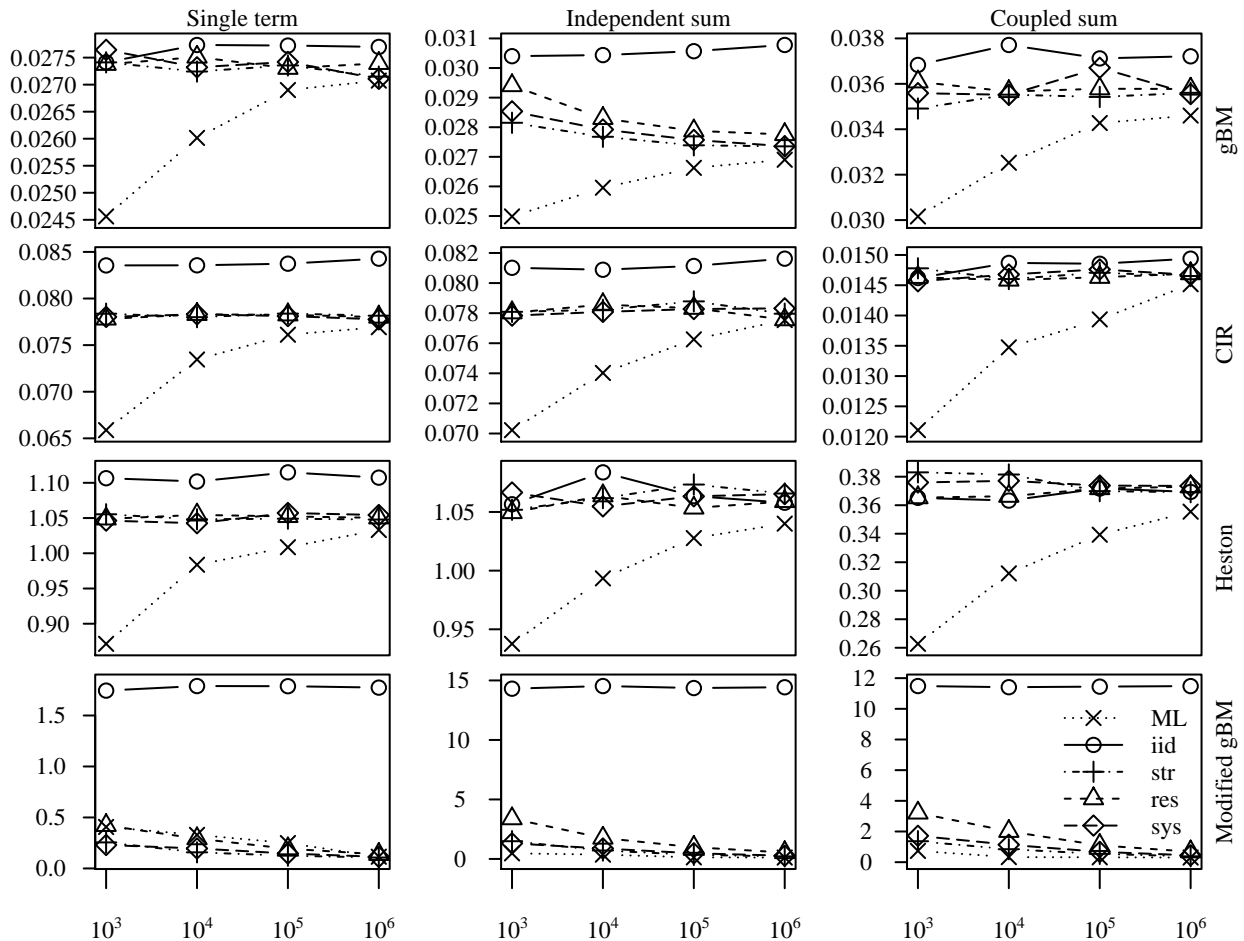


FIGURE 1. Estimated IRE over 10^5 replications of estimators with $n = 10^3$, 10^4 , 10^5 and 10^6 .

(2015), who employ the original version of the MLMC (Giles, 2008). The findings of Rhee and Glynn (2015) suggest that unbiased estimators applied with stopping rules may sometimes improve upon the original MLMC.

The differences in performance of single term and independent sum estimators with gBM, CIR and Heston examples are all relatively small, As n increases, the IREs of the new single term and sum estimators become negligible, as anticipated by the theory. The coupled sum estimator appears to admit greater efficiency with CIR and Heston examples, but the differences between independent average estimators and the new estimators are small. The modified gBM example demonstrates that the new estimators can be significantly more efficient. The numerical values shown in Table 1 indicate a 13–16 fold increase in terms of relative efficiency with the new single term estimators and similar performance with MLMC. The increase is 17–60 fold with sum estimators. In both cases, the stratified and systematic sampling estimators appear to perform slightly better than the residual sampling estimator.

9. DISCUSSION

This paper presented a general framework for unbiased MLMC estimation, which admits previous debiasing schemes as special cases, and accomodates new, lower variance unbiased estimators. The proposed stratified sampling and residual sampling schemes are promising

classes of estimators, as they enjoy good theoretical behaviour—they can not only improve on averages of independent estimators (Proposition 16 and Proposition 21), but also can have a significant gain in efficiency as illustrated in the experiments. Indeed, the stratified and residual sampling estimators can be made arbitrarily close to MLMC in efficiency under general assumptions (Theorem 19 and Proposition 22), highlighting the close connection between the MLMC and the debiasing schemes, and showing that unbiasedness may be achieved with virtually no sacrifice on efficiency. Unbiasedness is an important quality of estimators, when employed as part of stochastic optimisation algorithms (Borkar, 2008; Kushner and Yin, 2003), or in a ‘compound sampling’ context (Vihola et al., 2016). It also enables rigorous stopping rules, which can lead to benefits over MLMC stopping rules (Rhee and Glynn, 2015, §4).

While the theory presented in this paper does not give guarantees on the limiting efficiency of the systematic sampling, it is expected to behave often similar to stratified and residual sampling schemes, as illustrated by the experiments. However, as stratified and residual sampling enjoy good theoretical properties, and because systematic sampling appears to perform comparatively in practice, stratified or residual sampling are recommended as safer alternatives. The empirical evidence from the numerical experiments in Section 8 suggests that stratified sampling might sometimes perform slightly better than residual sampling. This, together with the straightforward implementation of stratified sampling, makes it appealing for practical purposes.

Averages of independent realisations of the single term and the independent sum estimators may have different efficiencies in general. In case of residual and stratified sampling, the optimally tuned estimators often coincide in asymptotic efficiency. In general, the single term estimator always dominates the independent sum estimator in terms asymptotic efficiency, rendering the single term estimator preferable over the independent sum estimator. Based on the experiments, the optimally tuned coupled sum estimator may sometimes lead into significant performance gains. This suggests also that dependent level estimators might be worth considering in the MLMC context, where independent level estimators are currently widely employed.

Despite the potential performance gain of the coupled estimators, it should be noted that tuning of $(p_i)_{i \geq 1}$ requires estimation of $\text{var}(Y - Y_i) \approx \text{var}(Y_m - Y_i)$, with m large, which is often computationally demanding. This is in sharp contrast with the single term estimator, where $\text{var}(\Delta_i)$ are easily accessible and inexpensive for moderate i . The simplicity of the single term estimator optimisation criterion suggests, as discussed in Section 6, an algorithm which automatically tunes the distribution $(p_i)_{i \geq 1}$ during repeated simulation of $Z_{\text{str}}^{(n)}$ or $Z_{\text{res}}^{(n)}$, based on earlier observed values for $\Delta_i^{(j)}$. Haji-Ali et al. (2016) suggest methods for finding optimal discretisation hierarchies in the MLMC context, which could also be explored in the debiasing context.

The MLMC literature provides many other potential further research topics. It is possible to employ essentially all techniques developed in the context of MLMC with debiasing. These include, for instance, quasi-Monte Carlo (Dick et al., 2013; Giles and Waterhouse, 2009), adaptive time steps (Hoel et al., 2012) or adaptive importance sampling scheme based on a drifted Brownian motion (Kebaier and Lelong, 2015). It is yet unclear how well the unbiased estimators can compete with MLMC in scenarios with slower than canonical rate of convergence. This was investigated by Rhee and Glynn (2015) and elaborated in Zheng et al. (2017), but quantifying the effect in the context of the new estimators requires further research. Recent work of Zheng and Glynn (2017) suggests an infinite stratification approach, which is slightly different what was proposed here.

ACKNOWLEDGEMENTS

This research was supported by the Academy of Finland (grants 274740 and 284513), and computational resources were provided by CSC, the IT Center for Science, Finland. The author wishes to thank Christel Geiss, Stefan Geiss, Peter Glynn, Chang-Han Rhee, Raul Tempone and Anni Toivola for useful discussions and remarks, and the associate editor and the reviewers of *Operations Research* for helpful comments, and for the suggestion to consider dependent randomisation, which led to the developments in Section 7.

REFERENCES

- S. Agapiou, G. O. Roberts, and S. J. Vollmer. Unbiased Monte Carlo: posterior estimation for intractable/infinite-dimensional models. Preprint [arXiv:1411.7713](https://arxiv.org/abs/1411.7713), 2014.
- M. B. Alaya and A. Kebaier. Central limit theorem for the multilevel Monte Carlo Euler method. *Ann. Appl. Probab.*, 25(1):211–234, 2015.
- A. Beskos and G. O. Roberts. Exact simulation of diffusions. *Ann. Appl. Probab.*, 15(4):2422–2444, 2005.
- V. S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008. ISBN 978-0-521-51592-4.
- N. Collier, A.-L. Haji-Ali, F. Nobile, E. von Schwerin, and R. Tempone. A continuation multilevel Monte Carlo algorithm. *BIT*, 55(2):399–432, 2015.
- B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, 27(1):94–128, 1999.
- S. Dereich and T. Mueller-Gronbach. General multilevel adaptations for stochastic approximation algorithms. Preprint [arXiv:1506.05482](https://arxiv.org/abs/1506.05482), 2015.
- J. Dick, F. Y. Kuo, and I. H. Sloan. High-dimensional integration: the quasi-Monte Carlo way. *Acta Numer.*, 22:133–288, 2013.
- R. Douc, O. Cappé, and E. Moulines. Comparison of resampling schemes for particle filtering. In *Proc. Image and Signal Processing and Analysis, 2005.*, pages 64–69, 2005.
- W. Feller. A limit theorem for random variables with infinite moments. *Amer. J. Math.*, 68(2):257–262, 1946.
- M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.
- M. B. Giles. Multilevel Monte Carlo methods. *Acta Numer.*, 24:259–328, 2015.
- M. B. Giles and L. Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional sdes without Lévy area simulation. *Ann. Appl. Probab.*, 24(4):1585–1620, 2014.
- M. B. Giles and B. J. Waterhouse. Multilevel quasi-Monte Carlo path simulation. In *Advanced Financial Modelling*, Radon Series on Computational and Applied Mathematics, pages 165–181. de Gruyter, 2009.
- P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer, 2003.
- P. Glynn. Randomized estimators for time integrals. Technical report, DTIC Document, 1983.
- P. W. Glynn and C.-H. Rhee. Exact estimation for Markov chain equilibrium expectations. *J. Appl. Probab.*, 51A:377–389, 2014.
- P. W. Glynn and W. Whitt. The asymptotic efficiency of simulation estimators. *Oper. Res.*, 40(3):505–520, 1992a.
- P. W. Glynn and W. Whitt. The asymptotic validity of sequential stopping rules for stochastic simulations. *Ann. Appl. Probab.*, 2(1):180–198, 1992b.

- A.-L. Haji-Ali, F. Nobile, E. von Schwerin, and R. Tempone. Optimization of mesh hierarchies in multilevel Monte Carlo samplers. *Stoch. Partial Differ. Equ. Anal. Comput.*, 4(1):76–112, 2016.
- M. H. Hansen, W. N. Hurwitz, and W. G. Madow. *Sample Survey Methods and Theory*, volume 1 Methods and Applications. Wiley, 1953.
- S. Heinrich. Multilevel Monte Carlo methods. In *Large-scale scientific computing*, pages 58–67. Springer, 2001.
- H. Hoel, E. Von Schwerin, A. Szepessy, and R. Tempone. Adaptive multilevel Monte Carlo simulation. In *Numerical Analysis of Multiscale Computations*, pages 217–234. Springer, 2012.
- M. L. Huber. *Perfect Simulation*. Chapman Hall/CRC, 2015.
- P. E. Jacob and A. H. Thiery. On nonnegative unbiased estimators. *Ann. Statist.*, 43(2):769–784, 2015.
- A. Jentzen, P. E. Kloeden, and A. Neuenkirch. Pathwise approximation of stochastic differential equations on domains: higher order convergence rates without global lipschitz coefficients. *Numer. Math.*, 112(1):41–64, 2009.
- C. Kahl and P. Jäckel. Fast stong approximation Monte Carlo schemes for stochastic volatility models. *Quant. Finance*, 6(6):513–536, 2006.
- A. Kebaier and J. Lelong. Coupling importance sampling and multilevel Monte Carlo using sample average approximation. Preprint [arXiv:1510.03590](https://arxiv.org/abs/1510.03590), 2015.
- P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer, 1992.
- H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Number 35 in Applications of Mathematics: Stochastic Modelling and Applied Probability. Springer-Verlag, 2nd edition, 2003. ISBN 0-387-00894-2.
- A.-M. Lyne, M. Girolami, Y. Atchade, H. Strathmann, and D. Simpson. On Russian roulette estimates for Bayesian inference with doubly-intractable likelihoods. *Statist. Sci.*, 30(4):443–467, 2015.
- D. McLeish. A general method for debiasing a Monte Carlo estimator. *Monte Carlo Methods Appl.*, 17(4):301–315, 2011.
- J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms*, 9(1-2):223–252, 1996.
- C.-H. Rhee and P. W. Glynn. A new approach to unbiased estimation for SDE’s. In *Proceedings of the Winter Simulation Conference*, page 17. Winter Simulation Conference, 2012.
- C.-H. Rhee and P. W. Glynn. Unbiased estimation with square root convergence for SDE models. *Oper. Res.*, 63(5):1026–1043, 2015.
- H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22:400–407, 1951.
- T. Rychlik. Unbiased nonparametric estimation of the derivative of the mean. *Statist. Probab. Lett.*, 10(4):329–333, 1990.
- T. Rychlik. A class of unbiased kernel estimates of a probability density function. *Appl. Math. (Warsaw)*, 22(4):485–497, 1995.
- H. Strathmann, D. Sejdinovic, and M. Girolami. Unbiased Bayes for big data: Paths of partial posteriors. Preprint [arXiv:1501.03326](https://arxiv.org/abs/1501.03326), 2015.
- M. Vihola. Unbiased estimators and multilevel Monte Carlo. Preprint [arXiv:1512.01022v2](https://arxiv.org/abs/1512.01022v2), 2015.
- M. Vihola, J. Helske, and J. Franks. Importance sampling type correction of Markov chain Monte Carlo and exact approximations. Preprint [arXiv:1609.02541](https://arxiv.org/abs/1609.02541), 2016.

- C. Walter. Point process-based Monte Carlo estimation. *Statist. Comput.*, 27(1):219–236, 2017.
- Z. Zheng and P. W. Glynn. A CLT for infinitely stratified estimators, with applications to debiased MLMC. Preprint, 2017.
- Z. Zheng, J. Blanchet, and P. W. Glynn. Rates of convergence and CLTs for subcanonical debiased MLMC. Preprint, 2017.

APPENDIX A. SUBSEQUENCE LEMMA

The following lemma is due to an argument by [Rhee and Glynn \(2015\)](#).

Lemma 29. *Suppose $\mathbb{E}(Y_i - Y)^2 \rightarrow 0$ as $i \rightarrow \infty$. Then, there exists strictly increasing $(m_k)_{k \geq 1}$ such that for all $k \geq 1$,*

$$\mathbb{E}(Y_i - Y_{m_k})^2 \leq 4\mathbb{E}(Y_i - Y)^2 \quad \text{for all } m_1 \leq i \leq m_k.$$

Proof. Let $\delta_i := Y_i - Y$. If $\mathbb{E}\delta_i^2 = 0$ for infinitely many i , choose them as $(m_k)_{k \geq 1}$. Otherwise, let m_1 be such that $\mathbb{E}\delta_i^2 > 0$ for all $i \geq m_1$, and define recursively $m_{k+1} := \min\{i > m_k : \mathbb{E}\delta_i^2 \leq \mathbb{E}\delta_{m_k}^2\}$. Now, for any k and any $m_1 \leq i \leq m_k$,

$$\mathbb{E}(Y_i - Y_{m_k})^2 \leq 2\mathbb{E}\delta_i^2 + 2\mathbb{E}\delta_{m_k}^2 \leq 4\mathbb{E}\delta_i^2. \quad \square$$

APPENDIX B. PROOF OF THEOREM 7

Define $Z_0 := 0$ and for $m \geq 1$

$$Z_m := \sum_{i=1}^m \frac{1}{\mathbb{E}N_i} \sum_{j=1}^{\infty} \Delta_i^{(j)} \mathbb{I}\{j \leq N_i\},$$

then by dominated convergence

$$\mathbb{E}Z_m = \sum_{i=1}^m \frac{\mathbb{E}\Delta_i}{\mathbb{E}N_i} \mathbb{E}\left[\sum_{j=1}^{\infty} \mathbb{I}\{j \leq N_i\}\right] = \sum_{i=1}^m \mathbb{E}\Delta_i = \mathbb{E}Y_m,$$

because $\sum_{j=1}^{\infty} \mathbb{I}\{j \leq N_i\} = N_i$.

For $i, k \geq 1$, by dominated convergence,

$$\begin{aligned} (5) \quad & \mathbb{E}[(Z_i - Z_{i-1})(Z_k - Z_{k-1})] \\ &= \frac{1}{\mathbb{E}N_i \mathbb{E}N_k} \mathbb{E}\left[\sum_{j,\ell=1}^{\infty} \Delta_i^{(j)} \Delta_k^{(\ell)} \mathbb{I}\{j \leq N_i\} \mathbb{I}\{\ell \leq N_k\}\right] \\ &= \frac{1}{\mathbb{E}N_i \mathbb{E}N_k} \left(\mathbb{E}(\Delta_i \Delta_k) \mathbb{E}\left[\sum_{j=1}^{\infty} \mathbb{I}\{j \leq N_i\} \mathbb{I}\{j \leq N_k\}\right] \right. \\ & \quad \left. + \mathbb{E}\Delta_i \mathbb{E}\Delta_k \mathbb{E}\left[\sum_{\substack{j,\ell=1 \\ j \neq \ell}}^{\infty} \mathbb{I}\{j \leq N_i\} \mathbb{I}\{\ell \leq N_k\}\right] \right) \\ &= \frac{(\mathbb{E}(\Delta_i \Delta_k) - \mathbb{E}\Delta_i \mathbb{E}\Delta_k) \mathbb{E}(N_i \wedge N_k) + \mathbb{E}\Delta_i \mathbb{E}\Delta_k \mathbb{E}[N_i N_k]}{\mathbb{E}N_i \mathbb{E}N_k}, \end{aligned}$$

because $\sum_{j,\ell=1}^{\infty} \mathbb{I}\{j \leq N_m\} \mathbb{I}\{\ell \leq N_k\} = N_m N_k$.

We deduce that for $0 \leq \ell < m$,

$$(6) \quad \mathbb{E}(Z_m - Z_\ell)^2 = \sum_{i,k=\ell+1}^m \mathbb{E}[(Z_i - Z_{i-1})(Z_k - Z_{k-1})] = v_{\ell,m} + (\mathbb{E}Y_m - \mathbb{E}Y_\ell)^2.$$

Therefore, by assumption $(Z_{m_k})_{k \geq 1}$ is Cauchy in L^2 . Because $\sum_{i \geq 1} N_i < \infty$ a.s., we have $Z_{m_k} \rightarrow Z$ a.s., and therefore $Z_{m_k} \rightarrow Z$ in L^2 . We deduce that $\mathbb{E}Z = \lim_{k \rightarrow \infty} \mathbb{E}Z_{m_k} = \lim_{k \rightarrow \infty} \mathbb{E}Y_{m_k} = \mathbb{E}Y$. Similarly we find the expression $\mathbb{E}Z^2 = \lim_{k \rightarrow \infty} \mathbb{E}Z_{m_k}^2 = \lim_{k \rightarrow \infty} [v_{0, m_k} + (\mathbb{E}Y_{m_k})^2]$. \square

APPENDIX C. PROOF OF PROPOSITION 16

Let us consider first $Z_{\text{str}}^{(n)}$, which is an average of independent random variables $(X_i^{(1)})_{i=1, \dots, n}$ which follow, respectively, the conditional distribution of the single term estimator $Z^{(1)}$, given the uniform random variable U generating $R = F^{-1}(U)$ takes value in $I_i := (\frac{i-1}{n}, \frac{i}{n})$. The desired variance bound follows from Lemma 12 (ii) applied with $(X_i^{(1)})$, $\ell_i = 1$, $\ell = n$ and $q_i = n^{-1}$. The sum estimator $Z_{\Sigma, \text{str}}^{(n)}$ is similarly stratified version of the average of n independent sum estimators.

The systematic sampling estimators $Z_{\text{sys}}^{(n)}$ and $Z_{\Sigma, \text{sys}}^{(n)}$, are averages as above, but with the difference that the uniformly distributed random variables on I_i which determine $X_i^{(1)}$ are not independent. We may apply Lemma 12 (i) which gives the (pessimistic) upper bound on the variance.

The residual sampling $Z_{\text{res}}^{(n)}$ and $Z_{\Sigma, \text{res}}^{(n)}$ may also be seen as stratification of $Z^{(1)}$ and $Z_{\Sigma}^{(1)}$, but instead of considering $R = F^{-1}(U)$, we now let $R = g(U)$, where

$$g(u) := \begin{cases} \min\{k \in \mathbb{N} : (r/n) \sum_{i \geq 1} p_i^* \geq u\}, & u \in (0, r/n), \\ \min\{k \in \mathbb{N} : n^{-1} \sum_{i \geq 1} n_i \geq u - n/r\}, & u \in [r/n, 1). \end{cases}$$

It is direct to verify that $g(U) \sim (p_i)_{i \geq 1}$. Let $X_1^{(1)}, \dots, X_r^{(1)}$ be independent random variables with conditional distribution of $Z^{(1)}$ (resp. $Z_{\Sigma}^{(1)}$) given $U \in I_i$, respectively, and let $X_{r+1}^{(1)}, \dots, X_{r+1}^{(n-r)}$ be, similarly, independent conditional on $U \in (\frac{r}{n}, 1)$. We may apply Lemma 12 (ii) with $\ell_1 = \dots = \ell_r = 1$, $\ell_{r+1} = n - r$, $\ell = n$, $q_1 = \dots = q_r = n^{-1}$ and $q_{r+1} = (n - r)/n$. \square

APPENDIX D. PROOF OF THEOREM 19

Consider (i), and let N_i correspond to $Z_{\text{res}}^{(n)}$. We have for any $i, k \geq 1$

$$\text{cov}(N_i, N_k) = \text{cov}(N_i^*, N_k^*) = r(p_i^* \mathbb{1}\{i = k\} - p_i^* p_k^*),$$

so $|\text{cov}(N_i, N_k)| \leq r$. In case of $Z_{\text{str}}^{(n)}$, it is not difficult to check that the number of strata $I_j := (\frac{j-1}{n}, \frac{j}{n})$ partially overlapping i , that is, such that $i \in F^{-1}(I_j)$ but $F^{-1}(I_j) \setminus \{i\} \neq \emptyset$, is at most two, so $\text{var}(N_i) \leq 1$, and consequently $|\text{cov}(N_i, N_k)| \leq 1$.

Denote $Z_*^{(n, m)} := n^{-1} \sum_{i=1}^m p_i^{-1} \sum_{i=1}^{N_i} \Delta_i^{(j)}$ for $m \geq 1$, where $(N_i)_{i \geq 1}$ correspond either to the residual sampling or stratified sampling scheme. We have (cf. Example 9)

$$\left| \text{var}(Z_*^{(n, m)}) - \frac{1}{n} \sum_{i=1}^m \frac{\text{var}(\Delta_i)}{p_i} \right| \leq \frac{r}{n^2} \sum_{i, k=1}^m \frac{|\mathbb{E}\Delta_i \mathbb{E}\Delta_k|}{p_i p_k}.$$

Because $r/n \rightarrow 0$ as $n \rightarrow \infty$, we deduce that $\lim_{n \rightarrow \infty} n \text{var}(Z_*^{(n, m)}) = \sum_{i=1}^m \text{var}(\Delta_i)/p_i$.

On the other hand, $\text{var}(Z_*^{(n, m)} - Z_*^{(n, \ell)})$ is no greater than the corresponding variance of single term estimators (2), and therefore for $1 < \ell < m$,

$$n \text{var}(Z_*^{(n, m)} - Z_*^{(n, \ell)}) \leq \sum_{i=\ell+1}^m \frac{\mathbb{E}\Delta_i^2}{p_i} - (\mathbb{E}Y_m - \mathbb{E}Y_\ell)^2.$$

We deduce that for any $m_0 > 1$, due to independence,

$$\begin{aligned} |n\text{var}(Z_*^{(n)}) - \sigma_\infty^2| &= |n \lim_{m \rightarrow \infty} \text{var}(Z_*^{(n,m)}) - \sigma_\infty^2| \\ &\leq |n\text{var}(Z_*^{(n,m_0)}) - \sigma_\infty^2| + \limsup_{m \rightarrow \infty} n\text{var}(Z_*^{(n,m)} - Z_*^{(n,m_0)}). \end{aligned}$$

Both terms on the right can be made arbitrarily small by choosing m_0 large enough and letting $n \rightarrow \infty$.

An easy calculation shows that

$$\begin{aligned} n\text{var}(Z_{\text{ML}}^{(n)}) &= n \sum_{i=1}^{m_n} \frac{\mathbb{I}\{n_i > 0\}}{n_i} \text{var}(\Delta_i) \\ (7) \quad &= \sum_{i=1}^{m_n} \frac{\text{var}(\Delta_i)}{p_i} + \sum_{i=1}^{m_n} \left(\frac{n\mathbb{I}\{n_i > 0\}}{n_i} - \frac{1}{p_i} \right) \text{var}(\Delta_i), \end{aligned}$$

where the first term converges to σ_∞^2 as $n \rightarrow \infty$. For the latter, observe that

$$0 \leq \frac{n}{n_i} - \frac{1}{p_i} = \frac{np_i - n_i}{n_i p_i} \leq \frac{1}{n_i p_i} \leq \frac{1}{p_i},$$

and $n_i \rightarrow \infty$ as $n \rightarrow \infty$, so by dominated convergence the last sum in (7) vanishes.

Let us then turn into (ii). Let us calculate first

$$\begin{aligned} \text{var}(Z_{\Sigma, \text{ML}}^{(n)}) &= \sum_{i,k=1}^{\tilde{m}_n} \frac{\text{cov}(\Delta_i, \Delta_k) \hat{n}_{i \vee k}}{\hat{n}_i \hat{n}_k} \\ &= \sum_{i=1}^{\tilde{m}_n} \left(\frac{\text{var}(\Delta_i)}{\hat{n}_i} + 2 \sum_{k=i+1}^{\tilde{m}_n} \frac{\text{cov}(\Delta_i, \Delta_k)}{\hat{n}_i} \right) \\ &= \sum_{i=1}^{\tilde{m}_n} \frac{\text{var}(\Delta_i) + 2\text{cov}(\Delta_i, Y_{\tilde{m}_n} - Y_i)}{\hat{n}_i}, \end{aligned}$$

and because

$$\text{var}(\Delta_i) + 2\text{cov}(\Delta_i, Y_{\tilde{m}_n} - Y_i) = \text{var}(Y_{\tilde{m}_n} - Y_{i-1}) - \text{var}(Y_{\tilde{m}_n} - Y_i) =: \xi_i^{(\tilde{m}_n)},$$

we obtain

$$(8) \quad n\text{var}(Z_{\Sigma, \text{ML}}^{(n)}) = \sum_{i=1}^{\tilde{m}_n} \frac{\xi_i^{(\tilde{m}_n)}}{\tilde{p}_i} + \sum_{i=1}^{\tilde{m}_n} \left(\frac{n}{\hat{n}_i} - \frac{1}{\tilde{p}_i} \right) \xi_i^{(\tilde{m}_n)}.$$

Note that $|\xi_i^{(\tilde{m}_n)}| \leq \text{var}(Y_{\tilde{m}_n} - Y_{i-1}) \leq 2\text{var}(Y - Y_{\tilde{m}_n}) + 2\text{var}(Y - Y_{i-1})$, so

$$\frac{|\xi_i^{(\tilde{m}_n)}|}{\tilde{p}_i} \leq 4 \sup_{k \geq i} \frac{\text{var}(Y - Y_k)}{\tilde{p}_k}.$$

The first sum in (8) therefore converges to $\sigma_{\Sigma, \infty}^2$ as $n \rightarrow \infty$ and the latter sum in (8) converges to zero as above.

Consider then $Z_{\Sigma, \text{res}}^{(n)}$. As above, $\text{cov}(\tilde{N}_i, \tilde{N}_k) = \text{cov}(\tilde{N}_i^*, \tilde{N}_k^*) = r(\tilde{p}_{i \vee k}^* - \tilde{p}_i^* \tilde{p}_k^*)$. In case of $Z_{\Sigma, \text{str}}^{(n)}$, at most one stratum $F^{-1}(I_j)$ contains some $\ell < i$ and i , so $\text{var}(\tilde{N}_i) \leq 1/4$. Therefore, $|\text{cov}(\tilde{N}_i, \tilde{N}_k)| \leq r$ for both schemes, and the variance of $Z_*^{(n,m)} := n^{-1} \sum_{i=1}^m \tilde{p}_i^{-1} \sum_{j=1}^{\tilde{N}_i} \Delta_i^{(j)}$

admits the bound

$$\left| \text{var}(Z_*^{(n,m)}) - \frac{1}{n} \sum_{i=1}^m \left(\frac{\text{var}(\Delta_i)}{\tilde{p}_i} + 2 \sum_{k=i+1}^m \frac{\text{cov}(\Delta_i, \Delta_k)}{\tilde{p}_i} \right) \right| \leq \frac{r}{n^2} \sum_{i,j=1}^m \frac{|\mathbb{E}\Delta_i \mathbb{E}\Delta_k|}{\tilde{p}_i \tilde{p}_k}$$

The second term on the left equals $n^{-1} \sum_{i=1}^m \xi_i^{(m)}/\tilde{p}_i$, and clearly $\text{var}(Z_{\Sigma, \text{res}}^{(n,m)}) \rightarrow \sum_{i=1}^m \xi_i^{(m)}/\tilde{p}_i$ as $n \rightarrow \infty$. Now, take $(m_k)_{k \geq 1}$ from Lemma 29, then stratification implies that for $k, j \geq 1$,

$$n \text{var}(\tilde{Z}_*^{(n, m_{k+j})} - \tilde{Z}_*^{(n, m_k)}) \leq 4 \sum_{i=m_k+1}^{m_{k+j}} \frac{\mathbb{E}(Y_{i-1} - Y_{m_k})^2}{\tilde{p}_i} + (\mathbb{E}Y_{m_{k+j}} - \mathbb{E}Y_{m_k})^2.$$

We conclude as above by writing for any $k \geq 1$,

$$\begin{aligned} |n \text{var}(\tilde{Z}_*^{(n)}) - \sigma_{\Sigma, \infty}^2| &\leq |n \text{var}(\tilde{Z}_*^{(n, m_k)}) - \sigma_{\Sigma, \infty}^2| + \limsup_{j \rightarrow \infty} n |\text{var}(\tilde{Z}_*^{(n, m_{k+j})} - \tilde{Z}_*^{(n, m_k)}) \\ &\quad + 2 \text{cov}((\tilde{Z}_*^{(n, m_{k+j})} - \tilde{Z}_*^{(n, m_k)}), \tilde{Z}_*^{(n, m_k)})|. \end{aligned}$$

Because $n |\text{cov}(A, B)| \leq \sqrt{n \text{var}(A) n \text{var}(B)}$, all terms on the right can be made arbitrarily small by choosing m_k large enough and letting $n \rightarrow \infty$.

Finally, the proof of (iii) follows similarly as (i) and (ii). \square

APPENDIX E. PROOF OF THEOREM 28

Consider (i) and notice that

$$\mathbb{E} \left(\sum_{j=1}^n |\Delta_i^{(j)}| \mathbb{I}\{j \leq N_i\} \mid \mathcal{F}_{i-1} \right) \leq C_i \sum_{j=1}^n \mathbb{P}(j \leq N_i \mid \mathcal{F}_{i-1}) \leq C_i \mathbb{E}(N_i \mid \mathcal{F}_{i-1}).$$

Denote $S_i := \mathbb{E}(N_i \mid \mathcal{F}_{i-1})^{-1} \sum_{j \geq 1} \Delta_i^{(j)} \mathbb{I}\{j \leq N_i\}$, then $\mathbb{E}|S_i| \leq c_i$ and by dominated convergence a similar calculation yields

$$\mathbb{E} \left(\frac{S_i}{\mathbb{E}(N_i \mid \mathcal{F}_{i-1})} \mid \mathcal{F}_{i-1} \right) = \lim_{n \rightarrow \infty} \mathbb{E} \left(\sum_{j=1}^n \frac{(\mathbb{E}Y_i - \mathbb{E}Y_{i-1}) \mathbb{P}(j \leq N_i \mid \mathcal{F}_{i-1})}{\mathbb{E}(N_i \mid \mathcal{F}_{i-1})} \right) = \mathbb{E}Y_i - \mathbb{E}Y_{i-1},$$

which leads to $\mathbb{E}Z_m = \mathbb{E}Y_m$. For the latter claim, note that only finitely many of $\{N_i\}_{i \geq 1}$ are non-zero, so Z is well-defined, and the result follows by the dominated convergence theorem.

The statement (ii) is a generalisation of Theorem 7, and the proof follows similarly. Indeed, Condition 26 (i) and (ii) hold by independence, with $c_i = C_i = \mathbb{E}|\Delta_i|$, so part (i) implies $\mathbb{E}Z_m = \mathbb{E}Y_m$. Denoting $\mathcal{F}_{i,k} := \sigma(\mathcal{F}_{i-1}, \mathcal{F}_{k-1})$, a straightforward calculation similar to (5) yields

$$\mathbb{E}[(Z_i - Z_{i-1})(Z_k - Z_{k-1}) \mid \mathcal{F}_{i,k}] = \frac{\text{cov}(\Delta_i, \Delta_k) \mathbb{E}(N_i \wedge N_k \mid \mathcal{F}_{i,k}) + \mathbb{E}\Delta_i \mathbb{E}\Delta_k \mathbb{E}(N_i N_k \mid \mathcal{F}_{i,k})}{\mathbb{E}(N_i \mid \mathcal{F}_{i-1}) \mathbb{E}(N_k \mid \mathcal{F}_{k-1})},$$

and as in (6), $\mathbb{E}(Z_m - Z_\ell)^2 = v_{\ell, m} + (\mathbb{E}Y_m - \mathbb{E}Y_\ell)^2$. \square

MATTI VIHOLA, DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF JYVÄSKYLÄ
P.O.BOX 35, FI-40014 UNIVERSITY OF JYVÄSKYLÄ, FINLAND

E-mail address: matti.vihola@iki.fi